

---

# Representaciones Alternativas de la Demanda para la Integración de Energías Renovables a la Planificación de la Expansión de Transmisión

---



TESIS DE GRADO PRESENTADA COMO REQUISITO  
PARCIAL PARA OPTAR AL GRADO DE MAGÍSTER  
EN CIENCIAS DE LA INGENIERÍA ELÉCTRICA

Elizabeth Andrea Paduro Williamson

**Director de Tesis: Esteban Gil Sagás**

**Co-Director de Tesis: Ignacio Aravena Solís**

Departamento de Ingeniería Eléctrica  
Universidad Técnica Federico Santa María  
Valparaíso - Chile

Noviembre 2019



# “Representaciones Alternativas de la Demanda para la Integración de Energías Renovables a la Planificación de la Expansión de Transmisión”

**Elizabeth Andrea Paduro Williamson**

Tesis presentada al Comité de Evaluación  
de la Tesis integrado por:

*Director de Tesis:*

**Esteban Gil**

*Co-director de Tesis:*

**Ignacio Aravena**

*Presidente del Comité:*

**Victor Hinojosa**

*Co-referente Externo:*

**Rodrigo Palma**

Para completar las exigencias del grado de  
Magíster en Ingeniería Eléctrica

**Departamento de Ingeniería Eléctrica  
Universidad Técnica Federico Santa María  
Valparaíso - Chile**

**Noviembre 2019**



# Agradecimientos

*Solo el amor con su ciencia  
nos vuelve tan inocentes.*

Violeta Parra

Culminando por fin este largo trabajo y a la vez enfrentando este difícil momento a nivel país, quiero detenerme para agradecer profundamente a mis padres, quienes me han entregado los valores y la fortaleza para llegar hasta aquí. Sé que sus enseñanzas guiarán por siempre mi camino como ingeniera, esperando poder aportar de la mejor forma a una mejor sociedad.

Muchas gracias familia por ser mi pilar fundamental. Muchas gracias mamá por mostrarme desde siempre que podía lograr cualquier meta y a ti papá por ser mi apoyo en cada proyecto que emprendo. Gracias a mi abuelita Elsa, porque de ti heredé muchas mañas y también muchas cosas buenas, y a mi tatita Choche que se que desde el cielo está muy orgulloso de su nieta mayor. Gracias a mis hermanitos Teby, Taty y Benjy, por entregarme siempre amor incondicional, y por ser mis referentes para saber que no hay limitantes en cualquier cosa que me proponga.

Gracias a mi querido Luchito por toda la paciencia y amor que me entregas día a día, y particularmente por soportarme durante todo el desarrollo de mi tesis. Espero seguir compartiendo muchas más mañas y sobre todo alegrías.

De todo corazón agradecer a todas esas personas que marcaron mi etapa universitaria, mis amigas y amigos de universidad, mi querido equipo de basketball y entrenadores, mis chicas eléctricas con las que formamos un lazo hermoso y sororo durante estos años, mis queridos amigos y amigas que formé en los centros de alumnos en que pude participar, los “cabros” con quienes compartí esos largos tramos de estudios y muchos otros momentos felices que nunca se olvidarán, entre muchas más personas, que sin duda me han marcado y que valoraré por siempre de una manera muy particular.

Gracias a cada profesor que ha pasado por este largo camino, desde mi querida tía Mariluz, el profe Maxi, y muchos otros profesores que me han marcado durante mi paso en el colegio y universidad. Gracias a don Nelson, que pese al poco tiempo, logró transmitirme lo hermosa que puede ser esta carrera. De manera particular agradecer a mi profesor guía, el Dr. Esteban Gil, por confiar en mí y ser un muy buen guía y motivador durante el periodo de tesis, gracias por cada comentario y consejo durante el desarrollo de este trabajo.

Agradezco también el apoyo y colaboración del Dr. Ignacio Aravena, quien con muy buena disposición brindó ayuda para la realización de este trabajo, y gestionó la colaboración del *Lawrence Livermore National Laboratory*, con financiamiento del Departamento de Energía de Estados Unidos de América bajo el Contrato DE-AC52-07NA27344. Finalmente agradecer a CONICYT a través de los proyectos Basal FB0008 (AC3E) y Fondecyt 1151270.

# Resumen Ejecutivo

El problema de la Planificación de la Expansión de la Transmisión (TEP) puede volverse difícil de tratar computacionalmente a medida que se incluyen más variables en su formulación, por lo cual los planificadores generalmente limitan la representación de las condiciones operacionales del sistema.

De este modo, este trabajo propone metodologías de *clustering* para la selección de condiciones operativas en el TEP, considerando además la creciente integración de energías renovables variables (ERV) en los sistemas. Para lograr dicho objetivo, se contrasta el método tradicional de descomposición de la curva de duración de la demanda (LDC) con técnicas de *clustering* como k-means y PAM, que consideran la diversidad espacial de la demanda neta.

Los métodos propuestos se prueban en el sistema IEEE de 24 barras para diferentes niveles de penetración renovable variable (0 %, 5 %, 12 %, 26 % y 50 %), ubicando parques eólicos y solares en diferentes localizaciones, con el objetivo de mostrar los efectos de la variabilidad espacial y temporal en los perfiles de demanda neta. Las representaciones obtenidas se ingresan como entradas en el problema TEP y los resultados se comparan con la solución del problema completo (sin reducción por *clustering*) utilizando diferentes métricas comparativas.

Los resultados numéricos muestran que el número óptimo de *clusters* depende de la variabilidad de los perfiles de demanda neta. Además, la selección de condiciones operativas para los estudios de TEP que utilizan las técnicas de *clustering* propuestas reducen el error con respecto al método LDC a medida que la demanda neta se vuelve más diversa espacialmente.

La segunda parte de este trabajo presenta una metodología alternativa para la reducción de escenarios de demanda, basando su *clustering* en un análisis post-óptimo de variables de interés para el problema de expansión de la transmisión. La metodología utilizada consiste en el pre-procesamiento de los datos de entrada a través de un flujo de potencia óptimo, cuyos resultados relacionados con los costos de operación y los flujos a través de las líneas se procesan y agrupan para rescatar grupos de demanda representativos a partir de ellos.

Las metodologías propuestas realizan su proceso de *clustering* basadas en las técnicas de k-means, PAM y la conocida LDC (demanda neta), siendo PAM la técnica con mayor requerimiento computacional asociado. De este modo, se agrupan los respectivos indicadores horarios de las variables de interés y a través de estos realizar los respectivos grupos de demanda del sistema. De manera adicional, se presenta una variable relativa a

los flujos del sistema que considera información de interés respecto a los flujos de líneas con y sin límites de capacidad de transmisión. Las metodologías propuestas se contrastan con la metodología propuesta en la primera parte, que consiste en el *clustering* directo en la demanda.

Los resultados numéricos indican que la selección de una metodología adecuada depende directamente del nivel de variabilidad del sistema. De manera general, el agrupar a partir de la demanda y flujos de potencia sin límites de capacidad son las metodologías con mejor desempeño cuando se utiliza k-means y PAM en sus *clusters*. Así, para los casos de estudio con baja variabilidad el estas dos tienen un rendimiento similar, sin embargo en los casos de alta variabilidad el agrupamiento basado en PAM tiene el mejor desempeño.

# Abstract

The Transmission Expansion Planning (TEP) problem it is known to become computationally intractable as the number of variables are included in the formulation increase, this is why planners usually restrict their analysis to consider only a few representative operational conditions.

This work proposes a clustering methodology for the selection of operational conditions in TEP. In particular we study the effect of the rising in penetration of variable renewable energy. The traditional Load Duration Curve (LDC) decomposition method is contrasted against clustering techniques such as k-means and PAM that consider the spatial diversity of the net load.

The proposed method is tested in the 24-bus IEEE Test System for different renewable penetration levels (0 %, 5 %, 12 %, 26 % and 50 %), locating wind and solar farms in different buses to show the effects of spatial and temporal variability in the net load profiles. The clusters are then used in the TEP problem and results are compared against the solution of the full problem without clustering using different metrics.

Numerical results show that the optimal number of clusters depends on the variability of the net load profiles. Furthermore, the selection of operational conditions for TEP studies using clustering techniques reduces the error with respect to the LDC method as the net load becomes more spatially diverse.

The second half of this work presents an alternative methodology to the reduction of load scenarios, basing its clustering on a post-optimal analysis of interesting variables for TEP. The methodology consists of a preprocessing of input data through an optimal power flow, whose outputs related to operating costs and flows through the lines are used to cluster load groups.

The proposed methodologies carry out their clustering process based on the techniques of k-means, PAM and the well-known LDC (net demand), being PAM the technique with the highest associated computational requirement. In this way, the respective time indicators of the analyzed variables and through these performing the respective demand groups of the system. Additionally, it is proposed to use a variable related to the system flows that considers relevant information regarding the flows of lines with and without capacity limits is presented. The proposed methodologies are tested in the same test system and load scenarios as the first half of the work.

The numerical results indicate that the proper selection of a methodology depends directly on the variability level of the system. In general, grouping based on demand and power flows without capacity constraints are the best performing methodologies

when using k-means and PAM in their clusters. Thus, for the cases of study with low variability both have similar performance, however, in the cases of high variability the clustering based on PAM provide better results.

# Abreviaturas

<i>ER</i>	Energía(s) Renovable(s)
<i>ERV</i>	Energía(s) Renovable(s) Variable(s)
<i>SEP</i>	Sistema(s) Eléctrico(s) de potencia
<i>OE</i>	Operación Económica
<i>OPF</i>	<i>Optimal Power Flow</i> o Flujo óptimo de potencia
<i>CEP</i>	<i>Capacity Expansion Planning</i>
<i>GEP</i>	<i>Generation Expansion Planning</i>
<i>GEPMS</i>	<i>Generation Expansion Planning Models</i>
<i>TEP</i>	<i>Transmission Expansion Planning</i>
<i>TEPMS</i>	<i>Transmission Expansion Planning Models</i>
<i>LDC</i>	Load Duration Curve o Curva de Duración de la Demanda
<i>CA</i>	Clustering Algorithm o Algoritmo de Clustering

## *Variables del problema*

<i>L</i>	Conjunto de trazos de líneas
<i>B</i>	Conjunto de barras del sistema
<i>Y</i>	Conjunto de nuevas líneas candidatas
<i>T</i>	Conjunto de estados del sistema
<i>G</i>	Conjunto de unidades generadoras del sistema
$L_{km,y}^{new}$	Variable binaria de decisión de construcción del nuevo tramo de línea $y$ , entre los nodos $k$ y $m$
$f_{km,t}^0$	Flujo de potencia entre las líneas iniciales del sistema
$f_{km,t}$	Flujo de potencia entre las líneas nuevas del sistema
$p_{j,t}$	Potencia en MW generada por el generador $j$ durante el tiempo $t$
$USE_{k,t}$	Energía no servida en MW, generada en el nodo $k$ durante el tiempo $t$
$p_{vert\ ERV\ k,t}$	Potencia en MW, vertida en el nodo $k$ durante el tiempo $t$
$\delta_{j,t}$	Ángulo relativo a la tensión de la barra $j$ en el tiempo $t$

*Parámetros del problema*

$\alpha_t$	Factor tasa de descuento del sistema en el periodo t. Se trabajará con su aproximación $\lim_{t \rightarrow \infty} \alpha_t = \frac{1}{r}$ , con r la tasa de descuento del sistema, y $\alpha_t$ el factor relativo a la tasa de descuento que lleva al valor presente el costo de inversión
$V_{oLL}$	<i>Value of Lost Load</i> o valor de pérdida de carga del sistema
$S_{Base}$	Potencia base del sistema
$N$	Número de intervalos de agrupamiento del problema
$h_t$	Ancho del intervalo asociado al grupo t
$c_{km}$	Costos de construcción de línea entre nodos k y m
$d_{k,t}$	Demanda en MW el nodo k, en el tiempo t
$n_{km}^0$	Número de circuitos iniciales entre los nodos k y m
$\bar{n}_{km}$	Número máximo de nuevos circuitos entre los nodos k y m
$b_{km}$	Susceptancia de línea entre nodos k y m
$\bar{f}_{km}$	Flujo máximo permitido para cada tramos de línea entre los nodos k y m
$\bar{p}_j$	Generación máxima disponible del generador j
$\underline{p}_j$	Generación mínima disponible del generador j
$\bar{\delta}_j$	Máximo valor del ángulo de tensión en la barra j
$C_{MW j}$	Costos variable de operación del generador j sistema en \$/MW

*Escenarios de estudio*

$ERV_0$	Escenario con 0% de generación ERV.
$ERV_{\frac{1}{5}}$	Escenario con 5% de generación ERV, ubicada en las barras 7, 15 y 17.
$ERV_{\frac{2}{5}}$	Escenario con 5% de generación ERV, ubicada en las barras 4, 13 y 22.
$ERV_{\frac{3}{5}}$	Escenario con 5% de generación ERV, ubicada en las barras 1, 6, 9, 15, 16 y 23.
$ERV_{\frac{1}{12}}$	Escenario con 12% de generación ERV, ubicada en las barras 7, 15 y 17.
$ERV_{\frac{2}{12}}$	Escenario con 12% de generación ERV, ubicada en las barras 4, 13 y 22.
$ERV_{\frac{3}{12}}$	Escenario con 12% de generación ERV, ubicada en las barras 1, 6, 9, 15, 16 y 23.
$ERV_{\frac{1}{26}}$	Escenario con 26% de generación ERV, ubicada en las barras 7, 15 y 17.
$ERV_{\frac{2}{26}}$	Escenario con 26% de generación ERV, ubicada en las barras 4, 13 y 22.
$ERV_{\frac{3}{26}}$	Escenario con 26% de generación ERV, ubicada en las barras 1, 6, 9, 15, 16 y 23.
$ERV_{50}$	Escenario con 50% de generación ERV (relativa al total de la demanda), distribuida en 9 barras del sistema.

# Índice

Agradecimientos	V
Resumen Ejecutivo	VII
Abstract	IX
Abreviaturas	XI
Índice	XIII
Índice de Figuras	XVI
Índice de Tablas	XVIII
<b>1 Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Hipótesis . . . . .	5
1.3. Objetivos . . . . .	6
1.3.1. Objetivo Principal . . . . .	6
1.3.2. Objetivos Específicos . . . . .	6
1.4. Metodología . . . . .	6
1.5. Alcances y estructura del documento . . . . .	8
<b>2 Marco Teórico y Estado del Arte</b>	<b>11</b>
2.1. Planificación y Operación de Sistemas de Potencia . . . . .	11
2.1.1. Operación Económica . . . . .	12
2.1.2. Planificación de Sistemas de Potencia . . . . .	15
2.2. Reducción de escenarios . . . . .	26
2.2.1. Tratamiento de Variables Aleatorias en TEP . . . . .	26
2.2.2. Temporalidad y Espacialidad . . . . .	27
2.2.3. Modelamiento de la Demanda . . . . .	30
2.2.4. Técnicas de Agrupamiento de Datos: Clustering . . . . .	33
<b>3 Reducción Dimensional de la Demanda Considerando Diversidad Es- pacial de Generación Renovable</b>	<b>39</b>
3.1. Introducción . . . . .	39

3.1.1.	Estado del Arte . . . . .	41
3.1.2.	Contribuciones y Organización del Capítulo . . . . .	43
3.2.	Metodología . . . . .	45
3.2.1.	El problema de la expansión de la Transmisión (TEP) . . . . .	45
3.2.2.	Curva de duración de la demanda . . . . .	46
3.2.3.	Técnicas de Agrupamiento de Datos: <i>Clustering</i> . . . . .	47
3.3.	Casos de estudio . . . . .	51
3.3.1.	Datos . . . . .	53
3.3.2.	Perfiles de demanda y generación ERV . . . . .	55
3.3.3.	Selección de escenarios de estudio . . . . .	58
3.4.	Resultados numéricos . . . . .	60
3.5.	Conclusiones . . . . .	68
3.6.	Anexos . . . . .	70
3.6.1.	Construcción de bloques de demanda [1]. . . . .	70
3.6.2.	Algoritmos de clustering [1]. . . . .	71
3.6.3.	Datos de entrada adicionales . . . . .	73
3.6.4.	Resultados adicionales . . . . .	75
<b>4</b>	<b>Reducción Dimensional de la Demanda en Base a Condiciones Operacionales de un Análisis Post-óptimo</b>	<b>79</b>
4.1.	Introducción . . . . .	80
4.1.1.	Estado del Arte . . . . .	81
4.1.2.	Contribuciones y Organización del Capítulo . . . . .	82
4.2.	Metodología . . . . .	83
4.2.1.	El problema de la expansión de la Transmisión (TEP) . . . . .	83
4.2.2.	Curva de duración de la demanda . . . . .	84
4.2.3.	Metodología de análisis post-óptimo . . . . .	87
4.3.	Casos de estudio . . . . .	92
4.3.1.	Datos . . . . .	94
4.3.2.	Selección de escenarios de estudio . . . . .	99
4.4.	Resultados numéricos . . . . .	102
4.4.1.	Análisis respecto a la Función Objetivo (FO) . . . . .	104
4.4.2.	Análisis respecto al MSE . . . . .	107
4.4.3.	Análisis respecto a la Función Objetivo Recalculada . . . . .	109
4.4.4.	Revisión detallada de casos particulares seleccionados . . . . .	112
4.5.	Conclusiones . . . . .	118
4.6.	Anexos . . . . .	120
4.6.1.	Resultados anexos . . . . .	120
<b>5</b>	<b>Conclusiones y Recomendaciones</b>	<b>129</b>
5.1.	Aportes . . . . .	129
5.2.	Conclusiones generales . . . . .	130
5.3.	Recomendaciones . . . . .	131
<b>A</b>	<b>Resumen</b>	<b>133</b>

ÍNDICE	XV
<b>Referencias</b>	<b>137</b>
<b>Bibliografía</b>	<b>137</b>

# Índice de Figuras

1.1.1. Esquema general del trabajo propuesto. . . . .	4
2.1.1. Mapa conceptual variables de estudio problema de Planificación Expansión de la Capacidad del Sistema. [Fuente: Elaboración propia] . . . . .	16
2.1.2. % Variación FO (formulación de McCormick) respecto a solución original relajada (McCormick) para diferentes niveles de generación renovable $ERV_0$ y $ERV_{50}$ . . . . .	24
2.1.3. % Variación FO (formulación de McCormick) respecto a solución original para diferentes niveles de generación renovable $ERV_0$ y $ERV_{50}$ . . . . .	24
2.1.4. % Variación FO (formulación no-lineal) respecto a solución original para diferentes niveles de generación renovable $ERV_0$ y $ERV_{50}$ . . . . .	25
2.2.1. Subdivisiones del área de Aprendizaje de Máquinas. [Fuente: Elaboración propia a partir de [2]] . . . . .	34
3.2.1. Esquema de reducción de la LDC. . . . .	47
3.2.2. Esquema de reducción para k-means y PAM. . . . .	50
3.3.1. Subdivisión de casos de estudio por localización y nivel de generación ERV. . . . .	51
3.3.2. Esquema del trabajo propuesto. . . . .	52
3.3.3. Esquema del trabajo propuesto. . . . .	53
3.3.4. Localización de generadores ERV en el sistema y costos de operación de centrales. [Fuente: Elaboración propia.] . . . . .	54
3.3.5. Gráficas de correlación espacial entre perfiles de demanda neta por barra. . . . .	55
3.3.6. Gráficas de perfiles de demanda neta promedio por barra. . . . .	56
3.3.7. Gráficas de perfiles de demanda neta promedio por barra. . . . .	57
3.4.1. Métricas de evaluación de casos de estudio para $ERV_5^1$ , $ERV_5^2$ y $ERV_5^3$ . . . . .	63
3.4.2. Métricas de evaluación de casos de estudio para $ERV_{12}^1$ , $ERV_{12}^2$ y $ERV_{12}^3$ . . . . .	64
3.4.3. Métricas de evaluación de casos de estudio para $ERV_{26}^1$ , $ERV_{26}^2$ y $ERV_{26}^3$ . . . . .	65
3.4.4. Métricas de evaluación de casos de estudio para $ERV_0$ y $ERV_{50}$ . . . . .	66
3.6.1. Métricas de evaluación de casos de estudio para $ERV_5^1$ , $ERV_5^2$ y $ERV_5^3$ . . . . .	75
3.6.2. Métricas de evaluación de casos de estudio para $ERV_{12}^1$ , $ERV_{12}^2$ y $ERV_{12}^3$ . . . . .	76
3.6.3. Métricas de evaluación de casos de estudio para $ERV_{26}^1$ , $ERV_{26}^2$ y $ERV_{26}^3$ . . . . .	77
3.6.4. Métricas de evaluación de casos de estudio para $ERV_0$ y $ERV_{50}$ . . . . .	78
4.2.1. Esquema típico de reducción de la LDC. . . . .	85
4.2.2. Esquema de reducción para k-means y PAM. . . . .	87

4.2.3. Esquema de reducción según Costos de Operación a través de LDC, k-means y PAM. . . . .	89
4.2.4. Esquema de reducción según Flujos por las líneas a través de LDC. . . . .	90
4.2.5. Esquema de reducción según Flujos por las líneas a través de k-means y PAM. . . . .	90
4.3.1. Subdivisión de casos de estudio por localización y nivel de generación ERV. . . . .	92
4.3.2. Esquema del trabajo propuesto. . . . .	93
4.3.3. Esquema del trabajo propuesto. . . . .	94
4.3.4. Localización de generadores ERV en el sistema y costos de operación de centrales. . . . .	95
4.3.5. Gráficas de correlación espacial entre perfiles de demanda por barra. . . . .	96
4.3.6. Gráficas de perfiles de demanda promedio por barra. . . . .	97
4.3.7. Gráficas de perfiles de demanda promedio por barra. . . . .	98
4.4.1. Error % FO para $ERV_5^1$ , $ERV_5^2$ y $ERV_5^3$ . . . . .	104
4.4.2. Error % FO para $ERV_{12}^1$ , $ERV_{12}^2$ y $ERV_{12}^3$ . . . . .	104
4.4.3. Error % FO para $ERV_{26}^1$ , $ERV_{26}^2$ y $ERV_{26}^3$ . . . . .	105
4.4.4. Error % FO para $ERV_0$ y $ERV_{50}$ . . . . .	105
4.4.5. MSE para $ERV_5^1$ , $ERV_5^2$ y $ERV_5^3$ . . . . .	107
4.4.6. MSE para $ERV_{12}^1$ , $ERV_{12}^2$ y $ERV_{12}^3$ . . . . .	107
4.4.7. MSE para $ERV_{26}^1$ , $ERV_{26}^2$ y $ERV_{26}^3$ . . . . .	108
4.4.8. MSE para $ERV_0$ y $ERV_{50}$ . . . . .	108
4.4.9. Error % FO recalculada para $ERV_5^1$ , $ERV_5^2$ y $ERV_5^3$ . . . . .	109
4.4.10. Error % FO recalculada para $ERV_{12}^1$ , $ERV_{12}^2$ y $ERV_{12}^3$ . . . . .	110
4.4.11. Error % FO recalculada para $ERV_{26}^1$ , $ERV_{26}^2$ y $ERV_{26}^3$ . . . . .	110
4.4.12. Error % FO recalculada para $ERV_0$ y $ERV_{50}$ . . . . .	111
4.4.13. Error % FO para $ERV_0$ y $ERV_{50}$ . . . . .	112
4.4.14. Error % FO para $ERV_{5,12,25}^{1,2,3}$ . . . . .	113
4.4.15. MSE para $ERV_0$ y $ERV_{50}$ . . . . .	113
4.4.16. MSE para $ERV_{5,12,25}^{1,2,3}$ . . . . .	114
4.4.17. Error % FO recalculada para $ERV_0$ y $ERV_{50}$ . . . . .	114
4.4.18. Error % FO recalculada para $ERV_{5,12,25}^{1,2,3}$ . . . . .	115
4.6.1. Error % nuevas líneas para $ERV_5^1$ , $ERV_5^2$ y $ERV_5^3$ . . . . .	120
4.6.2. Error % nuevas líneas para $ERV_{12}^1$ , $ERV_{12}^2$ y $ERV_{12}^3$ . . . . .	120
4.6.3. Error % nuevas líneas para $ERV_{26}^1$ , $ERV_{26}^2$ y $ERV_{26}^3$ . . . . .	121
4.6.4. Error % nuevas líneas para $ERV_0$ y $ERV_{50}$ . . . . .	121
4.6.5. MAE para $ERV_5^1$ , $ERV_5^2$ y $ERV_5^3$ . . . . .	123
4.6.6. MAE para $ERV_{12}^1$ , $ERV_{12}^2$ y $ERV_{12}^3$ . . . . .	123
4.6.7. MAE para $ERV_{26}^1$ , $ERV_{26}^2$ y $ERV_{26}^3$ . . . . .	124
4.6.8. MAE para $ERV_0$ y $ERV_{50}$ . . . . .	124
4.6.9. Correlación para $ERV_5^1$ , $ERV_5^2$ y $ERV_5^3$ . . . . .	126
4.6.10. Correlación para $ERV_{12}^1$ , $ERV_{12}^2$ y $ERV_{12}^3$ . . . . .	126
4.6.11. Correlación para $ERV_{26}^1$ , $ERV_{26}^2$ y $ERV_{26}^3$ . . . . .	127
4.6.12. Correlación para $ERV_0$ y $ERV_{50}$ . . . . .	127
A.0.1. Esquema del trabajo propuesto para estudio por caso. . . . .	135

# Índice de Tablas

2.2.1.	Cantidad de variables según número de <i>clusters</i> y formulación en el problema de expansión de la transmisión simplificado, implementado en el sistema RTS 96 zona A. . . . .	29
2.2.2.	Tabla resumen características principales de algunas metodologías de <i>Clustering</i> [2]. . . . .	37
3.4.1.	Tiempos de simulación para algunos cluster del caso $ERV_0$ . . . . .	61
3.4.2.	Tiempos de simulación para algunos clusters del caso $ERV_{50}$ . . . . .	61
3.4.3.	Resultados para 120 <i>clusters</i> . . . . .	62
3.6.1.	Datos de generadores del sistema . . . . .	73
3.6.2.	Datos de líneas del sistema . . . . .	74
4.2.1.	Tabla resumen características principales de variables de estudio escogidas . . . . .	91
4.4.1.	Tiempos de simulación para algunos <i>clusters</i> del caso $ERV_0$ . . . . .	102
4.4.2.	Tiempos de simulación para algunos <i>clusters</i> del caso $ERV_{50}$ . . . . .	103
4.4.3.	Variación porcentual de la FO para 120 <i>clusters</i> . . . . .	106
4.4.4.	MSE para 120 <i>clusters</i> . . . . .	109
4.4.5.	Variación porcentual de FO recalculada para 120 <i>clusters</i> . . . . .	111
4.4.6.	Ranking según FO . . . . .	115
4.4.7.	Ranking según MSE . . . . .	116
4.4.8.	Ranking según FO recalculada . . . . .	116
4.4.9.	N° de <i>clusters</i> mínimo para un 1% de error en FO recalculada . . . . .	116
4.6.1.	Error % Nuevas líneas construidas para 120 clusters . . . . .	122
4.6.2.	MAE para 120 clusters . . . . .	125
4.6.3.	Correlación para 120 clusters . . . . .	128

# Capítulo 1

## Introducción

### 1.1. Motivación

En la actualidad existe un aumento considerable en las inversiones relativas a energías renovables (**ER**) [3], debido tanto a razones ambientales como económicas, cuyo rápido desarrollo tecnológico ha permitido disminuir costos de manera significativa en su construcción e implementación. Es por ello que los Sistemas Eléctricos de Potencia (**SEP**) deben adaptarse a estas tecnologías de generación, incrementando los niveles de flexibilidad operativa del sistema, por lo cual se requiere de la integración de estas variables a la planificación eléctrica de los sistemas. De este modo, surge la necesidad de buscar nuevas metodologías que ayuden a simplificar sus cálculos, sin perder la precisión requerida en estos.

Por otra parte, las interconexiones de los SEP (integración energética entre sistemas) son cada vez más frecuentes, por lo cual el problema de la planificación de expansión de transmisión ha adquirido mayor dimensión y su implementación se requiere en sistemas cada vez más grandes.

En general, en los problemas de planificación de los SEP existen un gran número de variables dependientes e independientes que afectan su comportamiento, las cuales pueden ser indexadas de manera temporal, espacial, y por escenario, entre otras. Relativo al modelo del sistema, al tratarse de un estudio de mediano y largo plazo del SEP, este tipo de problemas poseen un gran número de variables de trabajo y probablemente de complejo manejo computacional, siendo su tratamiento dificultoso. Es por ello que el problema debe ser tratado de una forma simplificada, agregando variables para de este modo representar la operación de manera reducida, procurando perder la menor cantidad de información relevante en el proceso.

En general se pueden considerar tres categorías de simplificaciones comúnmente empleadas en problemas de planificación de expansión:

1. Omisión de ciertos fenómenos, tal como ignorar restricciones de pre-despacho o fenómenos dinámicos en formulación.
2. Representación simplificada de ciertos fenómenos, tales como el uso de flujo de carga en DC en vez de AC para linealizar ecuaciones.

3. Agregación de variables para reducir la dimensionalidad del problema de optimización. El cómo realizar esta reducción preservando las características principales del problema principal es el foco de esta propuesta.

Típicamente en problemas de planificación de la expansión las variables y parámetros se encuentran indexados de manera espacial (e.g. por barra, nodo o región), temporal (e.g. por hora, día, mes o año) y por escenario. Por ello, usualmente podemos destacar tres tipos de agregación de variables para reducir la dimensionalidad del problema y la disminución del número de variables de decisión, restricciones, y parámetros:

1. Espacial: Por ejemplo, representar varias barras relativamente cercanas por una única barra.
2. Temporal: Por ejemplo, representar un conjunto de horas con características similares de demanda por un bloque, o varios días por un día representativo.
3. Estocasticidad: Por ejemplo, representar varias posibles realizaciones de un fenómeno estocástico por un solo escenario.

Dados los antecedentes anteriormente mencionados, el estudio de variables independientes multi-dimensionales en los modelos de operación, tales como sus datos de entrada (demanda y generación renovable variable), serán el foco principal de esta tesis. Por ello se busca analizar el efecto que tiene el modelamiento y selección de estos datos de entrada en la planificación final del sistema, buscando establecer una relación entre el detalle de este proceso y la magnitud de los beneficios computacionales resultantes.

Por una parte, en el estudio de planificación de interconexión de sistemas eléctricos, una de las variables independientes principales por estudiar en esta tesis será la demanda, debido a que esta entrada del sistema considera aspectos espaciales y temporales muy relevantes en este tipo de problemas.

En general, en un sistema de potencia su operación económica se desarrolla buscando abastecer su demanda al mínimo costo, empleando para ello herramientas que de alguna forma predigan su comportamiento, siendo de este modo una variable de alta sensibilidad en el problema. Además, al considerar la interconexión de sistemas, se destaca el hecho que generalmente sistemas distintos poseen perfiles de consumo distintos, y en el caso de sistemas en distintas zonas horarias las demandas se encuentran, generalmente, desfasadas. El argumento anterior hace resaltar la importancia de considerar el factor intertemporal en estos casos de estudio.

Una simplificación comúnmente utilizada en estos estudios es la discretización de las variables de entrada en bloques. En el caso de la demanda, esta ha sido tratada históricamente en los problemas de planificación eléctrica de forma discreta, trabajándose en bloques a través de la Curva de Duración de la Demanda o Carga -*Load Duration Curve (LDC)*<sup>1</sup>-, cuyo objetivo es aproximar la curva de demanda de un sistema en  $L$

---

<sup>1</sup>La Curva de Duración de la Demanda ilustra la variación de la demanda de un sistema en orden descendiente, en lugar de su orden cronológico. Esta es utilizada generalmente para ilustrar la relación entre los requisitos de capacidad de generación y sus respectivos usos, dado un cierto incremento de carga.

bloques, donde cada bloque de demanda define un nivel y duración. La ventaja de esta metodología es que en vez de una representación detallada de 8760 horas al año, la demanda puede ser representada solo por  $L < 8760$ . Así, esta reducción en el tamaño del modelo permite expandir el horizonte de tiempo a más años (décadas) o aumentar el detalle de otros aspectos del modelo. Sin embargo, su desventaja radica en la pérdida de secuencialidad de los datos, y con ello información relevante relativa a temporalidad y causalidad del problema.

Otra variable de entrada del sistema con gran relevancia en este estudio es la generación de energías renovables variables (**ERV**)<sup>2</sup>, la cual se considera como ingresos de energía de disponibilidad limitada que posee variabilidad e incertidumbre, y cuyos perfiles de generación poseen alta correlación con su localización geográfica. Además, en conjunto con la integración de altos niveles de generación ERV en los sistemas de potencia, estos se han visto en la necesidad de generar incentivos económicos que busquen la inserción de recursos que permitan aumentar la flexibilidad del sistema<sup>3</sup>, y de este modo integrar nuevas restricciones operativas al problema de planificación<sup>4</sup>.

Es de considerar, además, que la influencia de estas variables sobre el sistema depende directamente de la escala del sistema que se estudia. En algunos casos se considera esta entrada en la *Demanda Neta*<sup>5</sup> del sistema, modelando su variabilidad directamente en la demanda, y simplificando de este modo espacialmente esta variable de entrada.

En resumen, el trabajo relativo a esta tesis se basará en el análisis de datos multivariados relativos a demanda y generación ERV y su aplicación en la resolución del problema de expansión de la transmisión. El trabajo se enfocará en proponer y aplicar metodologías para la representación reducida de estas variables, considerando formulaciones que simplifiquen sus relaciones de temporalidad y espacialidad, minimizando en el proceso la pérdida de información relevante para la resolución del problema. Relativo a la demanda y generación ERV, se contrastarán estrategias de representación reducida actualmente empleadas en la literatura, y se propondrá una nueva metodología basada en un análisis post-óptimo de las soluciones de un problema reducido con características similares a las del problema de expansión de transmisión eléctrica. En la Figura 1.1.1 se presenta un esquema que ilustra el trabajo propuesto.

---

<sup>2</sup>En particular, se considerarán como energías renovables variables la generación solar y eólica

<sup>3</sup>Entendida como la capacidad de todo sistema eléctrico para adaptarse a las crecientes fluctuaciones de la oferta y la demanda garantizando al mismo tiempo la fiabilidad del sistema

<sup>4</sup>Algunas restricciones operativas son el almacenamiento en base a baterías para Control Primario de Frecuencia (CPF), máquinas con alta tasa de toma de carga para Control Secundario de Frecuencia (CSF), condensadores sincrónicos, entre otras

<sup>5</sup>La demanda neta considera a la generación renovable como negativa y se la resta al perfil de demanda original

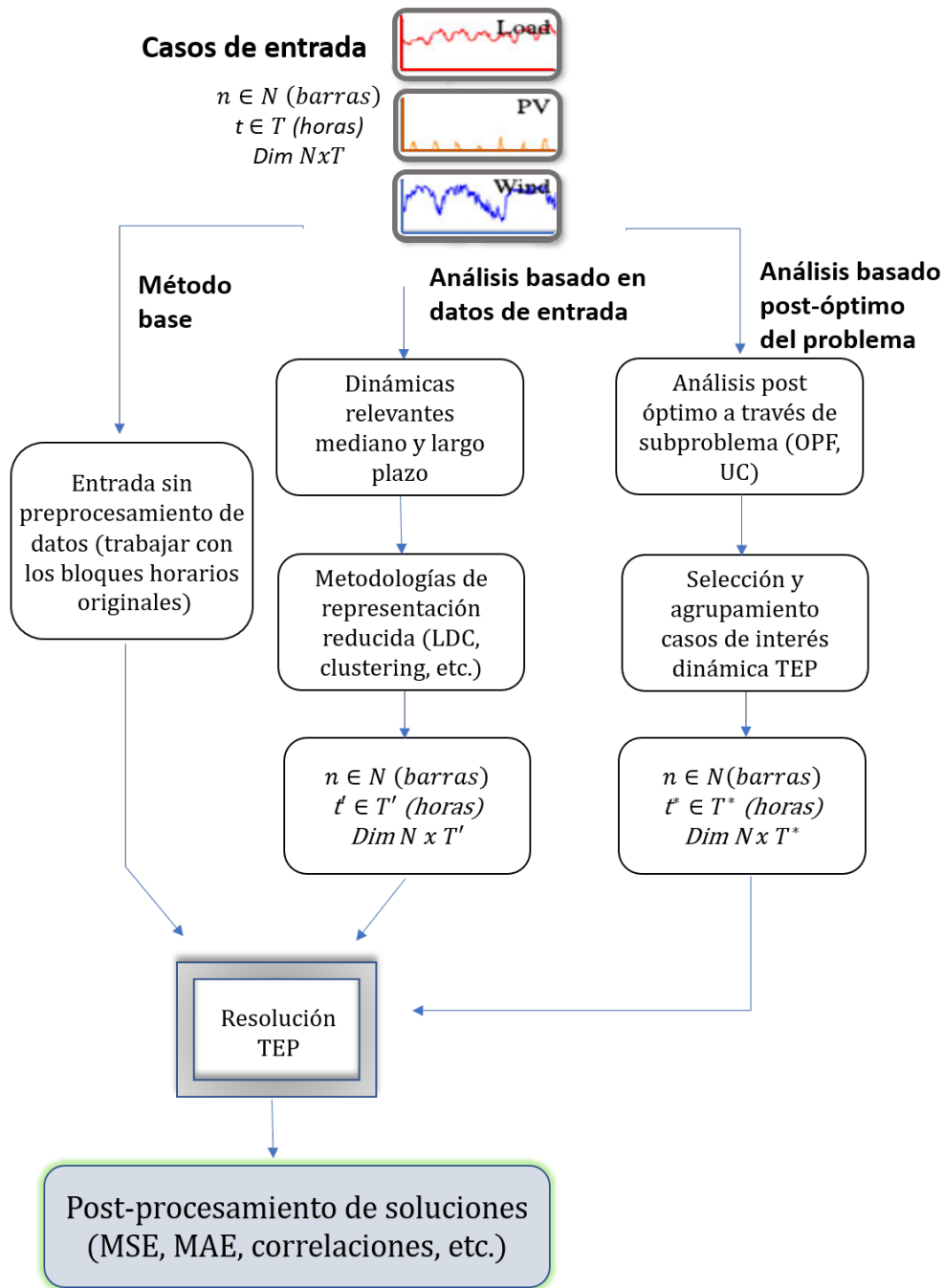


Figura 1.1.1: Esquema general del trabajo propuesto.

## 1.2. Hipótesis

Para el desarrollo de la presente tesis algunas de las principales hipótesis de trabajo fueron:

- El considerar las características del problema que se intenta resolver (en el presente caso expansión de la capacidad de transmisión), ayudaría a obtener representaciones reducidas de variables multi-dimensionales de demanda y generación renovable variable, que no afectarían demasiado la precisión de la solución final del problema.
- El problema completo (sin agregación de variables) no es resoluble en un tiempo razonable, por lo que se requieren representaciones reducidas de los datos de entrada y variables del problema agregadas.
- La aproximación empleada a través de los bloques de la curva de duración de la demanda no considera la temporalidad del problema; por lo cual afecta al espacio de solución, puesto que estas no considerarán necesariamente la temporalidad y causalidad<sup>6</sup> del problema. Esto incide principalmente sobre variables de decisión y fenómenos que requieren de este tipo de consideraciones, como lo es el caso del almacenamiento y restricciones de flexibilidad operacional del sistema.
- Respecto a la correlación espacial de la demanda en diferentes nodos del sistema:
  - Relativo a la distribución espacial de las variables, se esperan mejores resultados para casos de estudio que consideren la espacialidad del sistema, dada su alta relevancia dentro del problema de expansión de la transmisión (una descripción más detallada del sistema tendrá claramente una mejor representatividad de los efectos espaciales que una aproximación simplificada).
  - Para un escenario de estudio con alta correlación espacial entre variables, se espera que el método LDC sea lo suficientemente bueno, ya que la variable espacial pierde importancia en este caso.  
Para escenarios de estudio con baja correlación espacial la calidad del cluster dependerá de la forma en que se considere esta. Se propone un análisis post-óptimo que realice *clustering* según costos de operación y otro según flujos por las líneas, a modo tal de determinar que tan importante es la información que entregan estos respecto a la distribución espacial de la demanda.
- Respecto a las técnicas de *clustering* propuestas para el análisis post-óptimo:
  - Un agrupamiento basado en costos de operación considerando un OPF con límites de capacidad de líneas será mucho más representativo que el caso sin límites, dado que podrá considerar el agrupamiento según los altos costos de la energía no servida.

---

<sup>6</sup>Entiéndase causalidad como la relación de causa y efecto entre dos eventos.

- Un agrupamiento basado en flujos por las líneas considerando un OPF sin límites de capacidad de líneas sea mucho más representativo que el caso con límites, dado que el considerar agrupamiento de líneas con niveles muy superiores de flujo, puede entregar una buena señal de agrupamiento de demanda para el problema de expansión de la transmisión.
- El método de *clustering* basado en métricas de congestión (o método alternativo), que procesa de forma conjunta los datos de flujos por las líneas de un OPF con y sin límites de capacidad, debería poder rescatar de mejor forma la información espacial, debido a que posee información relevante del problema de expansión de la transmisión.

### 1.3. Objetivos

#### 1.3.1. Objetivo Principal

El objetivo principal de esta tesis es proponer y aplicar metodologías para la representación reducida de demanda y generación renovable variable de un sistema de potencia, considerando formulaciones que simplifiquen sus relaciones de temporalidad y espacialidad, para la aplicación de estas como entradas del problema de planificación de la transmisión eléctrica, minimizando en el proceso la pérdida de información relevante a este.

#### 1.3.2. Objetivos Específicos

- Diseñar estrategias para obtener representaciones reducidas de variables multi-dimensionales de demanda y generación renovable variable, en base a la revisión del estado del arte y a las necesidades relativas a su aplicación al problema de expansión de la transmisión.
- Proponer una nueva metodología para obtener representaciones reducidas de variables multi-dimensionales de la demanda y generación renovable variable basadas en el análisis post-óptimo de las soluciones de un problema reducido con características similares a las del problema de expansión de transmisión eléctrica.
- Desarrollar algoritmos para la implementación de las metodologías de reducción de variables multidimensionales de demanda y de generación renovable variable.
- Evaluar el desempeño de estrategias diseñadas para la generación de escenarios de demanda y generación renovable variable a través de los algoritmos propuestos e implementar en el problema de expansión de la transmisión.

### 1.4. Metodología

La metodología de trabajo utilizada se divide en 6 etapas, las cuales se describen a continuación:

1. Revisión bibliográfica y estado del arte: se realiza una revisión de los trabajos y publicaciones realizadas hasta la fecha respecto al problema planteado. Respecto a estas se plantean los alcances y objetivos generales del presente trabajo.
2. Selección de escenarios de estudio: dado que uno de los principales aportes de este estudio es la implementación de la reducción de escenarios en diferentes casos de demanda, se requiere de la elección correcta de un conjunto de estos para poder contrarrestar de mejor forma los resultados obtenidos. Se realizan análisis de correlaciones espaciales de los diferentes escenarios de estudio, para diferentes niveles de generación renovable y diferentes localizaciones de estas.
3. Implementación de algoritmos de *clustering*: para la reducción de escenarios propuesta se modifican e implementan los respectivos algoritmos de agrupación de k-means, LDC y PAM<sup>7</sup>, los cuales reducen la dimensión temporal de los datos de entrada de demanda ingresados en vectores multidimensionales. Estos algoritmos son trabajos en R y sus librerías asociadas.

Para el desarrollo de la segunda parte del estudio, la cual consiste en una reducción de escenarios basada en un análisis post-óptimo, se realiza una modificación de los algoritmos para implementar *clusters* según entradas obtenidas de las salidas de un subproblema (OPF), y que agrupe sus respectivos índices horarios, para de esta forma obtener los respectivos *clusters* de demanda. Las metodologías de agrupamiento de datos desarrolladas en esta etapa son:

- **Clustering basado en COp (con límites de flujos)**: Se considera el vector de costos de operación de OPF realizado con restricciones de capacidad de líneas.
- **Clustering basado en COp (sin límites de flujos)**: Se consideran el vector de costos de operación de OPF realizado sin restricciones de capacidad de líneas.
- **Clustering basado en Flujos (con límites de flujos)**: Se considera el conjunto de vectores de flujos por las líneas del OPF realizado con restricciones de capacidad de líneas.
- **Clustering basado en Flujos (sin límites de flujos)**: Se considera el conjunto de vectores de flujos por las líneas del OPF realizado sin restricciones de capacidad de líneas.
- **Clustering basado en Métricas de Congestión (o método alternativo)**: Se considera un conjunto de vectores de flujo “alternativos” ( $fl_{alt}$ ), el cual consiste en un agrupamiento respecto a flujos con ( $fl_{con}$ ) y sin ( $fl_{sin}$ ) límites de capacidad en las líneas y un tratamiento posterior de estos.

En los próximos capítulos se explicará con mayor detalle esta metodología y su aplicación.

---

<sup>7</sup>El trabajo realizado es la continuación del trabajo iniciado por Ignacio Aravena en [1], en donde se proponen las técnicas respectivas (k-means, PAM y LDC) para casos más pequeños. El presente trabajo propone algunas modificaciones a estos algoritmos y una implementación en sistemas de mayor escala.

4. Implementación de modelo de TEP: se crean algoritmos del problema OPF y del TEP en *PYOMO*<sup>8</sup>.
5. Selección de métricas de evaluación de resultados: se analizan las diferentes salidas obtenidas de la resolución del problema de expansión de transmisión y se procesan para obtener los datos relevantes para determinar cualitativa y cuantitativamente qué tan buena es la metodología implementada.
6. Análisis de resultados: una vez obtenidas las métricas de evaluación de resultados se realiza un contraste generalizado de lo obtenido y se concluye respecto a los aportes reales del trabajo.

## 1.5. Alcances y estructura del documento

El alcance de este trabajo es proponer e implementar metodologías de reducción de escenarios de demanda y generación renovable, cuya principal aplicación será en el problema de planificación de la expansión de la transmisión de sistemas de potencia. De este modo, se busca reducir la cantidad de variables necesarias para la representación del problema, reduciendo tiempos de simulación, pero procurando además una buena calidad de la solución.

El desarrollo de este trabajo ha requerido del manejo de modelos de optimización y técnicas de manejo de datos. Sin embargo su principal aporte es relativo a las metodologías de reducción de escenarios de demanda, quedando en un segundo plano la complejidad del modelo de planificación de la transmisión utilizado y los algoritmos de *clustering* implementados, a fin de centrar la atención en las aplicaciones de ingeniería eléctrica que los distintos problemas tienen.

La presente tesis se organiza en capítulos autocontenidos para facilitar la lectura del documento. Cada capítulo contiene de forma independiente su propia introducción, estado del arte, metodología y análisis, para de este modo separar los dos casos de estudio de manera independiente. Los capítulos del cuerpo de esta tesis se organizan como sigue:

- Capítulo 1: Introducción.  
Se presenta la introducción al problema estudiado, con los principales alcances, objetivos, metodología de trabajo y la presentación de la estructura del documento.
- Capítulo 2: Marco Teórico y Estado del Arte.  
Se desarrolla de manera general la teoría, con descripciones introductorias a los diferentes tópicos relativos a este trabajo y metodologías de trabajo utilizadas en los siguientes capítulos. De manera conjunta se desarrollará un estado del arte generalizado de los diferentes tópicos de interés.
- Capítulo 3: Reducción de Escenarios Considerando Diversidad Espacial de Generación Renovable.

---

<sup>8</sup>Pyomo es una colección de paquetes de software de Python para formular modelos de optimización.

Se presenta el primer trabajo relativo al empleo de técnicas de *clustering* directamente en los datos de entrada del problema de expansión, en donde se contrasta la metodología de la curva de duración de la demanda con la reducción de escenarios a través de k-means y PAM. Se presenta el estado del arte del problema, los casos de estudio desarrollados, la metodología empleada y los resultados generales. En la sección anexos se complementan estos resultados con otras gráficas de interés y tablas relativas a la simulaciones realizadas.

- Capítulo 4: Reducción de Escenarios en Base a Condiciones Operacionales de un Análisis Post-óptimo.

Se plantea un método de reducción de escenarios en base al procesamiento de datos relevante en el problema, obtenidos a través de un subproblema o problema reducido con características similares a las del problema de expansión de transmisión eléctrica. Luego se define un conjunto de resultados de interés y realizar la reducción de casos identificando las características relevantes que se buscan preservar en la representación reducida. Se propone utilizar como modelo sencillo un OPF (*Optimal Load Flow*), de modo tal que a través de este se pueda obtener de manera rápida los datos de realimentación para la selección de escenarios adecuados a simular en el problema de expansión de la transmisión real. La presentación del capítulo se desarrolla de manera similar al capítulo 3, presentando su propio estado del arte, metodología y resultados.

- Capítulo 5: Conclusiones.



## Capítulo 2

# Marco Teórico y Estado del Arte

**RESUMEN:** Se presenta el marco teórico y el estado del arte de los dos temas principales abarcados en el trabajo. Inicialmente se describe lo relativo al problema de planificación y operación de sistemas de potencia, en donde se describen las principales consideraciones relativas a estos temas y las respectivas formulaciones consideradas durante el trabajo. En la segunda parte del capítulo se describen técnicas de reducción de escenarios utilizadas en este tipo de problemas y la descripción generalizada de las técnicas escogidas.

### 2.1. Planificación y Operación de Sistemas de Potencia

La planificación del diseño y operación de un SEP requiere de un continuo análisis y estudio, con el objetivo de evaluar su operación y establecer la efectividad de su diseño o los planes alternativos, como por ejemplo, si la necesidad del sistema es algún tipo de expansión futura. Este tipo de decisiones requiere de estudios apropiadamente concebidos y conducidos con el objetivo de que sean cada vez más económicas, seguras, confiables y ambientalmente sostenibles. Aquí es donde surge con gran relevancia la figura del operador del sistema, cuyo deber es garantizar el cumplimiento de estos objetivos y llevar esto acorde al sistema regulatorio que lo rige.

Los modelos regulatorios bajo los que pueden funcionar los sistemas de energía eléctrica son básicamente dos: Modelo Centralizado, Regulado o *Pool*, y Modelo Liberalizado o Desregulado [4].

En el Modelo de Operación Centralizado, la explotación del sistema la lleva a cabo un operador central con el objetivo de minimizar los costos totales de generación, con el objetivo de reducir el precio de la energía y al mismo tiempo asegurar el cumplimiento de estándares de calidad de servicio. El Modelo Liberalizado en cambio, opera bajo las leyes de libre mercado, desapareciendo la figura de un operador central y aparecen otros agentes, entre los que se destacan compañías generadoras, comercializadores, compañías distribuidoras, y un operador independiente que coordina las transacciones de energía entre los agentes anteriores. En este modelo las compañías generadoras son libres de par-

participar en el mercado, no tienen obligación de servicio y su único objetivo es maximizar sus ingresos por venta de energía.

Así, con los nuevos paradigmas planteados por el Modelo Liberalizado, los precios de mercado son definidos por la oferta y la demanda. El operador independiente tendrá como objetivo lograr la coordinación del sistema igualando la producción de energía eléctrica de los generadores con la demandada por los consumidores, y que esta transacción de energía se lleve a cabo cumpliendo determinados criterios, como, por ejemplo, maximizar el beneficio social neto. En particular, en el caso chileno se tiene un mercado desregulado relativo a la generación y uno centralizado relativo a las decisiones de transmisión, lo cual se genera una dificultad a la hora de encontrar una decisión óptima en el plan de inversiones final del sistema.

Por otra parte, años atrás los problemas básicos tradicionalmente enfocados en los sistemas eléctricos de potencia eran: cortocircuitos, flujos de potencia y estabilidad, los cuales en el tiempo fueron resueltos mediante modelamientos matemáticos y físicos, y llevados a su implementación computacional. Hoy en día, los nuevos problemas enfocados en el control y operación de los sistemas eléctricos de potencia son los relacionados con la operación económica y expansión óptima de los sistemas; por lo cual su área de estudio se encuentra en constante expansión.

### 2.1.1. Operación Económica

La **Operación Económica (OE)** del sistema de potencia tiene como uno de sus principales objetivos el minimizar los costos de operación del sistema, sujeto a las restricciones de red y operación. Para ello se recurre al desarrollo de modelos matemáticos del sistema eléctrico, tales como los modelos de las centrales eléctricas -térmicas e hidráulicas-, de la red y la demanda.

El problema de la operación económica se basa en el conjunto de características de entrada - salida de las unidades de generación: térmica convencional - vapor, fuel, carbón, diesel y gas -, nuclear, hidráulica y actualmente se consideran con gran importancia las energías renovables. Este problema, conocido también con el nombre del **Despacho Económico -Economic Dispatch (ED)-**, busca un nivel de generación para cada uno de los generadores disponibles; de manera que el costo total de operación sea mínimo para satisfacer a toda la carga y las pérdidas. Las pérdidas de transmisión son importantes; luego, éstas se deberán incluir en el modelo del sistema según el nivel de exactitud de cálculos buscado. El propósito de este modelo, es encontrar una política óptima de operación para estas unidades.

Por otra parte, el problema de entrada en servicio de una unidad u otra central, es más complejo y se denomina **Pre despacho de Carga -Unit Commitment (UC)-**. Este problema posee mayor complejidad matemática, debido a que se involucran variables enteras binarias (1 - 0, encendido-apagado); lo que lo hace un problema de programación lineal entera mixto.

Puesto que, un generador en particular tiene que estar conectado o desconectado de la red, pero no todos los generadores estarán conectados al sistema; sino de acuerdo a los requerimientos operacionales, satisfacer la demanda y disponer de suficiente reserva en

giro para enfrentar una contingencia; sin embargo la cuestión es, ¿cuál será la política óptima? Para resolver el problema, existen tres técnicas ampliamente empleadas, las mismas son: la lista de prioridad, la programación entera y la programación dinámica.

Este tipo de problemas por lo general son trabajados sobre el supuesto de estado estacionario en sus variables eléctricas y a través de estas hipótesis es posible modelar restricciones técnicas de los SEP, restricciones de potencias activas y reactivas, dispositivos de control de flujo, rampas de carga en generadores [5, 6] y líneas, restricciones de encendido y apagado de unidades generadoras (tiempos mínimos y máximos), tiempos de estabilización, restricciones de seguridad (dando paso al problema de *Security Constrained Unit Commitment* (SCUC)), salidas forzadas o fallas de componentes, incertidumbres y variabilidad en generación intermitente, entre otras. Otro tipo de consideraciones en el problema son relativas a los costos de operación, costos combustibles, costos de personal, modelaciones de costos lineales [5] y convexas [6], desgastes de equipos, etc. Los mayores desafíos de modelamiento técnicos actuales que enfrenta esta área corresponden a la inclusión adecuada de nuevas tecnologías como los son *Flexible AC Transmission System* (FACTS) [7], *Resilient AC Distribution Systems* (RACDS) [8], el Respuesta en Demanda -*demand response* (DR)- [9], [10], la inclusión de generaciones ERV y el desarrollo de las Redes Inteligentes -*Smart Grids*-[9, 10].

Relativo a los métodos de resolución, los avances computacionales han llevado a resolver los problemas bases planteados, en los cuales se combinan los métodos de optimización convencionales con técnicas de aplicación de medidas y procedimiento de carácter técnico operacional, basados principalmente en experiencia y conocimiento del sistema. De este modo, el problema puede ser planteado de forma determinista, cuyos enfoques han evolucionado desde la *Dynamic Programming* (DP), combinaciones Heurísticas, combinaciones con técnicas de *Lagrangian Relaxation* (LR), *Integer Programming* (IP), *Mixed-Integer Programming* (MIP), *Mixed-Integer Linear Programming* (MILP), *Quadratic Programming* (QP), *Mixed-Integer Quadratic Programming* (MIQP) [11], combinaciones de métodos MILP con LR [12], entre otros. Actualmente, dentro de los últimos avances en la modelación del UC pueden nombrarse las relajaciones cónicas y semidefinida positiva (SDP) presentes en los trabajos [13, 14].

Es de interés, además, el enfoque dado a la incertidumbre del problema en donde surge el estudio y desarrollo de la Programación Estocástica - *Stochastic Programming* (SP)- [15], la cual consiste en la generación de escenarios donde el valor de los parámetros sujetos a incertidumbre está fijado. Dado que la incertidumbre se encuentra discretizada, estas técnicas resultan en resolver una gran cantidad de modelos deterministas, los cuales ya poseen una escala considerable. Otro tipo de técnicas como lo es la Optimización Robusta - *Robust Optimization* (RO)- [16]. El concepto detrás de este campo de optimización radica en construir modelos que sean prácticamente inmunes al conjunto de datos inciertos presentes en el sistema, y a diferencia de los modelos de programación estocástica, no precisan de distribuciones de probabilidad; sino que utilizan la noción de conjuntos de incertidumbre, los cuales poseen todos los eventos de los cuales se quiere que la solución del modelo se proteja.

El **Flujo de Potencia** -*Power Flow*-, se refiere al problema de resolución de las ecuaciones de la red, con el objeto de conocer todas las variables eléctricas para de-

terminar el estado de operación del sistema. A partir de este problema surge un caso particular llamado **Flujo Óptimo de Potencia** -*Optimal Power Flow* (OPF)-, el cual se refiere a un estado de operación en las que algunas variables son optimizadas, sujeto a limitaciones sobre las variables y funciones. Mayor desarrollo del estado de arte del problema en [17, 18].

### 2.1.1.1. Flujo Óptimo de Potencia: OPF

Como se mencionó anteriormente, el Flujo Óptimo de Potencia busca optimizar algunas de las variables de su formulación, y dado que se trabaja con un modelo de mercado desregulado, lo que se buscará en su formulación es minimizar los costos de operación del sistema.

En el presente trabajo se utilizará un modelo simplificado del OPF, en el cual se integrará la infraestructura eléctrica del sistema a través de sus líneas y respectivos límites de flujos, en combinación con la operación de los diferentes generadores de la red. A continuación se presenta la formulación utilizada para el OPF, la cual será implementada en la formulación del problema de expansión de la transmisión y además como problema base para la formulación post-óptima explicada en mayor detalle en el capítulo 4.

*Formulación OPF:*

$$\min_p \sum_{t \in T} \alpha_t S_{Base} \left( \sum_{j \in G} c_{MW}^j \sum_{t \in T} p_{j,t} + VoLL \sum_{k \in B} \sum_{t \in T} USE_{k,t} \right) \quad (2.1.1)$$

$$\text{s.t.} \quad \sum_{j \in G(k)} p_{j,t} - \sum_{km \in L(k, \cdot)} \sum_{y \in Y} f_{km,t,y} + \sum_{mk \in L(\cdot, k)} \sum_{y \in Y} f_{mk,t,y} = \frac{d_{k,t}}{S_{Base}} + p_{vert\ ERV\ k,t} - USE_{k,t} \quad \forall k \in B, \forall t \in T \quad (2.1.2)$$

$$-\bar{\delta} \leq \delta_{k,t} \leq \bar{\delta} \quad \forall k \in B, \forall t \in T \quad (2.1.3)$$

$$\delta_1 = 1 \quad (2.1.4)$$

$$f_{km,t} = -b_{km} n_{km} (\delta_{k,t} - \delta_{m,t}) \quad \forall km \in L, \forall t \in T \quad (2.1.5)$$

$$-n_{km} \bar{f}_{km} \leq f_{km,t} \leq n_{km} \bar{f}_{km} \quad \forall km \in L, \forall t \in T \quad (2.1.6)$$

$$\underline{p}_j \leq p_{j,t} \leq \bar{p}_j \quad \forall j \in G, t \in T \quad (2.1.7)$$

$$USE_{k,t} \leq \frac{d_{k,t}}{S_{Base}} \quad \forall k \in B, t \in T \quad (2.1.8)$$

$$p_{vert\ ERV\ k,t} \leq -\frac{d_{k,t}}{S_{Base}} \quad \forall k \in B, t \in T, \quad \text{ssi} \quad \frac{d_{k,t}}{S_{Base}} \leq 0 \quad (2.1.9)$$

La formulación utilizada en este trabajo considera en (2.1.1) su función objetivo, en donde se busca abastecer la demanda  $d_{k,j}$  al mínimo costo de operación. En esta ecuación se presentan los bloques de demanda a través del factor  $h_t$ , el cual representa el ancho del intervalo temporal que ocupa ese bloque de demanda en el total de las horas utilizadas en el problema.

En (2.1.2) se presentan las ecuaciones del balance nodal, en la cual se consideran los flujos bidireccionales en las líneas para todos los periodos de tiempo y en (2.1.7), (2.1.8), (2.1.9) se presentan los diferentes restricciones empleadas en esta.

Dado que la formulación del OPF implementada es una simplificación en DC, los flujos por las líneas vienen dados por (2.1.5), (2.1.6) y las respectivas restricciones nodales se presentan en (2.1.3), (2.1.4).

### 2.1.2. Planificación de Sistemas de Potencia

El crecimiento de la demanda y requisitos de calidad impuestos al suministro eléctrico obligan a las empresas del sector a realizar la ampliación de su parque generador y expandir sus sistemas de transporte y distribución, lo cual representa altos costos de inversión. La planificación de expansión de sistemas eléctricos consiste en determinar el número de centrales generadoras, líneas de transporte y redes de distribución que el sistema requiere construir en un futuro, mediante modelos de planificación que determinan políticas óptimas de inversión para satisfacer la demanda esperada.

Dado el gran número de variables y la diversidad de alternativas de solución a este tipo problemas, el operador se ve obligado a buscar modelos de planificación y formas de resolución adecuadas a cada caso. Las características de los modelos de planificación (lineales, no-lineales) dependen; por un lado, del tipo de sistema eléctrico que se esté intentando modelar, y por el otro, de la representación que se haga a cada uno de los elementos que lo conforman: sistemas de generación, transporte y distribución de potencia. Las metodologías de planificación aplicables a ampliar generación, sistema de transmisión y distribución son diferentes básicamente por las restricciones de operación y red propias de cada sistema.

Algunas de las variables presentes en los problemas de planificación de sistemas eléctricos tienen la particularidad de interrelacionarse entre sí, siendo, en este sentido, dependientes. Así, por ejemplo, existen variables del tipo temporal, espacial, probabilístico, entre otras, cuyo manejo debe considerar el efecto de simultaneidad entre ellas. Relativo al modelo del sistema, al tratarse de un estudio de mediano y largo plazo del SEP, el número de variables a trabajar es alto, de difícil manejo matemático y con alto requerimiento computacional, por lo que se debe tratar el problema de una forma simplificada y de este modo representar la operación de manera reducida. Entre las simplificaciones utilizadas generalmente se encuentran los Flujos de Potencia en continua y la solución no cronológica (desacoplada) del Predespacho.

El problema de la Planificación de la Expansión de Capacidad del Sistema - *Capacity Expansion Planning* (CEP)- relaciona múltiples dimensiones del problema, entre las cua-

les se tienen aspectos económicos (representación en el mercado), matemáticos y físicos (manejo de incertidumbres, tipos de variables de decisión, formulación del problema), manejo de datos (datos de entrada y salida del problema), como se observa en el esquema conceptual presentado en la Figura 2.1.1.



Figura 2.1.1: Mapa conceptual variables de estudio problema de Planificación Expansión de la Capacidad del Sistema. [Fuente: Elaboración propia]

El CEP puede dividirse en diferentes subproblemas, los cuales pueden ser tratados de manera conjunta o separada. Estos son: Modelos de Optimización de Sistemas de Energía -*Energy Systems Optimization Models* (ESOMs) [19]-, Modelos de Planificación de Expansión de la Generación -*Generation Expansion Planning Models* (GEPs)- o Modelos de Planificación de la Expansión de la Transmisión -*Transmission Expansion Planning Models* (TEPMs)-. Ejemplos de casos en que se realiza una co-optimización de planificación de generación y transmisión en [20, 21, 22].

En el CEP los costos de operación son modelados a través del OPF, cuya estructura cambia según las decisiones de inversión en generación y/o transmisión. Inversiones en generación modifican la curva de abastecimiento de la demanda en el OPF (modificación local OPF). Por otra parte, inversiones en transmisión modifican la matriz de admitancias del sistema (modificación global del OPF) [23]. De este modo, este tipo de problemas según su formulación y resolución tiene como objetivo generar señales económicas de inversión para los sistemas de potencia. Así pues, su desarrollo es creciente y posee aspectos de interés por desarrollar tales como: resolución tiempo y espacio, balance de incertidumbre y transparencia, manejo de la creciente complejidad del sistema de energía e integración del comportamiento humano y los riesgos y beneficios sociales [19]. Otras bibliografías de interés relativas a este tipo de problemas y su desarrollo en el mercado son [24, 25].

Otras variables de creciente interés en el estudio de este tipo de problemas son la integración de energías renovables [26, 27], construcción de almacenamiento [28, 29, 30, 31, 32] y respuesta en demanda [33].

El aumento de las capacidades computacionales han ayudado al desarrollo de este tipo de problemas, con lo cual su implementación se ha podido desarrollar en sistemas cada vez mayores, incluyendo variables de incertidumbre, co-optimización de problemas e incluso la solución cronológica del problema [34]. Sin embargo, la gran cantidad de variables involucradas en el problema sigue siendo un limitante, y se debe buscar un equilibrio entre requerimientos computacionales, tiempos de simulación y resultados económicos del problema.

#### 2.1.2.1. Planificación de la Expansión de la Transmisión (TEP)

Una adecuada planificación y expansión de la transmisión de un sistema eléctrico en un mercado competitivo debe permitir su desarrollo a un mínimo costo, y proveer de señales que incentiven la eficiencia económica y técnica de éste, adaptando los requerimientos de generadores y consumidores y remunerando adecuadamente a sus propietarios [35]. El TEP busca encontrar respuestas al dónde, cuándo y cómo nuevas líneas deberían ser instaladas en el sistemas para ser capaz de abastecer su demanda, sujeto a restricciones físicas y económicas [24]. De este modo, en su planteamiento matemático el TEP es un problema de programación no lineal entero mixto con variables de decisión binarias relativas a la construcción de nuevas líneas o su expansión.

En [36] se presenta una visión generalizada del estado de arte del problema, y se señalan bibliografías de referencias para cada una de las siguientes categorías de análisis del problema. Además, a continuación se señalan nuevas bibliografías de interés actualizadas por tema.

1. **Modelamiento:** Considerar ventajas y desventajas de los modelos en DC o AC del sistema. Desarrollos actuales intentan integrar la planificación en AC, un ejemplo se tiene en [37], en donde se desarrolla un TEP con un flujo AC linealizado. Otras consideraciones son el tipo de líneas a construir y su modelamiento. En [38] se presenta una formulación del problema que contiene planificación de líneas HVAC/HVDC, restricciones de seguridad y sus respectivas pérdidas.

2. **Métodos de resolución:** En la planificación tradicional, el objetivo del TEP es minimizar el costo de inversión; sin embargo, en la planificación más avanzada, el planificador debería buscar diferentes objetivos tales como:
  - a) Facilitar la competencia entre los diferentes participantes del mercado.
  - b) Proveer de un desarrollo no discriminatorio y competitivo.
  - c) Mitigar las congestiones de transmisión.
  - d) Minimizar costos de inversión.
  - e) Minimizar el riesgo.
  - f) Confiabilidad y seguridad.
  - g) Generación distribuida.
  - h) Minimizar el impacto ambiental.
3. **Confiabilidad:** Los indicadores comúnmente utilizados en el problema para poder dimensionar su confiabilidad son:
  - a) *Loss of load expectation* (LOLE)
  - b) *Hierarchical realibility assessment*
  - c) *Loss of load cost* (LOLC)
  - d) *Energy expected not supplied* (EENS)
  - e) *Reliability improvement index*
4. **Congestión de líneas:** La congestión de transmisión es un tópico de relevancia que incorpora restricciones en la operación de la red, y por tanto se deberán considerar costos asociados a esta. Además, se deberá considerar la capacidad de transmisión disponible, transmisiones sobre la capacidad del sistema y el manejo de la carga en las líneas.
5. **Planificación de reactivos:** En los sistemas de potencia reales, las cargas reactivas se suministran a través de generadores. En este caso, la potencia reactiva se transfiere a través de líneas de transmisión y la transferencia de tal cantidad de potencia reactiva puede reducir la capacidad de transferencia disponible y esto puede llevar a construir más nuevas líneas de transmisión. Sin embargo, al asignar fuentes localmente reactivas cerca de los centros de carga, es posible suministrar demandas reactivas y, por lo tanto, se aumenta la capacidad de las líneas de transmisión y también se reducen las pérdidas de potencia. Por consiguiente, el problema de TEP debe asociarse con la planificación de la energía reactiva o, de lo contrario, el TEP conduce a construir más líneas nuevas.
6. **Visión de mercado eléctrico:** Los resultados de la planificación eléctrica deben ser acordes a los requerimientos del sistema y además deben brindar las señales económicas adecuadas para los inversionistas y participantes del mercado.

7. **Incertidumbre:** Relativo al manejo de incertidumbre del sistema, estas se pueden subdividir en las siguientes:
- a) Incertidumbre en predicciones de demanda y precios.
  - b) Disponibilidad de las componentes del sistema.
  - c) Incertidumbre generada por las simplificaciones en el modelo y simulación.
  - d) Incertidumbres relativas al mercado, energía disponible y riesgo, disponibilidad de combustibles y sus costos.
  - e) Nuevas tecnologías de generación como el viento y las unidades fotovoltaicas.
  - f) Políticas gubernamentales y públicas y aspectos medioambientales y sus costos.

Respecto al manejo probabilístico de este tipo de variable se tienen los trabajos [39, 40] y un ejemplo de aplicación directa en el TEP en [41].

8. **Consideraciones de generación renovable distribuida:** La generación distribuida es un foco en desarrollo que entrega al sistema variaciones en sus perfiles de demanda, y que afecta además a la relación espacial entre variables. Es un tema en creciente desarrollo cuyo estudio es cada vez más relevante.
9. **Horizontes de tiempo:** Desde la perspectiva del horizonte temporal, el problema TEP se clasifica como planificación estática y dinámica. En la planificación estática, el horizonte temporal no se considera y el plan óptimo se determina para un solo año. En otras palabras, se supone que todas las líneas nuevas deben instalarse en el primer año del horizonte de planificación. Pero, en la planificación dinámica, los años de horizonte se estudian por separado y se indican nuevas líneas para cada año. De hecho, se supone que cada línea nueva debe instalarse en su año relativo del horizonte de planificación. Parece que la planificación dinámica conduce a una planificación mejor y más barata, pero es muy compleja, con gran número de variables y largos tiempos de resolución.
10. **Impacto ambiental:** Hoy en día, las emisiones ambientales como el carbono, el óxido de nitrógeno y los gases naturales están limitados en función de la regulación de las Agencias de Protección Ambiental (EPA). Se alienta a los planificadores a instalar energías renovables en lugar de energías convencionales como el petróleo.
11. **Flexible AC Transmission System (FACTS):** Los dispositivos FACTS ayudan a incrementar la capacidad de transmisión del sistema, por ello su integración puede ser utilizada en lugar de instalar nuevas líneas.
12. **Co-optimización TEP y GEP:** es un área de creciente desarrollo, en donde se realizan ambas planificaciones de manera conjunta, de modo tal de buscar el plan óptimo para el sistema.
13. **Integración de energía eólica en el problema:** la integración de la energía eólica en el sistema es un factor que entrega incertidumbre al sistema. Se estudian

muchas metodologías de implementación de este tipo de variables, dado su efecto en los perfiles de demanda neta del sistema.

14. ***Security constrained TEP problem:*** La planificación de la expansión de la transmisión con restricciones de seguridad ha sido ampliamente investigada en los últimos años. Se han considerado diferentes restricciones de seguridad en problemas TEP tales como margen de estabilidad de voltaje, flujos de línea y limitaciones de generaciones, límites de tensión en estado estable y criterio de seguridad [N-1]. A modo de ejemplo, en [42] se presenta un TEP considerando restricciones de seguridad de transmisión. En [43] se plantea el como modelar estas restricciones de transmisión desde la formulación del OPF.

En [44] se presenta una visión generalizada del estado de arte actual del problema, y se desarrolla una metodología para implementación en un sistema de escala real y en [45] se expone una revisión de los modelos de planificación de la transmisión, clasificándolos relativo a sus métodos de resolución, horizontes de planificación y el prospecto de modelado.

El objetivo principal de esta tesis es el análisis desde el punto de vista de los datos de entrada del TEP, por lo cual el enfoque principal será en la variable de entrada demanda del sistema y en la generación de energías renovables variables (ERV). Estas serán estudiadas como las variables multidimensionales el sistema a trabajar. Relativo a la demanda, se estudiará su entrada de forma no-cronológica (a través de técnicas de agrupamientos de datos), buscando orientar este trabajo para sentar las bases de futuros estudios en que se considere la cronología, como lo es por ejemplo a través de la metodología de días típicos.

Las bases de este estudio se dan en los trabajos [1] y [46], los cuales se espera complementar con las metodologías de análisis actuales y proponer una metodología innovadora.

#### 2.1.2.2. Planificación estática versus dinámica

En general la planificación del sistema, es un problema de optimización no lineal entero-mixto, el cual considera minimizar el costo de inversión de la construcción de nuevas líneas considerando una demanda proyectada en el horizonte de planificación.

La planificación estática del modelo de optimización considera que el plan de expansión se reduce a un período específico, el cual disminuye la complejidad de su resolución. De este modo, para la planificación estática se realiza la planificación de un año entero, en donde en el tiempo  $t_0$  se realiza inversión en la transmisión para el resto del año, para posteriormente llevar a valor presente los respectivos costos de operación del sistema, y de este modo, proyectar la inversión a tiempo futuro.

El objetivo de la planificación dinámica o multi-etapa de *generación* es minimizar el costo de inversión y operación de las distintas centrales eléctricas, esto es, dónde, cuándo y cuánto tipo de generación instalar (tecnología y capacidad) en el sistema eléctrico para el horizonte de planificación establecido.

En la presente tesis se trabajará con planificación estática, dado que se busca simplificar lo máximo posible el problema de expansión de la transmisión, tomando en cuenta que el principal aporte de este trabajo es relativo a la reducción de escenarios.

### 2.1.2.3. Formulación del problema de planificación de la transmisión

Los modelos de TEP se han desarrollado por lo general en manera conjunta a la planificación de largo plazo, sin embargo, presentan algunas diferencias importantes con respecto a otros modelos: (i) Los modelos TEP son más difíciles de resolver que los modelos de expansión de generación y (ii) los costos de inversión en transmisión son solo una fracción de la inversión en generación, por lo que generalmente se asume que las inversiones en transmisión seguirán a la inversión de generación en sistemas reales. Sin embargo, este no es el caso o la regla común en la práctica, y la congestión, que resulta de decisiones de expansión de transmisión subóptimas, puede socavar significativamente el funcionamiento económico del sistema.

Respecto al modelo de planificación, la formulación de este tipo de problemas es de naturaleza entera mixta, debido a que posee variables enteras en las decisiones de construcción de nuevas líneas. En el presente trabajo dado que se quiere trabajar con una formulación sencilla para poder comparar la calidad de las soluciones de expansión de transmisión respecto a los datos de entrada, se ha escogido ocupar una formulación relajada de este problema. Para ello, se han considerando las variables de decisión como continuas buscando hacer el problema más tratable, debido a que el tratamiento de variables enteras en este tipo de formulaciones es de alto requerimiento computacional.

El modelo del TEP se presenta en las ecuaciones (2.1.10)-(2.1.22). Debido a su formulación no-lineal (ya que posee una multiplicación de variables en la ecuación (2.1.27)), se deben utilizar técnicas alternativas de resolución, dado que es difícil encontrar un óptimo global en este tipo de formulaciones.

Para la resolución del problema, se utiliza el solver *Ipopt + HSL MA86*<sup>1</sup>, el cual a través del método de punto interior busca soluciones locales del problema no lineal. Debido a que esta metodología ayuda a obtener soluciones locales del problema, es necesario definir un punto de partida adecuado para su resolución.

Para la definición de los puntos de partida adecuados se utilizará un modelo alternativo presentado en [49], el cual consiste en una formulación disyuntiva, cuya relajación de variables de construcción de líneas se conoce a través de la literatura como relajación de McCormick [50, 51], a través de la cual, el problema lineal entero mixto se relaja y es posible tratarlo como un problema lineal, cuyos requerimientos computacionales son menores. Esta formulación se presenta en (2.1.23)-(2.1.36). El problema de la relajación de McCormick radica en que sus soluciones solo son cotas inferiores de la solución requerida, por ello es que no se utiliza este modelo para las simulaciones.

Una forma de hacer más precisa la relajación es a través de envolturas de McCormick [51], las cuales se implementaron para la corroboración de los resultados obtenidos.

<sup>1</sup>*Interior Point OPTimizer* [47, 48], es una biblioteca de software para la optimización no lineal a gran escala de sistemas continuos.

Formulación no-lineal TEP:

$$\begin{aligned} \min_{p, L^{new}} \quad & \sum_{t \in T} \alpha_t S_{Base} (8760 h_t) \left( \sum_{j \in G} c_{MW}^j \sum_{t \in T} p_{j,t} + VoLL \sum_{k \in B} \sum_{t \in T} USE_{k,t} \right) + \\ & \sum_{km \in L} L_{km}^{new} c_{km} \end{aligned} \quad (2.1.10)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{j \in G(k)} p_{j,t} - \sum_{km \in L(k, \cdot)} f_{km,t} + \sum_{mk \in L(\cdot, k)} f_{mk,t} = \\ & \frac{d_{k,t}}{S_{Base}} + p_{vert} ERV_{k,t} - USE_{k,t} \quad \forall k \in B, \forall t \in T \end{aligned} \quad (2.1.11)$$

$$-\bar{\delta} \leq \delta_{k,t} \leq \bar{\delta} \quad \forall k \in B, \forall t \in T \quad (2.1.12)$$

$$\delta_1 = 1 \quad (2.1.13)$$

$$\theta_{km,t} = \delta_{k,t} - \delta_{m,t} \quad \forall km \in L, \forall t \in T \quad (2.1.14)$$

$$-\bar{\theta} \leq \theta_{km,t} \leq \bar{\theta} \quad \forall km \in L, \forall t \in T \quad (2.1.15)$$

$$f_{km,t} = -b_{km} (n_{km}^0 + L_{km}^{new}) \theta_{km,t} \quad \forall km \in L, \forall t \in T \quad (2.1.16)$$

$$-(n_{km}^0 + L_{km}^{new}) \bar{f}_{km} \leq f_{km,t} \leq (n_{km}^0 + L_{km}^{new}) \bar{f}_{km} \quad \forall km \in L, \forall t \in T \quad (2.1.17)$$

$$\underline{p}_j \leq p_{j,t} \leq \bar{p}_j \quad \forall j \in G, t \in T \quad (2.1.18)$$

$$USE_{k,t} \leq \frac{d_{k,t}}{S_{Base}} \quad \forall k \in B, t \in T \quad (2.1.19)$$

$$p_{vert} ERV_{k,t} \leq -\frac{d_{k,t}}{S_{Base}} \quad \forall k \in B, t \in T, \quad \text{ssi} \quad \frac{d_{k,t}}{S_{Base}} \leq 0 \quad (2.1.20)$$

$$L_{km}^{new} \leq \bar{n}_{km} \quad \forall km \in L \quad (2.1.21)$$

$$L_{km}^{new} \text{ continua} \quad \forall km \in L \quad (2.1.22)$$

La función objetivo (2.1.10) considera los costos de inversión de nueva infraestructura y los costos asociados a la operación del sistema en los diferentes bloques temporales, considerando la totalidad de un año (8760 horas). Cada bloque temporal será definido según el tipo de criterio (temporal o atemporal de ancho  $h_t$  con técnicas de *clustering*). Los costos de operación son evaluados para cada periodo según su correspondiente demanda. Las restricciones (2.1.11)–(2.1.17) modelan la red utilizando un modelo de flujo en continua. Restricciones en flujos de potencia se presentan en (2.1.18).

Respecto a la formulación disyuntiva con relajación de McCormick en el TEP, se presenta una estructura similar, siendo la diferencia mayor en la ecuación (2.1.29), en donde

a través de una penalización se determinan los valores de las líneas nuevas construidas.

*Formulación disyuntiva con relajación de McCormick en TEP:*

$$\begin{aligned} \min_{p, L^{new}} \quad & \sum_{t \in T} \alpha_t S_{Base}(8760h_t) \left( \sum_{j \in G} c_{MW} j \sum_{t \in T} p_{j,t} + VoLL \sum_{k \in B} \sum_{t \in T} USE_{k,t} \right) + \\ & \sum_{km \in L} \sum_{y \in Y} L_{km,y}^{new} c_{km} \end{aligned} \quad (2.1.23)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{j \in G(k)} p_{j,t} - \sum_{km \in L(k,\cdot)} (f_{km,t}^0 + \sum_{y \in Y} f_{km,t,y}) + \sum_{mk \in L(\cdot,k)} (f_{mk,t}^0 + \sum_{y \in Y} f_{mk,t,y}) = \\ & \frac{d_{k,t}}{S_{Base}} + p_{vert\ ERV\ k,t} - USE_{k,t} \quad \forall k \in B, \forall t \in T \end{aligned} \quad (2.1.24)$$

$$-\bar{\delta} \leq \delta_{k,t} \leq \bar{\delta} \quad \forall k \in B, \forall t \in T \quad (2.1.25)$$

$$\delta_1 = 1 \quad (2.1.26)$$

$$f_{km,t}^0 = -b_{km} n_{km}^0 (\delta_{k,t} - \delta_{m,t}) \quad \forall km \in L, \forall t \in T \quad (2.1.27)$$

$$-n_{km}^0 \bar{f}_{km} \leq f_{km,t}^0 \leq n_{km}^0 \bar{f}_{km} \quad \forall km \in L, \forall t \in T \quad (2.1.28)$$

$$\begin{aligned} -2\bar{\delta}(1 - L_{km,y}^{new}) \leq \frac{f_{km,y,t}}{b_{km}} + (\delta_{k,t} - \delta_{m,t}) \leq 2\bar{\delta}(1 - L_{km,y}^{new}) \\ \forall km \in L, \forall y \in Y, \forall t \in T \end{aligned} \quad (2.1.29)$$

$$-\bar{f}_{km} L_{km,y}^{new} \leq f_{km,t} \leq \bar{f}_{km} L_{km,y}^{new} \quad \forall km \in L, \forall y \in Y, \forall t \in T \quad (2.1.30)$$

$$L_{km,y}^{new} \leq L_{km,y-1}^{new} \quad \forall km \in L, \forall y \in Y \quad (2.1.31)$$

$$\underline{p}_j \leq p_{j,t} \leq \bar{p}_j \quad \forall j \in G, t \in T \quad (2.1.32)$$

$$USE_{k,t} \leq \frac{d_{k,t}}{S_{Base}} \quad \forall k \in B, t \in T \quad (2.1.33)$$

$$p_{vert\ ERV\ k,t} \leq -\frac{d_{k,t}}{S_{Base}} \quad \forall k \in B, t \in T, \text{ ssi } \frac{d_{k,t}}{S_{Base}} \leq 0 \quad (2.1.34)$$

$$\sum_{y \in Y} L_{km,y}^{new} \leq \bar{n}_{km} \quad \forall km \in L \quad (2.1.35)$$

$$L_{km,y}^{new} \text{ continua (entre 0 y 1)} \quad \forall km \in L, \forall y \in Y \quad (2.1.36)$$

Es importante recalcar que el foco principal de este trabajo es la reducción de escenario de demanda, por lo que la formulación escogida del TEP es sencilla para poder facilitar la comparación de las diferentes estrategias de selección de escenarios.

### Comentarios generales y algunos ejemplos relativos a las formulaciones escogidas

Las figuras 2.1.2-2.1.4 contrastan diferentes resultados obtenidos según el modelo matemático utilizado. La solución real del problema es la solución del problema no lineal para las 8760 horas del año, la cual será denominada como solución original, mientras que la solución de este caso con la formulación de McCormick será denominada como su solución relajada.

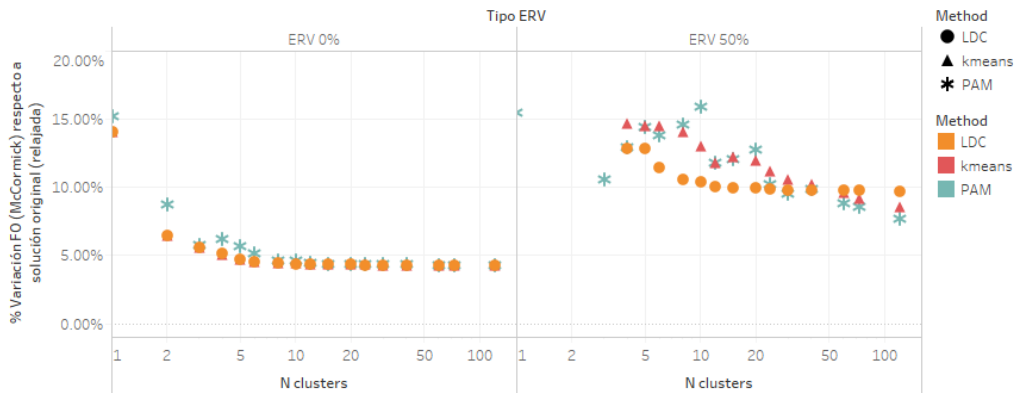


Figura 2.1.2: % Variación FO (formulación de McCormick) respecto a solución original relajada (McCormick) para diferentes niveles de generación renovable  $ERV_0$  y  $ERV_{50}$ .

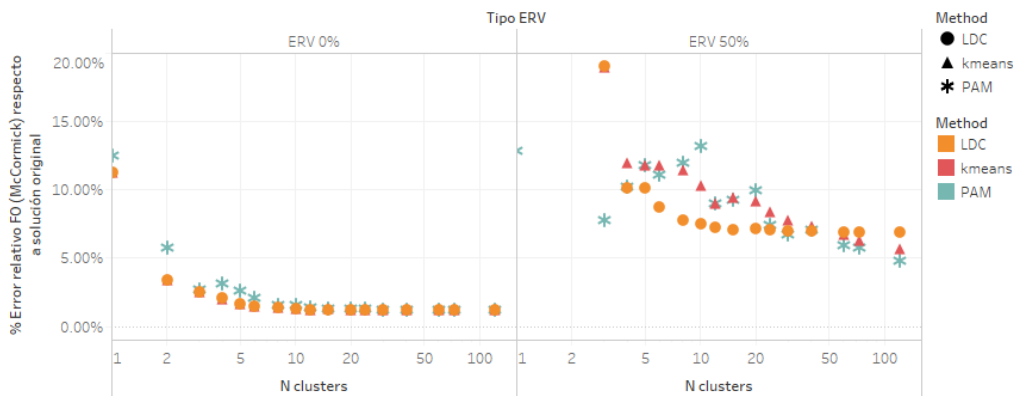


Figura 2.1.3: % Variación FO (formulación de McCormick) respecto a solución original para diferentes niveles de generación renovable  $ERV_0$  y  $ERV_{50}$ .

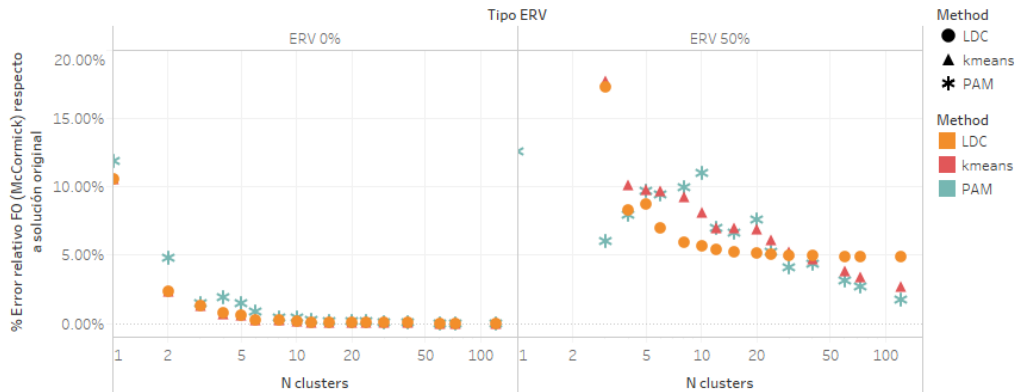


Figura 2.1.4: % Variación FO (formulación no-lineal) respecto a solución original para diferentes niveles de generación renovable  $ERV_0$  y  $ERV_{50}$ .

En las gráficas se puede apreciar que en las figuras 2.1.2 y 2.1.3, al utilizar la formulación de McCormick para resolver el TEP, se obtiene un cierto gap respecto a la solución original al evaluar el error relativo respecto a las soluciones con el total de horas. Es por ello que estas soluciones no son una referencia adecuada para la evaluación de los resultados del problema planteado, ya que este margen de diferencia puede ser significativo a la hora de comparar la calidad de las soluciones obtenidas. Así, como se observa en la figura 2.1.4 es posible conseguir soluciones más comparables utilizando la formulación no-lineal.

El uso de la formulación no-lineal es uno de los aportes que presenta este trabajo respecto a trabajos similares como [52]. En [52] utilizan como modelo el de McCormick, el cual como se observa en la figura 2.1.2, presenta un error muy alto respecto a la solución real, dado que es básicamente una cota inferior de la solución real del problema. Sin embargo, al utilizar la formulación no lineal se obtienen óptimos locales, los cuales bajo un adecuado ajuste de puntos de partida, son soluciones mejores y permiten comparar de mejor manera las diferentes soluciones obtenidas.

## 2.2. Reducción de escenarios

En diversas áreas de la ingeniería es necesario el estudio de variables independientes multi-dimensionales, dado que cada vez es más necesario el considerar el efecto de la simultaneidad de eventos en los diferentes casos de estudio, y tanto su correlación como su efecto en el resto de variables del problema. En particular, relativo al problema TEP, las variables de interés para este trabajo son la demanda de entrada del problema y la generación ERV. Estos casos de estudio consideran la simultaneidad temporal y espacial; por lo cual la representatividad de su conjunto de entrada afecta a los resultados finales del problema, es por esto que se buscará perder la menor cantidad de información relevante en el proceso.

En los estudios de mediano y largo plazo las variaciones estacionales y diarias en demanda y generación son representadas comúnmente a través de un número limitado de intervalos temporales, lo cual conlleva generalmente un problema de sub o sobrestimación de la planificación. Bajo este punto, surge el problema de cómo modelar la serie de tiempo relativa a la entrada, minimizando este error de estimación.

### 2.2.1. Tratamiento de Variables Aleatorias en TEP

El problema de planificación de la expansión de la transmisión cuenta con variables de gran incertidumbre como lo son el pronóstico de demanda del sistema, la generación, disponibilidad de tecnologías, precios del mercado, entre otras. En términos prácticos, a la hora de realizar la modelación del sistema estas variables se consideran de manera simplificada, y se intenta mantener un equilibrio entre la simplicidad y la representatividad del problema modelado. En el caso de estudio, se trabaja con un problema complejo y de difícil manejo computacionalidad dadas las no linealidades que su modelo presenta.

Además, es de considerar que los niveles de incertidumbre de variables del tipo generación y demanda son por general comunes a todos los procesos de planificación eléctrica de un sistema. Dentro de estos, existen incertidumbre consideradas de bajo y alto nivel, siendo las de bajo nivel cuantificables en base a modelos probabilísticos (e.g. hidrologías, niveles de radiación, viento, etc.), y por ende posibles de abordar en términos metodológicos a través del uso de métricas de las variables de interés, mientras que la incertidumbre de alto nivel no es cuantificable y es difícil de predecir, por lo cual es a su vez compleja de analizar ya que depende de muchos factores, entre ellos, económicos, sociales, tecnológicos, entre otros (e.g. costos de inversión, costos de combustibles).

En este último punto es relevante destacar la importancia de incorporar escenarios de expansión que consideren la incertidumbre de alto nivel, como lo es para el caso del costo de inversión de tecnologías de generación, donde las localizaciones de los recursos asociados a distintas tecnologías se encuentran distantes entre sí, y por ende sensibilizaciones en los costos de inversión de las tecnologías podrían resultar en requerimientos de desarrollo de la transmisión radicalmente opuestos [53]. En [40] se presenta una visión generalizada de como enfrentar este tipo de problemáticas en la planificación de sistemas de potencia.

El desarrollo de técnicas matemáticas relativo al manejo de incertidumbres en los problemas de optimización ha sido un gran aporte en la ingeniería eléctrica, dando

solución a este tipo de problemas y proponiendo alternativas a la manipulación de sus datos [54, 55]. En la actualidad ya hay trabajos que han implementado este tipo de desarrollos en sistemas reales [56].

En particular, el manejo de la incertidumbre es tratado a través de la generación de escenarios. Así, la programación estocástica a través del análisis de escenarios busca modelar la incertidumbre de una manera realista y desarrollar una metodología práctica, cuyo éxito depende de la facilidad en la construcción de un árbol de escenarios relativamente sencillo. En caso que el número de escenarios generados impida la resolución del problema, existen metodologías para la reducción de escenarios tales como muestreos, igualación de momentos, métricas de probabilidad, entre otras, cuyo objetivo es que el conjunto de escenarios finales afecte de manera mínima a la solución final.

### 2.2.2. Temporalidad y Espacialidad

Bajo una visión general de la problemática anteriormente expuesta, surge el problema general del cómo considerar la simultaneidad de procesos que consideran variables temporales y espaciales, como demanda y generación. Las propuestas hasta ahora mencionadas consideran diversas aproximaciones y técnicas de manejo de datos cuyo fin es lograr una mayor precisión en el modelo de modo tal que los resultados sean lo más realistas posibles.

Algunos de los artículos actuales se enfocan en el aumento de la resolución temporal relativo a los intervalos de tiempos seleccionados, aumentando su cantidad. Sin embargo, esto no es suficiente para capturar la inherente variabilidad de las ERV (se debe relacionar especialidad y temporalidad). Otras plantean técnicas de manejo de la especialidad como lo son la Demanda Neta (demanda menos generación renovable) como se ve en los casos [57, 58], sin embargo esto no soluciona problemáticas relativas a relaciones intra-temporales de datos. Es así como métodos como la aproximación por estados del sistema propuesta en [59] da cabida a estas consideraciones, pero que sin embargo requieren de mayor detalle en el modelamiento del sistema. Un contraste de esta metodología con la metodología de días representativos se presenta en [60].

Esta ha sido un área de interés en el desarrollo de modelos durante los últimos años, en donde algunos desarrollos se ven en los siguientes artículos: en [61] se presenta una metodología que busca relacionar la curva de duración de la demanda y la demanda neta del sistema a través de un algoritmo de k-means (*clustering*) y además representar la incertidumbre de la correlación entre la demanda y la producción eólica. Otro ejemplo de tratamiento de este problema a través de técnicas de *clustering* se presenta en [29], en donde se propone un nuevo método de agrupación de períodos de tiempo que supera los inconvenientes mencionados manteniendo la cronología de las series de tiempo de entrada a lo largo de todo el horizonte de planificación. Por lo tanto, el método propuesto puede evaluar correctamente el valor económico de combinar la generación de energía renovable con dispositivos de almacenamiento entre días. En [62] se desarrolla el problema anterior aplicado a un TEP, sin embargo el modelo de demanda varía y la integración ERV es modelada a través de factores de penalización en la reducción del viento.

Otras perspectivas se basan en la generación de escenarios de demanda y generación

ERV. En [63] se presenta un modelo dinámico, en donde a través del ruido blanco, y un polinomio matricial, que representa la correlación temporal y geográfica entre la carga y la potencia del viento, se busca preservar la correlación entre generación y demanda.

Otros desarrollos actuales respecto a la aplicación directa en el TEP se presentan en [30, 22, 21].

### 2.2.2.1. Espacialidad: Integración ERV y su localización

Actualmente, en conjunto con la integración de altos niveles de generación ERV en los sistemas de potencia, estos se han visto en la necesidad de generar incentivos económicos que busquen la inserción de recursos que permitan aumentar la flexibilidad del sistema<sup>2</sup>, como lo es el caso de almacenamiento en base a baterías para **Control Primario de Frecuencia (CPF)**, otras opciones de almacenamiento y/o máquinas con alta tasa de toma de carga para **Control Secundario de Frecuencia (CSF)**, condensadores sincrónicos o centrales de base para requerimientos de inercia mínima, y centrales flexibles en términos de ciclaje o almacenamientos para dar complementariedad a las ERV para el abastecimiento seguro de la demanda [53].

Otros ejemplos son la incorporación de **restricciones operativas** en el modelo de optimización de la generación, con el fin de cuantificar niveles de flexibilidad, identificar niveles de factibilidad técnica de inserción de ERV, y evaluar económicamente posibles requerimientos de flexibilidad que permitirían aumentar los niveles de inserción factible de ERV. La incorporación de **restricciones de seguridad operativa** (CPF, CSF, Inercias mínimas, mínimos técnicos, etc) y de **seguridad transmisión** (N-1) en el modelo de simulación de la operación económica, para realizar evaluaciones y ajuste de las propuestas de transmisión, cuyos órdenes de magnitud son obtenidos a partir de los modelos de optimización conjunta generación-transmisión, y adicionalmente para realizar evaluaciones económicas de los recursos que proveen flexibilidad al sistema.

Es así como, dado que la generación ERV posee alta incertidumbre y que las características de sus perfiles de generación tienen alta correlación con su localización geográfica (perfiles de viento, niveles de irradiancia solar), es necesario que su modelamiento considere un alto nivel de temporalidad y detalle geográfico (espacialidad). De este modo, la correlación espacial y temporal es uno de los datos importantes a rescatar en la representación simplificada del sistema.

En [27] se presenta el modelo Switch - herramienta de planificación para sistemas con grandes cantidades de generación renovable intermitente - el cual considera un modelo de planificación lineal por periodo, cuyo objetivo es minimizar el valor presente del costo de construcción de plantas, capacidad de transmisión, combustibles, producciones de dióxido de carbono, entre otras variables. Otros modelos consideran la optimización conjunta generación-transmisión para la propuesta de los planes de obras de generación y posterior evaluación de propuestas de transmisión, con lo cual se incorporaría de manera implícita la señal de localización en generación, y de esta forma, promover o impulsar el desarrollo de la oferta óptima a través del desarrollo de la transmisión. En [26, 21] se

<sup>2</sup>Entendida como la capacidad de todo sistema eléctrico para adaptarse a las crecientes fluctuaciones de la oferta y la demanda garantizando al mismo tiempo la confiabilidad del sistema

presenta la co-optimización de la planificación y métodos de manejo de la correlación entre variables de demanda y generación renovable, para introducirlas en el problema.

### 2.2.2.2. Temporalidad:

Relativo a la formulación de problema de expansión de la transmisión, la temporalidad es una variable determinante, debido a que de ella dependen la gran cantidad de sus variables de decisión. Además, respecto al tipo de restricciones formuladas en el problema, puede ser relevante o no la consideración inter-temporal de cada bloque de demanda (por ejemplo, restricciones de almacenamiento, encendido y apagado de plantas, etc). Sin embargo, estas no serán consideradas en el modelo de trabajo utilizado, pero son futuros desafíos que se pueden abarcar a partir de este trabajo y su implementación en metodologías como por ejemplo días típicos.

Para visualizar lo anterior se presenta la tabla 2.2.1, en la cual se representa la reducción dimensional del problema de expansión de la transmisión en el sistema RTS 96 zona A [64], al considerar el uso de diferentes bloques de demanda y en diferentes formulaciones (formulación entera y formulación relajada).

Tabla 2.2.1: Cantidad de variables según número de *clusters* y formulación en el problema de expansión de la transmisión simplificado, implementado en el sistema RTS 96 zona A.

N° <i>cluster</i>	No-lineal		No-lineal relajada
	Variables continuas	Variables enteras	Variables continuas
1	162	34	196
2	324	34	358
3	486	34	520
4	648	34	682
5	810	34	844
6	972	34	1,006
8	1,296	34	1,330
10	1,620	34	1,654
12	1,944	34	1,978
15	2,430	34	2,464
24	3,888	34	3,922
30	4,860	34	4,894
40	6,480	34	6,514
60	9,720	34	9,754
73	11,826	34	11,860
120	19,440	34	19,474
⋮	⋮	⋮	⋮
8,760	1,419,120	34	1,419,154

### 2.2.3. Modelamiento de la Demanda

Dentro del problema de planificación existe un subproblema relacionado con la modelación de la demanda como variable de entrada del sistema de estudio. El desafío es minimizar el error entre el modelo de demanda y la demanda real, considerando según el tipo de estudio las correlaciones de estas en términos espaciales y temporales. De este modo, por ejemplo, en modelos de corto plazo, la demanda es representada generalmente de manera horaria, sin embargo, en modelos de largo plazo esto no es siempre factible, dado el alto requerimiento computacional que su tratamiento implica.

Es así como se definen técnicas de representación alternativas, como lo son la discretización de la demanda a través de bloques, o técnicas un tanto más avanzadas como la generación de días representativos. En [59], por ejemplo, se plantea un modelo que realiza una descripción de estados del sistema y entorno a estos, se define una matriz de transición de sistema que ayuda a capturar de mejor forma la cronología y relación entre los datos.

Como se mencionó anteriormente, la integración de energías renovables, construcción de almacenamiento y respuesta en demanda, son variables que han impuesto nuevos requerimientos en la planificación de los SEP, ya que su dependencia temporal y espacial requiere de consideraciones cronológicas y geográficas entre las variables de estudio, como los son generación, demanda, y las instalaciones existentes del SEP. A modo de ejemplo, en trabajos como [32] se representa la interacción horaria entre el almacenamiento y la red a través de bloques de demanda agregados, los cuales se encuentran correlacionados con la generación ERV horaria, simplificando significativamente el modelo. Otro caso se presenta en [33], en donde el modelamiento se realiza en torno a la integración de la respuesta en demanda e integración de ERV al problema.

Actualmente en Chile, la Comisión Nacional de Energía realiza su planificación de transmisión según [53], proyectando su demanda como se indica en [65]. Básicamente se generan perfiles típicos de demanda de potencia media horaria, de acuerdo a los tipos de clientes y de días definidos, esto es: lunes, trabajo (martes a viernes), sábado y domingo. Estos perfiles se determinan en base a las demandas horarias reales del año anterior, para cada mes y día tipo, por cada barra de consumo modelada, de modo de poder representar de buena manera la distribución espacial y temporal de la demanda proyectada. Asimismo, para efectos de permitir una adecuada proyección de la demanda futura, estos perfiles son estimados para cada tipo de cliente, de acuerdo a la definición entregada previamente, permitiendo con esto contar con tasas de crecimiento diferenciadas para cada uno de ellos, e incluso por barra.

A partir de estos perfiles se generan curvas de carga mensual para todo el horizonte de proyección, de acuerdo a la composición de días tipo de cada mes del horizonte. De esta forma, se recoge la cantidad de días de cada tipo en cada mes del horizonte, reemplazando los feriados conocidos por días tipo domingo, en los casos que corresponda. Una vez generadas, son estas curvas de carga las que se ponderan de acuerdo a las tasas de proyección de demanda para cada tipo de cliente.

### 2.2.3.1. Curva de Duración de la Demanda

En modelos de planificación de mediano y largo plazo es común aproximar la curva de demanda a través de niveles, de modo tal de hacer a estos computacionalmente tratables. Esta curva es conocida como **Curva de Duración de la Demanda -Load Duration Curve (LDC)-** y típicamente es modelada por el ajuste de una función continua. Posteriormente, a esta función se le determinan los bloques de operación, consistentes en discretizaciones que aproximan toda la curva por un número reducido de segmentos. El problema central de esta metodología consiste en la irregularidad del comportamiento de los datos originales y por lo mismo, existe un error de ajuste de las funciones que la describen.

Las primeras publicaciones respecto a esta metodología datan de 1949 en un artículo publicado por M. Boiteux [66, 67] relativo al problema de tarificación a costo marginal. La evolución de esta aproximación a través del tiempo se desarrolla en [68, 69, 70, 71, 72]. El objetivo es aproximar la curva de demanda de un sistema en  $L$  bloques, donde cada bloque de demanda define un nivel y duración. Para lograr este objetivo, se resuelve un problema de optimización que divide la curva de carga continua en  $L$  grupos. Esta partición de datos busca minimizar la suma, sobre todos los grupos, de las distancias punto a punto entre el grupo y un “centro”<sup>3</sup>. El resultado de este proceso produce una representación de la demanda discreta que parece una función escalonada.

La ventaja de esta metodología es que en vez de una representación detallada de 8760 horas al año, la demanda puede ser representada solo por  $L < 8760$ . Así esta reducción en el tamaño del modelo permite expandir el horizonte de tiempo a más años (décadas). Sin embargo, su desventaja radica en la pérdida de secuencialidad de los datos, la que cobra más importancia en planificaciones de corto plazo.

Algunos casos de aplicación de esta metodología se presentan en [30, 32, 62]

**Días Típicos** Debido a las restricciones computacionales, comúnmente los modelos de planificación suelen representar variaciones intra-anales en la demanda mediante el uso de un número limitado de los llamados intervalos de tiempo, generalmente discretizado a través de la LDC. En [73] se realiza un contraste de ambas metodologías.

Capturar el carácter variable inherente de las fuentes de ERV en este número limitado de segmentos de tiempo es un desafío, dado que el agrupamiento temporal debe considerar el factor espacial en su desarrollo. Un enfoque para hacer esto es usar un pequeño conjunto de días históricos para representar el año completo (los llamados días representativos), sin embargo, no hay modelos de optimización disponibles en la literatura para optimizar la selección de días representativos, por lo cual los desarrollos de estos tienden a ser obtenidos a través de simple heurística, pero en otros casos se complementan seleccionando aleatoriamente algunos días adicionales.

Algunos enfoques diferentes para seleccionar días representativos [74]:

#### 1. Heurísticas.

---

<sup>3</sup>Al existir diferentes metodologías de agrupamiento de datos o *clustering*, el tipo de centro se escogerá según la técnica empleada. La metodología más comúnmente empleada para este tipo de casos es el método de mínimos cuadrados.

2. Algoritmos de *Clustering* Jerárquicos [75].
3. Selección Aleatoria: un subconjunto aleatorio de días y retiene el subconjunto con los errores más bajos.
4. Modo de optimización de programación lineal de enteros mixtos (MILP OPT): es un enfoque recientemente desarrollado que emplea un modelo de optimización MILP para identificar qué días se seleccionan (variables binarias), así como el peso asignado a cada día (variables lineales).
5. Enfoque híbrido (HYB): Selección Aleatoria seguida de una ponderación óptima: el peso asignado a cada día se optimiza.

Otra agrupación de técnicas de selección de días representativos son:

- **Algoritmos de Clustering (CA):** para agrupar periodos con similar demanda, velocidad del viento y/o irradiancia solar dentro de *clusters*. El análisis de agrupamientos de datos o *Clustering* es la tarea de agrupar un conjunto de objetos de tal manera que los objetos en el mismo grupo (llamado grupo) sean más similares (en cierto sentido) entre sí que a los de otros grupos (grupos). Es una tarea principal de la minería de datos exploratoria y una técnica común para el análisis de datos estadísticos, utilizada en muchos campos, incluido el aprendizaje automático, el reconocimiento de patrones, el análisis de imágenes, la recuperación de información, la bioinformática, la compresión de datos y los gráficos de computadora [76]. Ejemplos de aplicaciones en [77, 78]
- **Optimizar la selección de periodos representativos con respecto a un criterio definido por el usuario y un indicador numérico externo:** la selección es basada en las características de cada periodo de tiempo de manera individual o la similitud entre periodos de tiempos.

Generalmente, la necesidad de la implementación de estas metodologías está directamente ligada a la integración de ERV, almacenamiento o respuesta en demanda en los problemas de planificación eléctrica, entre otros. En [27, 75, 29, 28] se desarrollan ejemplos de estos casos. Además en [75, 74] se plantean otros desarrollos y metodologías de días típicos y en [77, 78] se presentan aplicaciones concretas en modelos de expansión.

### 2.2.3.2. Aproximación por bloques de la curva de duración de la demanda

La curva de duración de la demanda, o curva de carga, es un método comúnmente utilizado en la planificación de sistemas eléctricos de potencia cuya finalidad es representar la demanda del sistema de una manera reducida y representativa. Esta se construye ordenando todas las demandas totales de cada hora del sistema de mayor a menor, perdiendo su cronología, pero conservando su equivalente en horas porcentuales de duración. De este modo, se identifican los valores máximos y mínimos de la demanda y, posteriormente, mediante un procedimiento de cuadratura, es posible convertir la curva en bloques de demanda representativos, tales como la demanda base (la cual ocurre las

8760 horas del año); la demanda intermedia, que se encuentra por arriba de la mínima y que ocurre un par de horas al año; y la demanda de punta que solo ocurre algunas pocas horas al año.

Este método ha sido muy empleado en los problemas de operación y planificación de sistemas de potencia, ya que en sistemas puramente térmicos -y en algunos casos hídricos dentro de cada intervalo  $t$ -, luego es completamente equivalente emplear la demanda cronológica o la curva de duración de la demanda en ambos problemas. Esta metodología ha sido empleada además en sistemas de transmisión y decisiones de inversión enteras.

Su mayor ventaja se debe a que, al ser una curva unidimensional y monótona, puede aproximarse a una función constante por tramos, con un error relativamente bajo permitiendo resolver las ecuaciones de flujo de potencia para solo un número acotado de valores, llamados "bloques de demanda" (los que además son ponderados por un factor de tiempo representativo), en lugar de utilizar todos los valores de la curva de duración. De este modo, según la cantidad de bloques elegidos, y el criterio de agrupamiento de estos, es posible obtener diferentes soluciones óptimas en el problema. Sin embargo, dado el gran número de variables y su respectiva naturaleza entera-mixta, resulta muy complejo el utilizar todos los valores de la curva de duración, debido a la gran cantidad de variables que se debe considerar en el problema de optimización.

#### 2.2.4. Técnicas de Agrupamiento de Datos: Clustering

Como se mencionó anteriormente, el gran número de variables manejadas en este tipo de problemas conllevan el uso de técnicas avanzadas de manejo de datos e información. Es así como surge la importancia de la ciencia de datos -*Data Science*- o el Análisis de Datos -*Data Analytics*- como procesos de extracción de conocimiento significativos de estos. Sin embargo, al ir un poco más allá, resulta interesante el encontrar una estructura interesante de estos, lo que hace surgir una nueva área de interés que es la Minería de Datos -*Data Mining*- y de este modo la estructura de datos puede tomar muchas formas, incluyendo un conjunto de reglas, un gráfico o red, un árbol, una o varias ecuaciones, entre otras [79].

Relativo a las aplicaciones en ingeniería eléctrica, para problemas que consisten en el manejo de datos, existen una serie de técnicas relativas al Aprendizaje de Máquinas -*Machine Learning*-, como se observa en la figura 2.2.1.

En este trabajo es de particular interés el *clustering*, la cual es una técnica de aprendizaje de máquinas que involucra la agrupación de puntos de datos. Dado un conjunto de datos, es posible utilizar un algoritmo de *clustering* para clasificar cada punto en un grupo específico. En teoría, los puntos de datos que están en el mismo grupo deberían tener propiedades similares y/o características a fines, mientras que los puntos en diferentes grupos de datos deberán ser altamente diferentes. *Clustering* es un método de aprendizaje no supervisado y es una técnica común para análisis de datos estáticos usada en muchos campos.

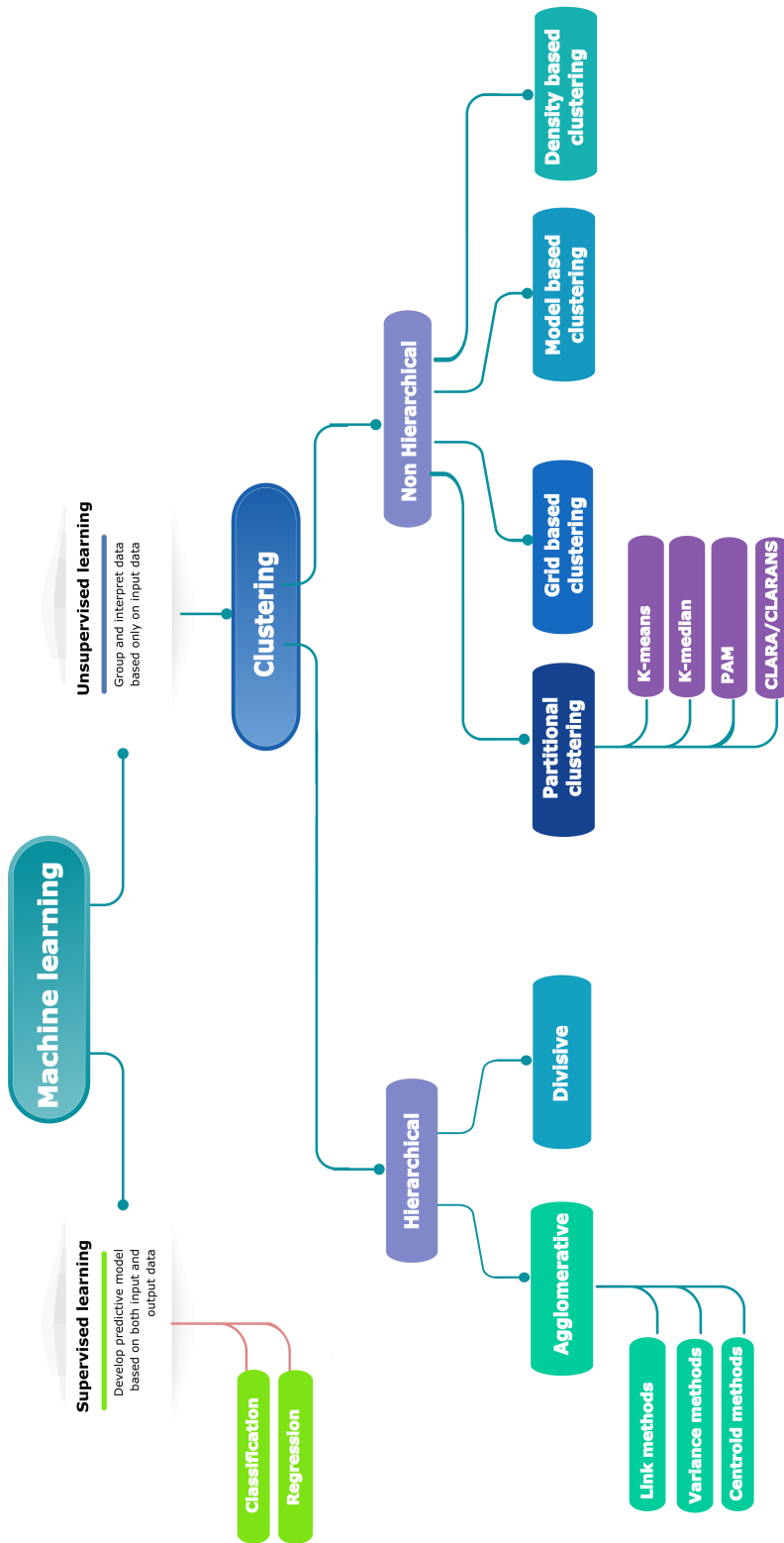


Figura 2.2.1: Subdivisiones del área de Aprendizaje de Máquinas. [Fuente: Elaboración propia a partir de [2]]

### 2.2.4.1. Clustering aplicado a la reducción de escenarios.

El término *clustering* hace referencia a una variedad de técnicas *unsupervised* cuya finalidad es encontrar patrones o grupos (*clusters*) dentro de un conjunto de observaciones. Las particiones se establecen de forma que, las observaciones que están dentro de un mismo grupo, son similares entre ellas y distintas a las observaciones de otros grupos. Se trata de un método *unsupervised*, ya que el proceso ignora la variable respuesta que indica a que grupo pertenece realmente cada observación (si es que existe tal variable). Esta característica diferencia al *clustering* de las técnicas estadísticas conocidas como análisis discriminante, que emplean un set de entrenamiento en el que se conoce la verdadera clasificación.

Dada la utilidad del *clustering* en disciplinas muy distintas (genómica, marketing...), se han desarrollado multitud de variantes y adaptaciones de sus métodos y algoritmos. Pueden diferenciarse en diferentes grupos, como por ejemplo:

- **Métodos que combinan o modifican los anteriores:** *hierarchical k-means, fuzzy clustering, model based clustering y density based clustering*
- **Según su definición de similitud:** Puedes establecer diferentes definiciones de similaridad, basados en distancia, densidad, semejanza a distribuciones de probabilidad, etc.
- **Según partición o jerarquización** Algoritmos que establecen los *clusters* sobre una muestra de datos es un único cálculo (particionandolo) o en forma jerárquica, vía particiones o uniones sucesivas. De este modo se definen:
  - **Partitioning Clustering:** Este tipo de algoritmos requieren que el usuario especifique de antemano el número de *clusters* que se van a crear (k-means, k-medoids, CLARA).
  - **Hierarchical Clustering:** Este tipo de algoritmos no requieren que el usuario especifique de antemano el número de *clusters*. (*agglomerative clustering, divisive clustering*).
  - **Según tipo de partición blanca o dura:** La agrupación puede realizarse construyendo *clusters* duros - *hard partition* - en los que los elementos de la muestra pueden pertenecer a un *cluster* u otro en forma excluyente; o *cluster* difusos - *soft partition o fuzzy clustering* - en que los elementos pueden pertenecer a varios *clusters* a la vez, con distintos grados de pertenencia a cada uno.

Al igual que en [1], en el desarrollo de esta tesis se trabaja con la definición de similitud en base a distancia, puesto que se emplean datos numéricos con unidades físicas similares. Además, basados en base a esto se han escogido las metodologías de *clustering* para emplear en este trabajo, las cuales buscan verificar el impacto de la simultaneidad y espacialidad de las variables de entrada de los modelos de SEP, por lo cual se utilizarán solo técnicas duras de *clustering*.

A continuación se realizará una pequeña descripción de los algoritmos de *clustering* utilizados en esta tesis, cuyo mayor desarrollo se presenta en la tesis [1] y en la sección anexos. No se discute mayormente su implementación, puesto que se utilizan implementaciones ya desarrolladas para el software R y incluidas en los paquetes *cluster* [80] y *fpc* [81].

En la tabla 2.2.2 se presenta un resumen con las principales técnicas *clustering* utilizadas en este trabajo y sus respectivas ventajas y desventajas.

Técnica	Método	Ventaja/desventaja
<b>K-means</b> [82]	<p>Este es uno de los algoritmos más antiguos y data de la década de 1950. Requiere saber de antemano el número de <i>clusters</i> en los datos a trabajar. Organiza los datos en torno a los centros de los <i>clusters</i>, respecto a alguna medida de distancia. Inicialmente se selecciona empíricamente la cantidad K de clases/grupos de datos a usar, y se inician al azar sus puntos centrales. Los datos se clasifican calculando la distancia entre ese punto y el respectivo centro del grupo, y se clasifica en el grupo cuyo centro esté más cercano a él. Luego, en función de los puntos ya clasificados, se recalcula el centro del grupo a través de la media de todos los vectores del grupo. Se itera en el mismo procedimiento hasta que los centros de grupo no cambian significativamente entre iteraciones.</p> <p>El algoritmo entrega soluciones mínimos locales. Este a través de la distancia euclídea organiza los <i>clusters</i> en torno a sus centroides.</p> <p><i>Algoritmo:</i></p> <ol style="list-style-type: none"> <li>1. Escoge n puntos como «centros» de los K <i>clusters</i>.</li> <li>2. Asocia cada elemento al centro más cercano.</li> <li>3. Calcula los nuevos «centros» usando los elementos asociados en el paso anterior.</li> <li>4. Repetir pasos 2-3 hasta convergencia.</li> </ol>	<p><i>Ventajas de K-Means:</i></p> <ol style="list-style-type: none"> <li>1. Si se tiene un gran número de variables, <i>k-means</i> la mayor parte del tiempo es más rápido que los <i>Hierarchical Clustering Methods</i>.</li> <li>2. <i>K-Means</i> produce <i>clusters</i> más estrictos que los <i>Hierarchical Clustering Methods</i>.</li> </ol> <p><i>Desventajas de k-Means:</i></p> <ol style="list-style-type: none"> <li>1. Su valor <b>k</b> (o número de particiones) es difícil de predecir.</li> <li>2. Al iniciar aleatoriamente los centros del <i>cluster</i> puede producir resultados muy diferentes en diferentes ejecuciones del algoritmo, haciendo que los resultados a veces carezcan de consistencia y lo hace difícil de comparar versus otros <i>clusters</i>.</li> </ol>
<b>K-medians</b>	<p>Al igual que <i>k-means</i>, realiza un proceso similar, pero calcula el centro del grupo a través de la media.</p>	<p>Este método es menos sensible a valores atípicos, pero es mucho más lento para conjunto de datos grandes, por que requiere clasificación en cada iteración cuando se calcula el vector mediana.</p>
<b>K-Medoid:</b> <b>PAM (Partitioning Around Medoids)</b> [82]	<p>Planteado a principio de la década de 1990</p> <p>Algoritmo basado en programación dinámica</p> <p>Soporta cualquier tipo de medida de distancia/disimilitud</p> <p>Algoritmo:</p> <p>FASE BUILD</p> <ol style="list-style-type: none"> <li>1. Elige el elemento respecto al cual se minimiza la suma de distancias como primer medoide</li> <li>2. Toma un elemento no medoide y los elementos que ya son medoides, y asigna los elementos restantes a <i>clusters</i> según el medoide más cercano. Calcula la suma de disimilitudes dentro de los <i>clusters</i>.</li> <li>3. Repite el paso dos para todos los elementos no medoides y selecciona aquel que entrega la menor suma de disimilitudes como nuevo medoide.</li> <li>4. Repite los pasos 2:3 según número de <i>clusters</i> a formar</li> </ol> <p>FASE SWAP: Una vez terminada la Fase <i>Build</i>, prueba mover de <i>cluster</i> cada uno de los elementos en forma individual. Si se disminuye la suma de disimilitudes con alguno de estos movimientos se reasigna el elemento.</p>	<p>Ventajas de k-medoid:</p> <p>Entrega resultados más estables y robustos que k-means</p> <p>Las distancias entre los elementos puede calcularse en forma previa y una sola vez, lo que permite usar cualquier tipo de distancia.</p> <p>Asignación determinista de elementos</p> <p>Es simple de entender y fácil de implementar</p> <p>Tiene rápido y de convergencia en un número fijo de pasos</p> <p>Es menos sensible a valores atípicos que otros algoritmos de partición.</p> <p>Desventajas de k-medoid:</p> <ol style="list-style-type: none"> <li>1) Es más costoso que k-means dada su complejidad temporal</li> </ol> <p>No funciona bien para grandes cantidades de datos</p> <p>Sus resultados y su tiempo de simulación dependen de la partición inicial.</p>

Tabla 2.2.2: Tabla resumen características principales de algunas metodologías de *Clustering* [2].



## Capítulo 3

# Reducción Dimensional de la Demanda Considerando Diversidad Espacial de Generación Renovable

**RESUMEN:** El problema de la Planificación de la Expansión de la Transmisión (TEP) puede volverse complejo de resolver computacionalmente a medida que se incluyen más variables en la formulación. Es por ello que los planificadores generalmente limitan sus estudios a unas pocas condiciones operativas representativas. Por lo tanto, este capítulo propone una metodología de agrupamiento para la selección de condiciones operativas en TEP, considerando la creciente penetración de energía renovable variable. El método tradicional de descomposición de la curva de duración de carga (LDC) se contrasta con las técnicas de agrupamiento como k-means y PAM que consideran la diversidad espacial de la demanda neta. El método propuesto se prueba en el sistema de prueba IEEE de 24 barras para diferentes niveles de penetración renovable (0 %, 5 %, 12 %, 26 % y 50 %), ubicando parques eólicos y solares en diferentes localizaciones para mostrar los efectos de la variabilidad espacial y temporal en los perfiles de demanda neta. Los *clusters* se utilizan como entradas en el problema TEP y los resultados se comparan con la solución del problema completo (entradas sin *clusters*) utilizando diferentes métricas. Los resultados numéricos muestran que el número óptimo de grupos depende de la variabilidad de los perfiles de carga neta. Además, la selección de condiciones operativas para los estudios de TEP utilizando técnicas de agrupamiento reduce el error con respecto al método LDC a medida que la demanda neta se vuelve más diversa espacialmente.

### 3.1. Introducción

El rápido aumento en las inversiones relativas a energías renovables (**ER**), debido tanto a razones ambientales como económicas, se ha desarrollado de la mano con su tecnología, permitiendo disminuir costos de manera significativa en su construcción e

implementación [3, 83]. Es por ello que los Sistemas Eléctricos de Potencia (**SEP**) deben adaptarse a estas tecnologías de generación, incrementando los niveles de flexibilidad operativa del sistema<sup>1</sup>, por lo que se requiere de la integración de estas variables a la planificación eléctrica de los sistemas. De este modo, surge la necesidad de buscar nuevas metodologías que ayuden a simplificar sus cálculos, sin perder la precisión requerida en estos.

Relativo al modelo del sistema, al tratarse de un estudio de mediano y largo plazo del SEP, este tipo de problemas poseen un gran número de variables de trabajo y probablemente de complejo manejo computacional. Es por ello que el problema debe ser tratado de una forma aproximada, agregando variables con la finalidad de representar la operación de manera reducida, procurando perder la menor cantidad de información relevante en el proceso.

Típicamente en problemas de planificación de la expansión, las variables y parámetros se encuentran indexados de manera espacial (e.g. por barra, nodo o región), temporal (e.g. por hora, día, mes o año) y por escenario. Por ello, usualmente podemos destacar tres tipos de agregación de variables para reducir la dimensionalidad del problema y la disminución del número de variables de decisión, restricciones, y parámetros:

1. Espacial: Por ejemplo, representar varias barras relativamente cercanas por una única barra.
2. Temporal: Por ejemplo, representar un conjunto de horas con características similares de demanda por un bloque, o varios días por un día representativo.
3. Estocasticidad: Por ejemplo, representar varias posibles realizaciones de un fenómeno estocástico por un solo escenario.

Dados los antecedentes anteriormente mencionados, el estudio de variables independientes multi-dimensionales en los modelos de operación, tales como sus datos de entrada (demanda y generación de energías renovables variables (ERV)), serán el foco principal de este trabajo. Por ello se busca analizar el efecto que tiene el modelamiento y selección de estos datos de entrada en la planificación final del sistema, buscando establecer una relación entre el detalle de este proceso y la magnitud de los beneficios computacionales resultantes.

Una simplificación comúnmente utilizada en estos estudios es la discretización de las variables de entrada en bloques. En el caso de la demanda, esta ha sido tratada históricamente en los problemas de planificación eléctrica de forma discreta, trabajándose en bloques a través de la Curva de Duración de la Demanda o Carga -*Load Duration Curve (LDC)*<sup>2</sup>-, cuyo objetivo es aproximar la curva de demanda de un sistema en  $L$  bloques, donde cada bloque de demanda define un nivel y duración. La ventaja de esta

<sup>1</sup>Entendida como la capacidad de todo sistema eléctrico para adaptarse a las crecientes fluctuaciones de la oferta y la demanda garantizando al mismo tiempo la confiabilidad del sistema

<sup>2</sup>La Curva de Duración de la Demanda ilustra la variación de la demanda de un sistema en orden descendiente, en lugar de su orden cronológico. Esta es utilizada generalmente para ilustrar la relación entre los requisitos de capacidad de generación y sus respectivos usos, dado un cierto incremento de carga.

metodología es que en vez de una representación detallada de 8760 horas al año, la demanda puede ser representada solo por  $L < 8760$ . Así, esta reducción en el tamaño del modelo permite expandir el horizonte de tiempo a más años (décadas) o aumentar el detalle de otros aspectos del modelo. Sin embargo, su desventaja radica en la pérdida de secuencialidad de los datos, y con ello información relevante relativa a temporalidad y causalidad del problema.

Otra variable de entrada del sistema con gran relevancia en este estudio es la generación de energías renovables variables (**ERV**)<sup>3</sup>, la cual se considera como ingresos de energía de disponibilidad limitada que posee variabilidad e incertidumbre, y cuyos perfiles de generación poseen alta correlación con su localización geográfica.

El presente trabajo se basa en el análisis de datos multi-variables relativos a demanda y generación ERV y su aplicación en la resolución del problema de expansión de la transmisión. Se busca proponer y aplicar metodologías para la representación reducida de estas variables, considerando formulaciones que simplifiquen sus relaciones de temporalidad y espacialidad, minimizando en el proceso la pérdida de información relevante para la resolución del problema. Para ello, se propone realizar la reducción de la dimensión temporal de la demanda a través de técnicas de *clustering* en escenarios de demanda-generación ERV, conservando en este proceso la variable espacial. Las técnicas utilizadas serán k-means y PAM, las cuales serán comparadas con la representación de la LDC bajo diferentes escenarios.

El método propuesto se prueba en el sistema de prueba IEEE de 24 barras para diferentes niveles de penetración renovable (0 %, 5 %, 12 %, 26 % y 50 %), ubicando parques eólicos y solares en diferentes localizaciones para mostrar los efectos de la variabilidad espacial y temporal en los perfiles de demanda neta.

Finalmente, para el análisis de los beneficios relativos a la metodología propuesta se utilizan diferentes métricas de comparación de los resultados obtenidos respecto a la solución del problema sin reducción de escenarios. Las métricas utilizadas son: MSE, MAE y correlaciones para análisis de los vectores de nuevas líneas construidas, comparaciones porcentuales de variación respecto a función objetivo y nuevas líneas, y además, se introduce una métrica que se denominará “FO recalculada”, la cual consiste en comparar el valor de la función objetivo de la solución real (8760 horas) con la solución de las nuevas líneas construidas obtenida aplicando la reducción de escenarios, y evaluando esta en un OPF, de modo tal de obtener la operación bajo la escala temporal original.

### 3.1.1. Estado del Arte

#### 3.1.1.1. Modelamiento de la Demanda

Dentro del problema de planificación existe un subproblema relacionado con la modelación de la demanda como variable de entrada del sistema de estudio. El desafío es minimizar el error entre el modelo de demanda y la demanda real, considerando según el tipo de estudio las correlaciones de estas en términos espaciales y temporales. De este modo, por ejemplo, en modelos de corto plazo, la demanda es representada generalmente

---

<sup>3</sup>En particular, se considerarán como energías renovables variables la generación solar y eólica

de manera horaria. Sin embargo, en modelos de largo plazo esto no es siempre factible, dado el alto requerimiento computacional que su tratamiento implica.

Es así como se definen técnicas de representación alternativas, como lo son la discretización de la demanda a través de bloques, o técnicas un tanto más avanzadas como la generación de días representativos [84]. Otro tipo de representaciones son, por ejemplo, el modelo presentado en [59], donde se plantea una descripción de estados del sistema y respecto a estos se define una matriz de transición, que ayuda a capturar de mejor forma la cronología y relación entre los datos<sup>4</sup>.

El presente trabajo se basa en una metodología sencilla de procesamiento de datos de demanda, considerando de forma particular la información relevante respecto a los perfiles de entrada del problema, y rescatando a partir de ellos y su variabilidad lo relevante para el problema de expansión de la transmisión. De este modo, además se analizan diferentes casos de estudios, para variados perfiles de generación ERV y diferentes localizaciones, de modo tal de capturar los reales beneficios de la metodología en diferentes escenarios de ERV.

Trabajos similares se presentan en [85, 86], en donde se plantea una metodología de trabajo similar a la desarrollada, sin embargo el planteamiento de la reducción de escenarios es respecto a la variable estocástica, tema en el cual se han desarrollado numerosos trabajos respecto al tratamiento de la variabilidad e incertidumbre en el problema de planificación de la transmisión. Otros desarrollos más orientados a la reducción de la temporalidad del problema han sido desarrollados respecto al problema de expansión de generación del sistema, como se presenta en [29].

Relativos a la reducción de la espacialidad del problema se tienen trabajos como [87, 88] en donde consideran agrupamientos por zonas.

Finalmente, el trabajo más reciente respecto a esta temática se presenta en [89], en donde se desarrollan tres nuevas técnicas para reducir el tamaño del problema TEP en sus tres dimensiones principales: la representación espacial [88], la representación operacional (representación temporal) [52] y el número de elementos de la red a considerar. Si bien el desarrollo de esta tesis posee aspectos similares a los desarrollados en [89] en cuanto a la reducción temporal, las técnicas de reducción de escenarios propuestas son distintas, intentando replicar algunos casos comparativos para poder rescatar los beneficios de las nuevas técnicas propuestas. Además se propone un mayor número de casos de estudio y un modelamiento diferente del problema de expansión de la transmisión considerando nuevas métricas de comparación de resultados.

**Curva de Duración de la Demanda** En modelos de planificación de mediano y largo plazo es común aproximar la curva de demanda a través de niveles, de modo tal de hacer a estos computacionalmente tratables. Esta curva es conocida como **Curva de Duración de la Demanda** -*Load Duration Curve* (LDC)- y típicamente es modelada por el ajuste de una función continua.

Las primeras publicaciones respecto a esta metodología datan de 1949 en un artículo publicado por M. Boiteux [67, 66] relativo al problema de tarificación a costo marginal.

---

<sup>4</sup>Si bien este modelo entrega una reducción de escenarios, su planteamiento no está desarrollado entorno al TEP, el cual requiere de consideraciones adicionales respecto a la naturaleza del problema.

La duración (pesos) de cada caso de demanda es calculada minimizando el error cuadrático entre el nivel de demanda y la data disponible. Una metodología un tanto más sofisticada para ello es la aplicación de técnicas de *clustering*, específicamente el algoritmo de k-means para la selección de los niveles de demanda y sus respectivos pesos [90].

La evolución de esta aproximación a través del tiempo se desarrolla en [72, 70, 69, 71, 68]. Algunos casos de aplicación de esta metodología se presentan en [30, 32, 62].

**Temporalidad y Espacialidad** Esta ha sido un área de interés en el desarrollo de modelos durante los últimos años, en donde algunos desarrollos se ven en los siguientes artículos: en [61] se presenta una metodología que busca relacionar la curva de duración de la demanda y la demanda neta del sistema a través de un algoritmo de k-means (*clustering*) y además representar la incertidumbre de la correlación entre la demanda y la producción eólica. Otro ejemplo de tratamiento de este problema a través de técnicas de *clustering* se presenta en [29], en donde se propone un nuevo método de agrupación de períodos de tiempo que supera los inconvenientes mencionados manteniendo la cronología de las series de tiempo de entrada a lo largo de todo el horizonte de planificación.

En [62] se desarrolla el problema anterior aplicado a un TEP, sin embargo el modelo de demanda varía y la integración ERV es modelada a través de factores de penalización en la reducción del viento.

Algunos estudios se basan en la generación de escenarios de demanda y generación ERV. En [63] se presenta un modelo dinámico, en donde a través del ruido blanco, y un polinomio matricial, que representa la correlación temporal y geográfica entre la carga y la potencia del viento, se busca preservar la correlación entre generación y demanda. Otros desarrollos actuales respecto a la aplicación directa en el TEP se presentan en [30, 21, 22].

Cabe mencionar que un aspecto importante en este tipo de problema es la relación intertemporal de variables, que se presenta en casos de integración de almacenamiento y variables operacionales en el sistema, y cuya representación se hace generalmente a través de metodologías como los días típicos. Si bien, el presente trabajo no considera la relación intertemporal de variables en la planificación eléctrica, su aporte relativo a la consideración de la espacialidad y variabilidad al generar bloques representativos de demanda puede ser provechoso para este tipo de aplicaciones, y por tanto para el posterior desarrollo de metodologías de este tipo.

### 3.1.2. Contribuciones y Organización del Capítulo

#### 3.1.2.1. Contribuciones

Las principales contribuciones de este trabajo son:

1. Una metodología simple de representación reducida de la demanda basada en técnicas de *clustering* para los datos de entrada del problema. Se propone emplear las técnicas de k-means y PAM para la reducción de la temporalidad de la demanda, buscando de paso rescatar la información espacial presente en el sistema.

2. Contrastar los resultados de la aplicación de estas representaciones reducidas de la demanda con la técnica de la LDC en el sistema de RTS 96 de 24 barras, bajo diferentes perfiles de demanda y generación renovable, implementando además casos de estudio con diferentes localizaciones de fuentes renovables con el objetivo de comparar los efectos de su localización en el problema.
3. El uso de un modelo de planificación de la transmisión con variables de decisión relajadas, es decir, se utilizarán variables de decisión continuas en vez de enteras. De este modo se busca contrastar de mejor manera las diferentes soluciones obtenidas y no ver limitadas las soluciones a la discretización de estas variables.
4. A diferencia de otros trabajos, se implementará un modelo de planificación con variables relajadas no lineal, el cual será resuelto a través del método de punto interior, obteniendo los puntos iniciales a través de un problema disyuntivo relajado (formulación de McCormick), cuyas soluciones finales son más comparables dada su mayor precisión.
5. Un análisis de la importancia de la variabilidad temporal y espacial al seleccionar una técnica de representación reducida en el TEP.

### 3.1.2.2. Organización del Capítulo

El presente capítulo se organiza según:

- Introducción: se plantea la problemática abordada.
- Estado del Arte: se hace un breve resumen respecto a la literatura actual referente al tema, y a los principales trabajos utilizados como referentes para el desarrollo de esta tesis.
- Metodología: se realiza una visión general respecto a las variables y simplificaciones del problema de expansión de transmisión consideradas en este trabajo. Además, se describen las metodologías utilizadas para la reducción de dimensionalidad a través de LDC, k-means y PAM.
- Casos de estudio: de manera general se describen los casos de estudio seleccionados con sus respectivos datos considerados. Se presentan además las métricas de evaluación de resultados utilizadas en el trabajo.
- Resultados Numéricos: se presentan los principales resultados numéricos a través de gráficas representativas y tablas resumen de resultados principales.
- Conclusiones: se indican las principales conclusiones de este trabajo y las posibles extensiones futuras.
- Anexos: se adjuntan 4 secciones anexas, las cuales consisten en desarrollo de las metodologías de LDC y *clustering*, datos específicos de los sistemas de estudio y resultados anexos de las simulaciones realizadas.

## 3.2. Metodología

### 3.2.1. El problema de la expansión de la Transmisión (TEP)

Una adecuada planificación y expansión de la transmisión de un sistema eléctrico en un mercado competitivo debe permitir su desarrollo a un mínimo costo, y proveer de señales que incentiven la eficiencia económica y técnica de éste, adaptando los requerimientos de generadores y consumidores, remunerando adecuadamente a sus propietarios [35].

El TEP busca encontrar respuestas al dónde, cuándo y cómo nuevas líneas deberían ser instaladas en el sistemas para ser capaz de abastecer su demanda, sujeto a restricciones físicas y económicas [24]. De este modo, en su planteamiento matemático el TEP es un problema de programación no lineal entero mixto con variables de decisión binarias relativas a la construcción de nuevas líneas o su expansión.

En [44] se presenta una visión generalizada del estado de arte actual del problema, y se desarrolla una metodología para implementación en un sistema de escala real y en [45] se expone una revisión de los modelos de planificación de la transmisión, clasificándolos relativo a sus métodos de resolución, horizontes de planificación y el prospecto de modelado.

El objetivo principal de esta tesis es el análisis desde el punto de vista de los datos de entrada del TEP, por lo cual el enfoque principal es relativo a dos variables multidimensionales de entrada del sistema, las cuales son la demanda y la generación ERV. Relativo a la demanda, se estudia de forma no-cronológica (a través de técnicas de agrupamientos de datos), buscando orientar este trabajo para sentar las bases de futuros estudios en que se considere la cronología, como lo es por ejemplo a través de la metodología de días típicos.

Para el presente trabajo se utiliza un modelo de planificación estático, es decir, se define el horizonte de planificación (se utiliza un año, es decir, 8760 horas), en donde en el tiempo  $t_0$  se realiza inversión en la transmisión para el resto horizonte temporal. Del mismo modo, se realizan las proyecciones de demanda para este tiempo y se valorizan los costos de operación del sistema para dicho año. Posteriormente, se repite el proceso para años futuros, lo cual queda representado al llevar a su valor presente la inversión, multiplicando los costos operacionales de un año del sistema por el factor uno sobre la tasa de descuento utilizada. De este modo es posible hacer comparable costos de inversión y operación del sistema en un horizonte temporal.

Respecto al modelo de planificación, se ha escogido un modelo de planificación no lineal, cuyas variables de decisión respecto a cantidad de nuevas líneas son enteras. Sin embargo, dada la naturaleza entera del problema, se ha escogido además realizar una relajación a este problema. Para ello se han considerando las variables de decisión de nuevas líneas construidas como continuas, haciendo de este modo el problema más tratable computacionalmente.

Es importante notar que según el tipo de modelo de planificación y sus restricciones, es posible que a medida que se considera un mayor número de *clusters* el problema aumente su cantidad de variables enteras de decisión, y por ende aumenten de manera

exponencial sus tiempos de simulación, por lo que encontrar la solución para el horizonte temporal total se hace altamente costoso computacionalmente. Sin embargo, dado el modelo de planificación estático escogido, se han utilizado restricciones que mantienen fijo el número de variables enteras<sup>5</sup>, y que por tanto simplifican esta problemática.

Además, al considerar las variables de decisión de construcción de líneas como continuas se tendrán menos requerimientos computacionales para la determinación de sus soluciones y un mejor contraste respecto a la solución real del problema, ya que pequeñas variaciones en el vector solución podrán ser percibidas por las métricas utilizadas. Más detalles respecto al modelo de transmisión utilizado se presentan en la sección 2.1.2.3.

### 3.2.2. Curva de duración de la demanda

La metodología de aproximación de la demanda a través de su curva de duración (LDC) ha sido ampliamente utilizada en los diferentes problemas de planificación y operación económica de los sistemas de potencia, por lo cual se ha escogido como un punto base para contrastar las nuevas metodologías propuestas.

El objetivo de la LDC es aproximar la curva de demanda de un sistema en  $L$  bloques, donde cada bloque de demanda define un nivel y duración. Para lograr este objetivo, se resuelve un problema de optimización que divide la curva de carga continua en  $L$  grupos. Esta partición de datos busca minimizar la suma, sobre todos los grupos, de las distancias punto a punto entre el grupo y un “centro”<sup>6</sup>. El resultado de este proceso produce una representación de la demanda discreta que parece una función escalonada.

La ventaja de esta metodología es que en vez de una representación detallada de 8760 horas al año, la demanda puede ser representada solo por  $L < 8760$ . Así, esta reducción en el tamaño del modelo permite expandir el horizonte de tiempo a más años (décadas) y/o utilizar un modelo más detallado de la red o de la operación del sistema. Sin embargo, su desventaja radica en la pérdida de secuencialidad y espacialidad de los datos, la que cobra más importancia en planificaciones de corto plazo.

Relativo al presente trabajo, para la construcción de la LDC se empleará el esquema presentado en la figura 3.2.1, en donde inicialmente se tienen los perfiles de demanda del sistema separados de forma temporal (horas) y espacial (barras). Luego se agrupan espacialmente a través de su demanda neta, perdiendo la información espacial relativa a su localización.

Posteriormente, se realiza el agrupamiento unidimensional, a través de alguna técnica de *clustering*, en este caso k-means, el cual realiza un agrupamiento de datos basado en la minimización de la distancia al centro de los *clusters*, en donde se rescata su índice horario agrupado según *clusters*, para poder asignarlos posteriormente a las respectivas barras desagregadas del sistema.

Así, el tercer paso es desagregar la demanda por barras, a través de sus respectivos índices horarios asignados por *clusters*, para luego, a través del promedio de cada uno,

<sup>5</sup>Las decisiones de inversión solo se toman en el tiempo  $t_0$ , por lo que son fijas para el resto del horizonte temporal.

<sup>6</sup>Al existir diferentes metodologías de agrupamiento de datos o *clustering*, el tipo de centro se escogerá según la técnica empleada. La metodología más comúnmente empleada para este tipo de casos es el método de mínimos cuadrados.

obtener su centroide representativo. De este modo se obtiene finalmente una matriz reducida temporalmente, que si bien conserva la dimensión espacial, ha perdido en el proceso de agrupación información relevante relativa a la localización de la demanda y generación ERV, siendo esta la principal diferencia con los nuevos métodos propuestos, que buscan conservar dicha información en su proceso de agrupación.

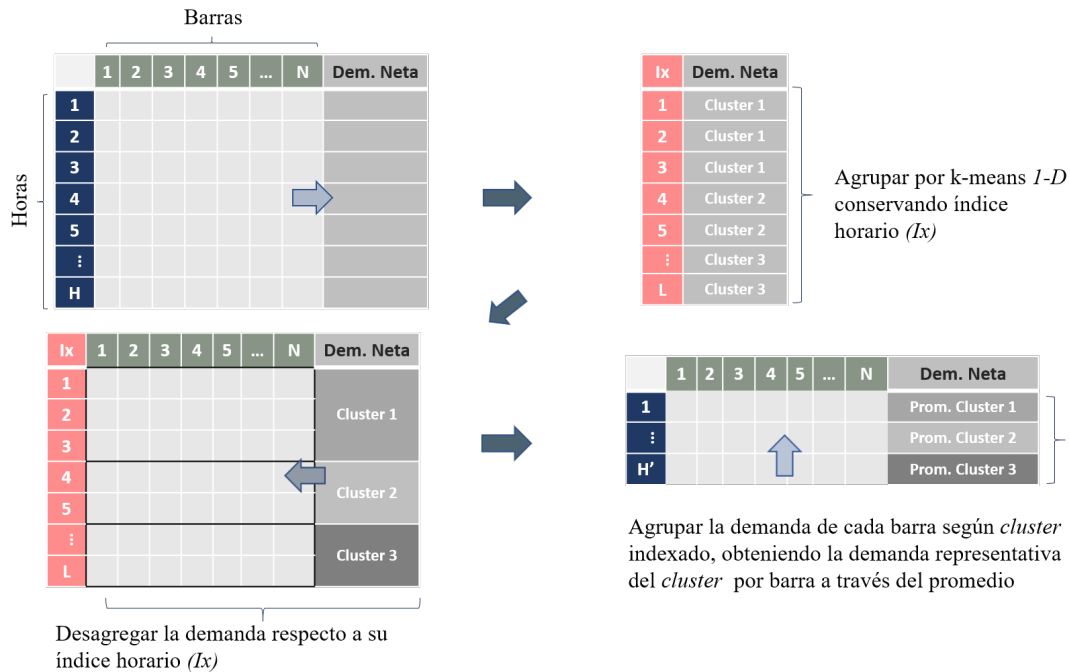


Figura 3.2.1: Esquema de reducción de la LDC.

Una descripción más detallada respecto a esta metodología se presenta en [1] y en el anexo 3.6.1.

### 3.2.3. Técnicas de Agrupamiento de Datos: *Clustering*

Como se mencionó anteriormente, el gran número de variables manejadas en este tipo de problemas conllevan el uso de técnicas avanzadas de manejo de datos e información. En este trabajo de particular interés el *clustering* [91, 2, 82], el cual es una técnica de *Machine Learning* que involucra la agrupación de puntos de datos. Dado un conjunto de datos, es posible utilizar un algoritmo de *clustering* para clasificar cada punto en un grupo específico.

En teoría, los puntos de datos que están en el mismo grupo deberían tener propiedades similares y/o características a fines, mientras que los puntos en diferentes grupos de datos deberán ser altamente diferentes.

A continuación, se presenta un resumen de las técnicas *clustering* empleadas en este trabajo y sus respectivas ventajas y desventajas. La descripción de los algoritmos de *clustering* utilizados en este trabajo se presentan en 3.6.2.

### 3.2.3.1. K-means

El método *k-means clustering* (MacQueen, 1967) agrupa las observaciones en  $K$  *clusters* distintos, donde el número  $K$  lo determina el analista antes de ejecutar del algoritmo. *k-means clustering* encuentra los  $K$  mejores *clusters*, entendiendo como mejor *cluster* aquel cuya varianza interna (*intra-cluster variation*) sea lo más pequeña posible. Se trata por lo tanto de un problema de optimización, en el que se reparten las observaciones en  $K$  *clusters* de forma que la suma de las varianzas internas de todos ellos sea lo menor posible. Para poder solucionar este problema es necesario definir un modo de cuantificar la varianza interna.

Requiere saber de antemano el número de *clusters* en los datos a trabajar. Organiza los datos en torno a los centros de los *clusters*, respecto a alguna medida de distancia. Inicialmente se selecciona empíricamente la cantidad  $K$  de clases/grupos de datos a usar, y se inician al azar sus puntos centrales.

Los datos se clasifican calculando la distancia entre ese punto y el respectivo centro del grupo, y se clasifica en el grupo cuyo centro esté más cercano a él. Luego, en función de los puntos ya clasificados, se recalcula el centro del grupo a través de la media de todos los vectores del grupo. Se itera en el mismo procedimiento hasta que los centros de grupo no cambien significativamente entre iteraciones. El algoritmo entrega soluciones mínimos locales. Este a través de la distancia euclidiana organiza los *clusters* en torno a sus centroides. El algoritmo en el que se basa esta técnica se puede encontrar en [82].

*Algoritmo:*

1. Escoge  $n$  puntos como «centros» de los  $K$  *clusters*.
2. Asocia cada elemento al centro más cercano.
3. Calcula los nuevos «centros» usando los elementos asociados en el paso anterior.
4. Repetir pasos 2-3 hasta convergencia.

*K-means* [82]:

- *Ventajas de k-Means:*

1. Si se tiene un gran número de variables, *k-means* la mayor parte del tiempo es más rápido que los *Hierarchical Clustering Methods*.
2. *K-Means* produce *clusters* más estrictos que los *Hierarchical clustering Methods*.

- *Desventajas de K-Means:*

1. Su valor  $K$  (o número de particiones) es difícil de predecir.
2. Al iniciar aleatoriamente los centros del *cluster* puede producir resultados muy diferentes en diferentes ejecuciones del algoritmo, haciendo que los resultados a veces carezcan de consistencia y lo hace difícil de comparar versus otros *clusters*.

### 3.2.3.2. PAM

El algoritmo PAM, propuesto en 1987 [92, 93], es una metodología heurística para resolver en forma casi óptima el problema de agrupamiento frente a cualquier medida de distancia o disimilitud. Es mucho más robusta que *k-means*, no viéndose afectada fuertemente por ruido u *outliers*, debido a que utiliza directamente la medida de disimilitud en lugar de su cuadrado. Otra diferencia importante respecto a *k-means* es que el representante de cada *cluster* no es su centroide sino su medoide, es decir, el elemento del *cluster* para el cual la disimilitud media respecto a los otros elementos es mínima. El algoritmo en el que se basa esta técnica se puede encontrar en [82].

*Algoritmo:*

#### FASE BUILD

1. Elige el elemento respecto al cual se minimiza la suma de distancias como primer medoide.
2. Toma un elemento no medoide y los elementos que ya son medoides, y asigna los elementos restantes a *clusters* según el medoide más cercano. Calcula la suma de disimilitudes dentro de los *clusters*.
3. Repite el paso dos para todos los elementos no medoides y selecciona aquel que entrega la menor suma de disimilitudes como nuevo medoide.
4. Repite los pasos 2:3 según número de *clusters* a formar

**FASE SWAP:** Una vez terminada la Fase *Build*, prueba mover de *cluster* cada uno de los elementos en forma individual. Si se disminuye la suma de disimilitudes con alguno de estos movimientos se reasigna el elemento.

*PAM (Partitioning Around Medoids)* [91]:

#### ■ *Ventajas de PAM:*

1. Entrega resultados más estables y robustos que k-means
2. Las distancias entre los elementos puede calcularse en forma previa y una sola vez, lo que permite usar cualquier tipo de distancia.
3. Asignación determinista de elementos.
4. Es simple de entender y fácil de implementar.
5. Tiene rápida convergencia en un número fijo de pasos.
6. Es menos sensible a valores atípicos que otros algoritmos de partición.

#### ■ *Desventajas de PAM:*

1. Es más costoso que K-Means dada su complejidad temporal.
2. No funciona bien para grandes cantidades de datos.
3. Sus resultados y su tiempo de simulación dependen de la partición inicial.

### 3.2.3.3. Implementación de k-means y PAM para reducción directa de la demanda

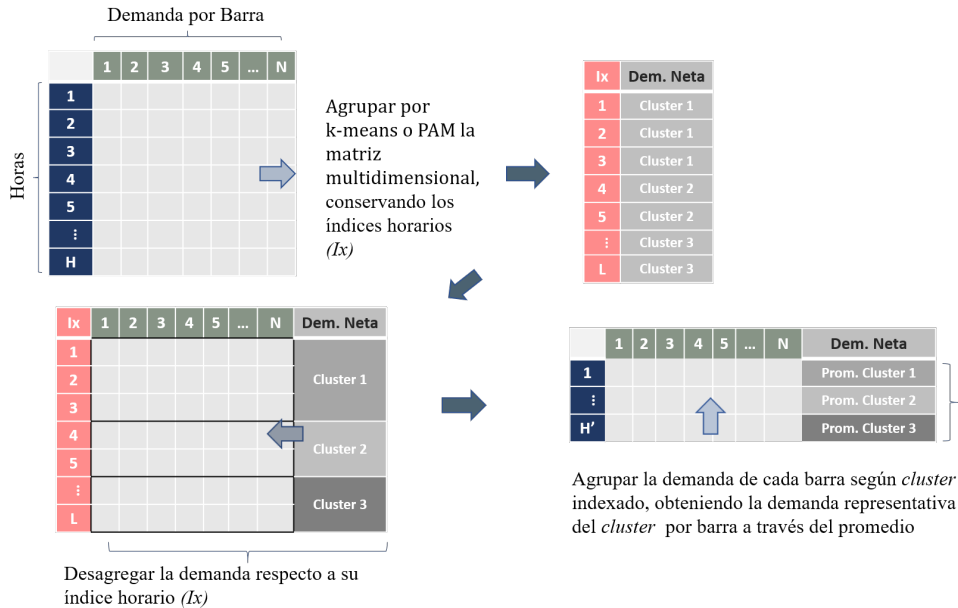


Figura 3.2.2: Esquema de reducción para k-means y PAM.

Relativo al presente trabajo, la reducción de dimensionalidad de la demanda a través de *clustering* se empleará el esquema presentado en la figura 3.2.2, en donde inicialmente se tienen los perfiles de demanda del sistema separados de forma temporal (horas) y espacial (barras). Luego se agrupan multidimensionalmente a través del algoritmo de *clustering* seleccionado, relacionando de esta forma la espacialidad y temporalidad del problema y obteniendo como salida los respectivos índices ( $I_x$ ) de los respectivos grupos de *clusters*.

Posteriormente, a través de los índices horarios que generan los nuevos *clusters* se asignarán los nuevos grupos representados a través de las respectivas demandas netas horarias. Así, el tercer paso es desagregar la demanda por barras, a través de sus respectivos índices horarios asignados, para luego, a través del promedio de cada uno, obtener su centroide representativo. De este modo se obtiene finalmente una matriz reducida temporalmente, pero que conserva la dimensión espacial.

A diferencia del esquema de la LDC aplicado directamente en la demanda, en el caso de los métodos de k-means y PAM la selección de los índices de los *clusters* representativos se realiza observando las variables temporales y espaciales del problema, conservando información espacial relevante en la selección de los respectivos grupos de demanda.

### 3.3. Casos de estudio

En la figura 3.3.1 se presenta un esquema de las metodologías empleadas y la forma de reducción de escenarios utilizada en este trabajo. Esta metodología es empleada en 5 escenarios de generación ERV (0 %, 5 %, 12 %, 26 % y 50 %), en los cuales los escenarios con 5 %, 12 % y 26 % de ERV se subdividen en 3 casos de estudio con diferentes localizaciones de los generadores renovables. Para hacer referencia a estos casos se utiliza la siguiente notación  $ERV_{0,5,12,25,50}^{1,2,3}$  descrita en la sección Abreviaturas.

En la figura 3.3.1 se presentan las divisiones de los diferentes casos de estudio, según nivel de generación renovable, y según localización de estas.

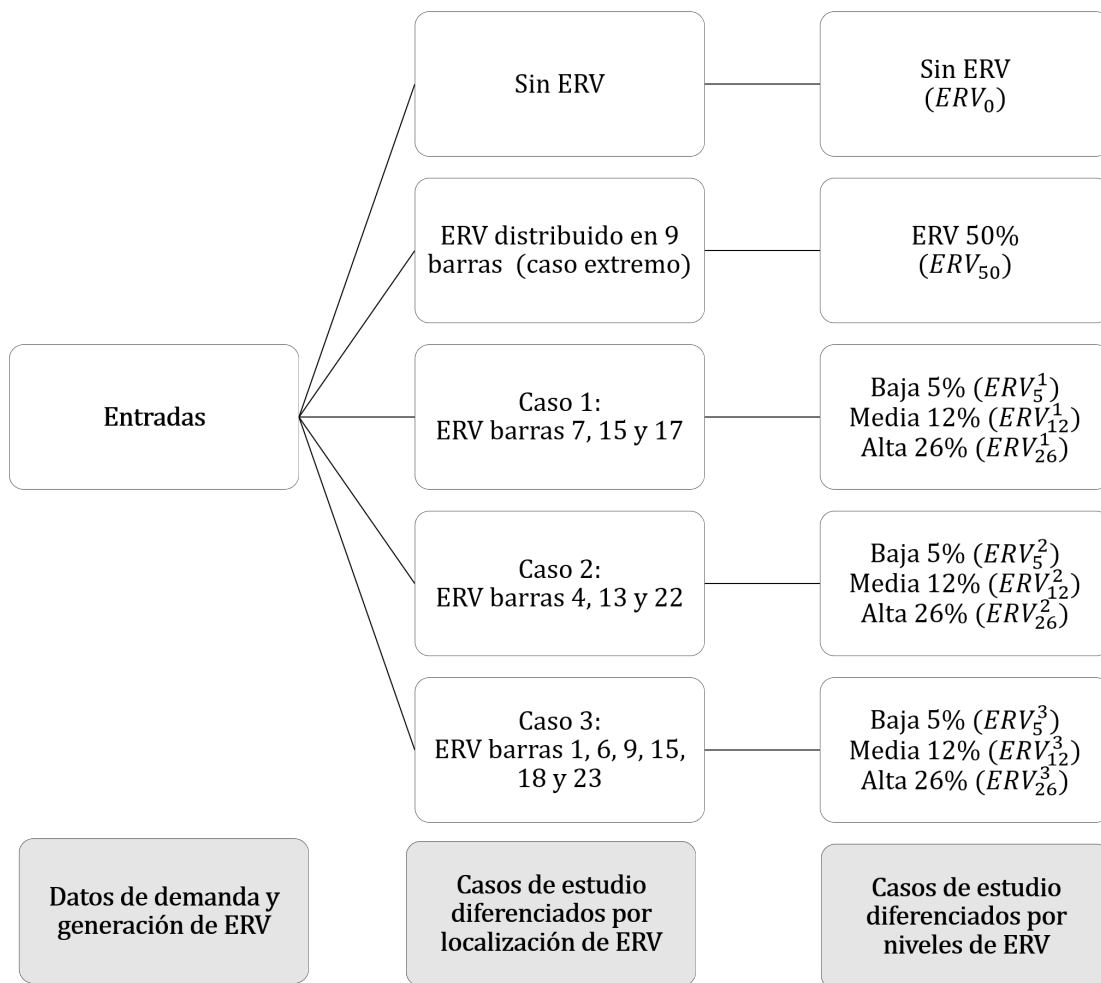


Figura 3.3.1: Subdivisión de casos de estudio por localización y nivel de generación ERV.

En las figuras 3.3.2 y 3.3.3 se presenta una especificación de las metodologías de *clustering* empleadas en cada caso de estudio, en donde se presenta el método base,

con el cual serán contrastados los escenarios reducidos. Inicialmente se determinará la solución relativa a un caso base, en donde se trabajará con los bloques horarios originales (8760 horas), y las soluciones de este caso serán contrastadas posteriormente con los casos de dimensionalidad reducida.

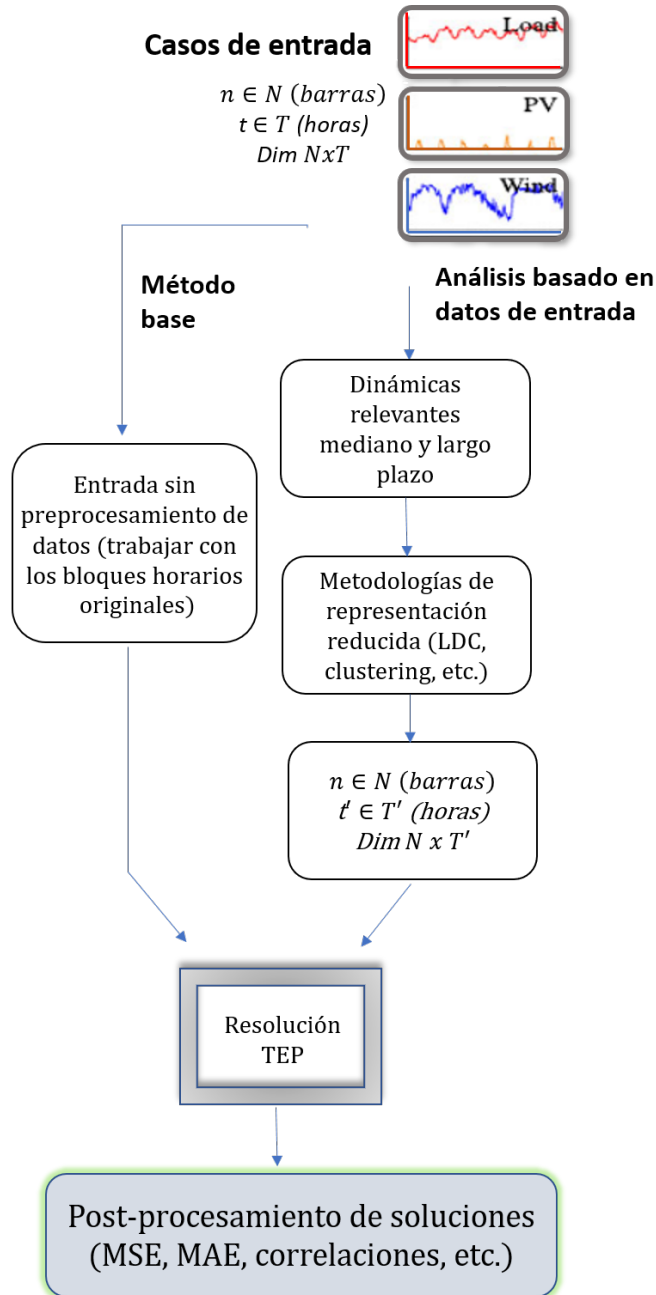


Figura 3.3.2: Esquema del trabajo propuesto.

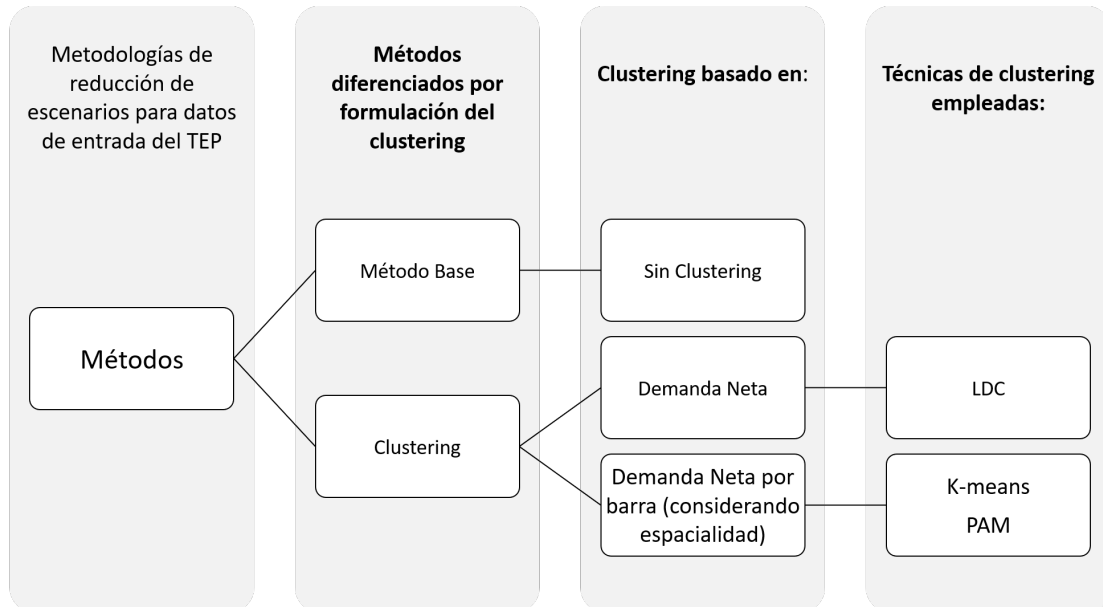


Figura 3.3.3: Esquema del trabajo propuesto.

### 3.3.1. Datos

Se utilizan los datos del sistema RTS-96, área A (24 barras) [64], considerando:

- Generadores: 32
- Barras: 24
- Tramos de líneas: 34
- Nuevos tramos por línea: 3 por línea
- VoLL: 2000 [\$/MW]
- Sbase: 100 [MVA]
- Tasa de descuento: 7 %

Relativo a la demanda y generación convencional, se utilizaron los datos del sistema *RTS-96*, y se realizaron las siguientes modificaciones:

- **Demanda:** Perfiles zona A *RTS-96*, multiplicados por 4.
- **Generación convencional:** Generación máxima sistema *RTS-96* multiplicada por 5.

Estas modificaciones al sistema se realizan debido a que se desea congestionar la red, para que de este modo el problema de planificación de la transmisión posea un mayor número de opciones de construcción, y por ende, tener un mayor número de soluciones

para contrastar. Relativo a los costos de generación del sistema, estos se presentan en la figura 3.3.4d. En la subsección anexa 3.6.3 se presentan los datos de la respectivas líneas y generadores del sistema de estudio.

En la figura 3.3.4 se presenta el sistema de estudio, y las diferentes localizaciones de generadores ERV en estudio.

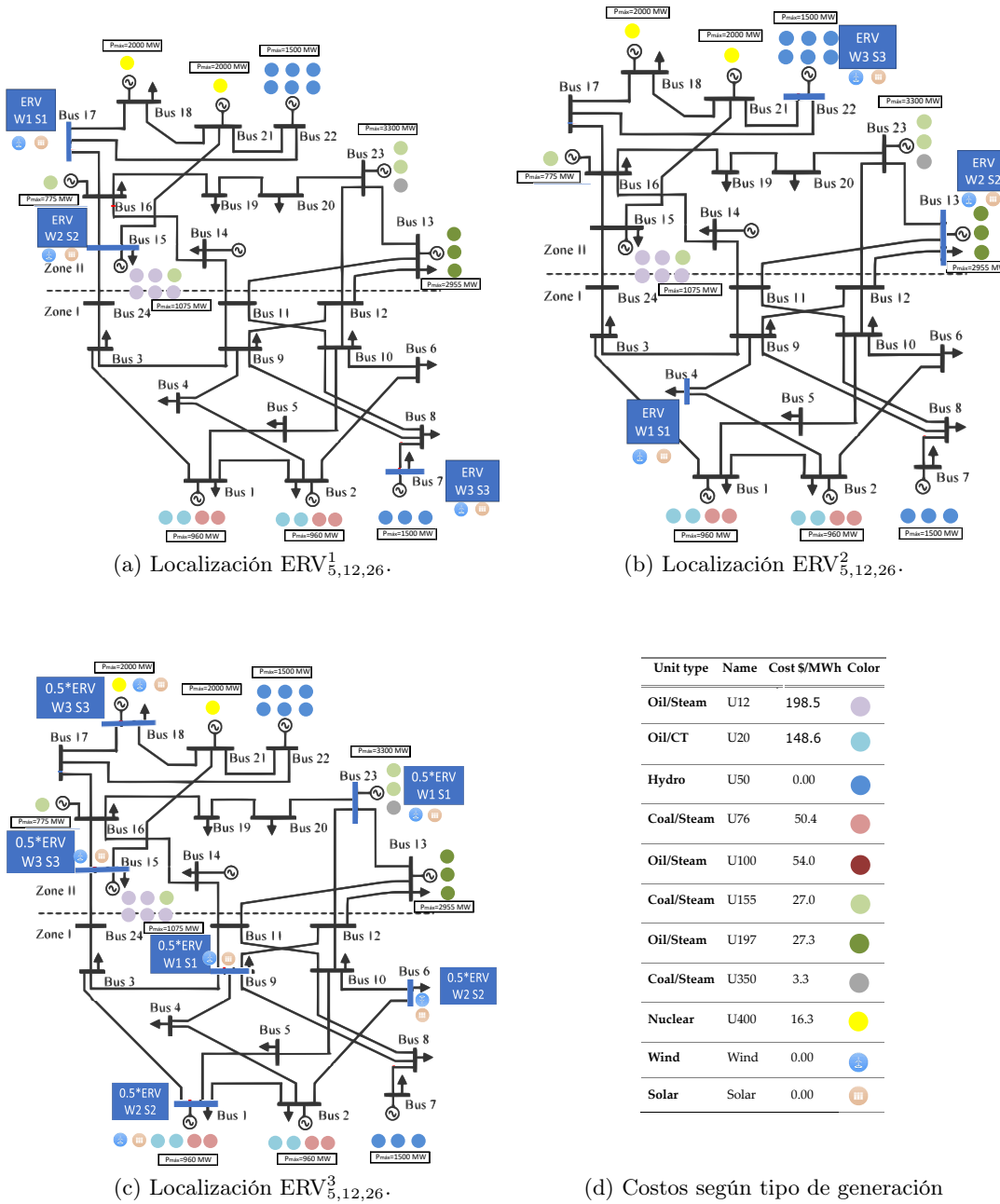


Figura 3.3.4: Localización de generadores ERV en el sistema y costos de operación de centrales. [Fuente: Elaboración propia.]

### 3.3.2. Perfiles de demanda y generación ERV

En la figura 3.3.5 se presentan las correlaciones entre los diferentes perfiles de demanda neta en cada barra según los casos de estudio. La generación renovable se considera restando su generación en la demanda de la barra en la cual está ubicada. De este modo, los perfiles de demanda por barra aumentan su variabilidad espacial y temporal. Adicionalmente, en las figuras 3.3.6 y 3.3.7 se presentan gráficamente los perfiles de demanda promedio a nivel horario para los diferentes casos de estudio.

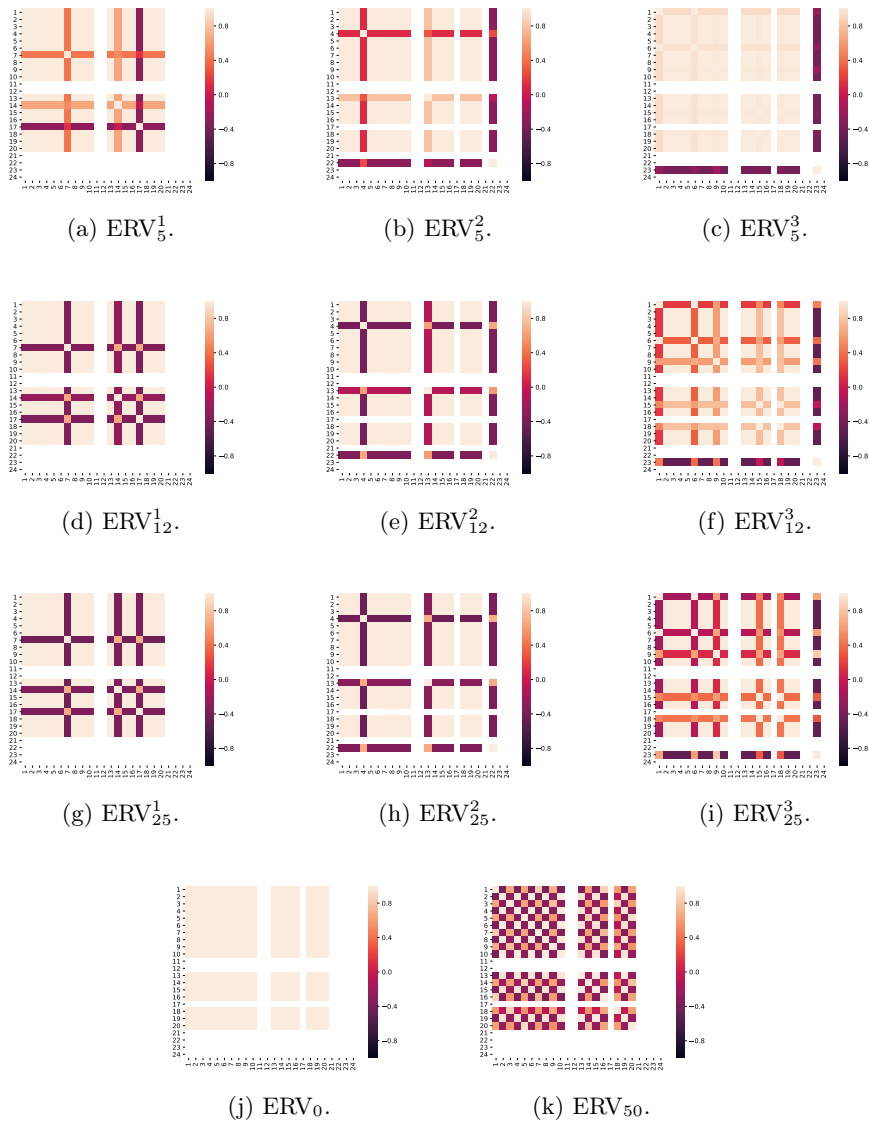
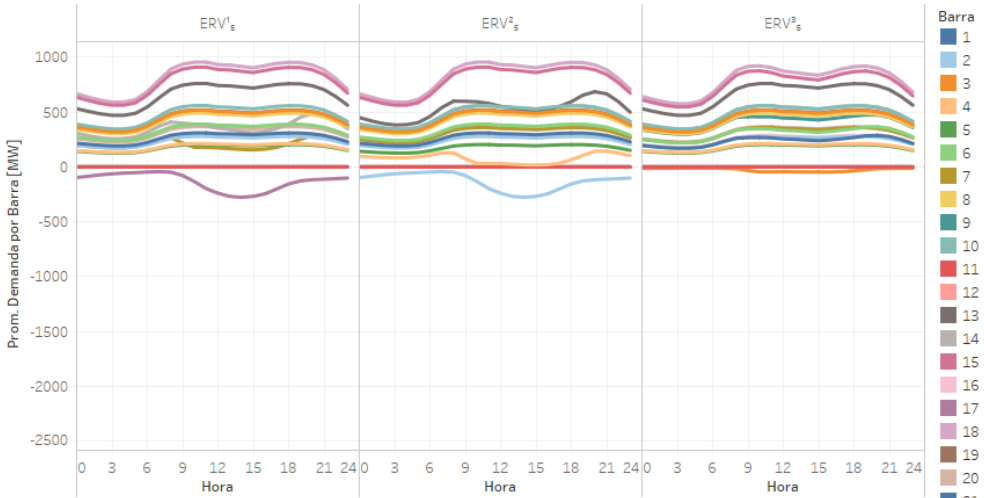
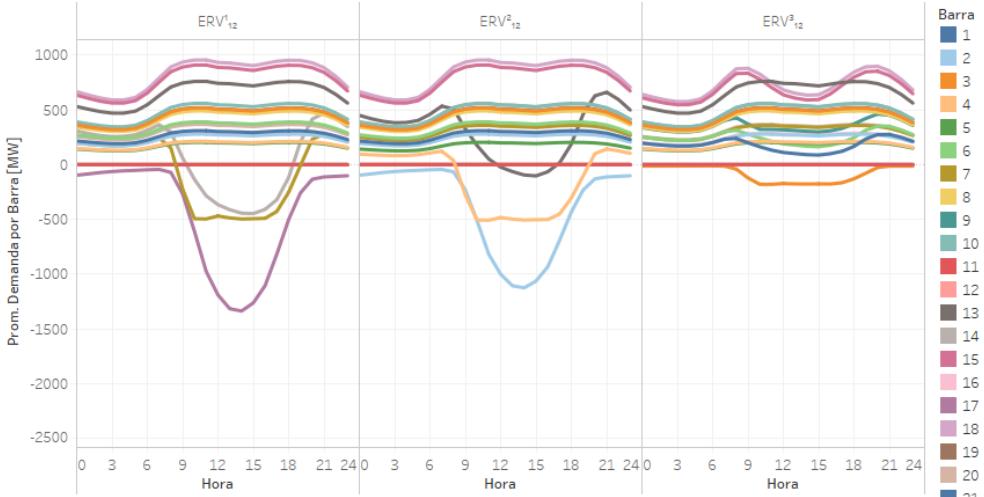


Figura 3.3.5: Gráficas de correlación espacial entre perfiles de demanda neta por barra.

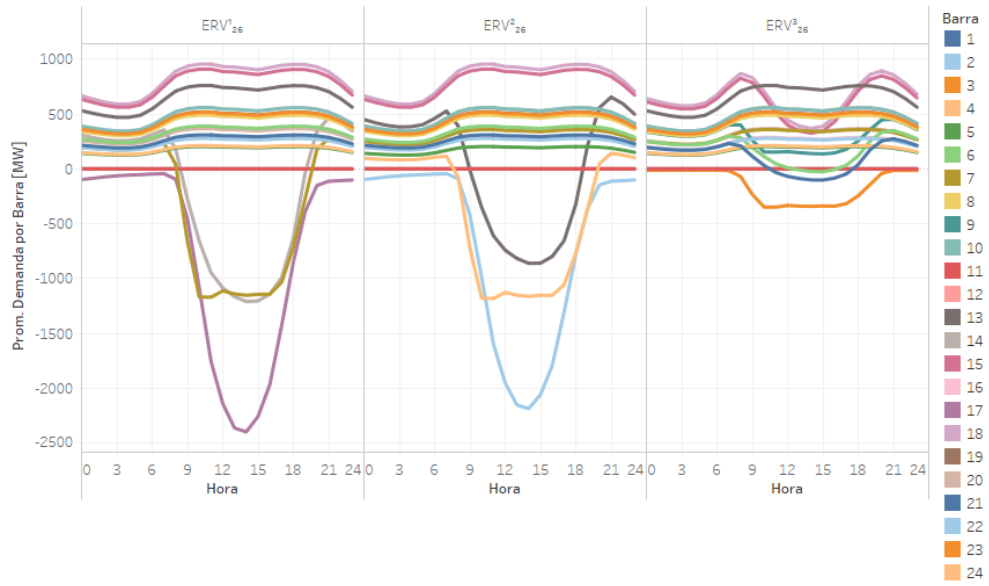


(a)  $ERV_5^{1,2,3}$ .

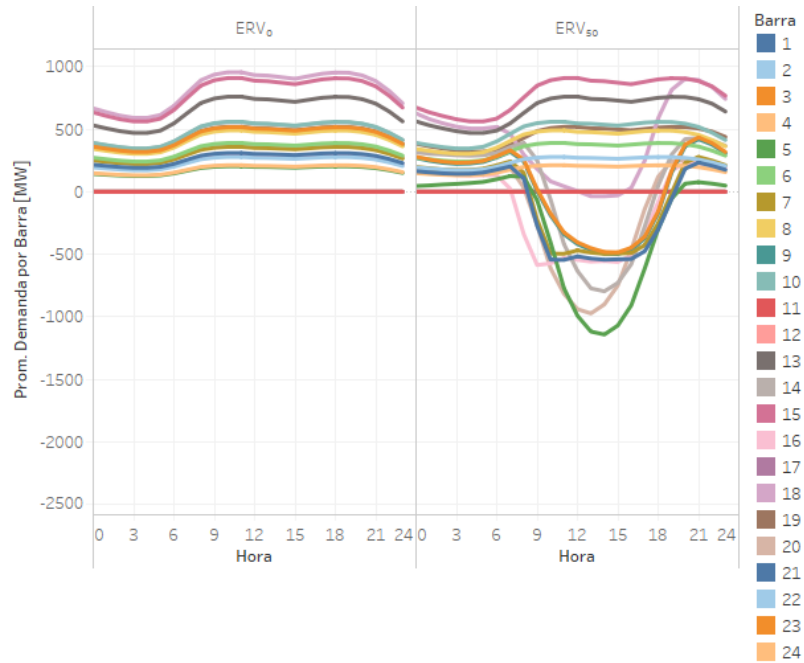


(b)  $ERV_{12}^{1,2,3}$ .

Figura 3.3.6: Gráficas de perfiles de demanda neta promedio por barra.



(a)  $ERV^{1,2,3}_{26}$ .



(b)  $ERV_0$  y  $ERV_{50}$ .

Figura 3.3.7: Gráficas de perfiles de demanda neta promedio por barra.

### 3.3.3. Selección de escenarios de estudio

- **Demanda:** Para el caso base de estudio se utilizan los perfiles de demanda dados por el sistema *RTS-96*. Luego estos son modificados de modo tal que el sistema en estudio requiera de considerable expansión de la transmisión en la planificación. Según esto, se determinó utilizar la demanda multiplicada por 4.
- **Generación convencional:** Para lograr abastecer la mayor demanda planteada en el sistema, se trabaja con los niveles de demanda dados por el sistema *RTS-96* y estos se multiplican por 5, como se observa en la tabla de la sección anexo 3.6.1.
- **Generación ERV:** Se tomaron perfiles de generación de ERV eólica y solar, y estos se adaptaron según los diferentes casos de estudio, de modo tal de proveer diferentes perfiles de demanda en el sistema, cuya variabilidad espacial fuese en aumento y se pudieran observar los fenómenos buscados en el estudio.
- **Ubicación generación ERV:** Para determinar el primer caso de ubicación de las centrales de generación ERV, primero se realiza una pequeña prueba del sistema a través de un flujo de potencia sin límite en las líneas. Además, por medio de esta información se detectan las direcciones promedio de los flujos y sus respectivos gradientes en el sistema, dimensionando también los tiempos en que se supera el límite real de capacidad de cada línea.  
De este modo se observó que las direcciones de los flujos con mayor saturación eran las dadas entre las barras 17-16, 16-14, 24-15, 3-24, 3-9, 9-8 y 7-8. Así, según la dirección de los flujos obtenida se determinó una metodología de localización estratégica de fuentes renovables con el objetivo de congestionar la red, incentivando de este modo la construcción de nuevas líneas. Este consiste básicamente en posicionar la generación renovable en las barras 17, 15 y 7, como se observa en la figura 3.3.4a.  
Respecto al segundo caso de ubicación de renovables, se utilizó la misma que en [89], de modo tal de poder contrastar resultados (ver figura 3.3.4b). Finalmente, el tercer caso de ubicación de renovables fue planteado buscando repartir en un mayor número de barras la generación renovable, eligiendo de este modo 9 barras (ver figura 3.3.4c).
- **Escenarios de estudio:** Dado que el objetivo de este estudio es observar el comportamiento de las diferentes técnicas de *clustering* seleccionadas sobre la demanda, se requerirá además contar con escenarios de demanda lo suficientemente variados para observar los efectos buscados. De este modo, será relevante considerar la correlación espacial entre los perfiles de demanda y sus barras, y de este modo poder analizar la importancia de esta en la selección de una técnica de *clustering* adecuada. Se considera además que la integración de energías renovables en el sistema será trabajada como una *demanda negativa*<sup>7</sup>. Así, durante el desarrollo de esta tesis se definieron inicialmente 4 perfiles de integración renovable en las 3 barras mencionadas anteriormente, estos son:

<sup>7</sup>Existen diversas formas de tratar esta información, pero resulta más práctico para el *clustering* esta simplificación.

- **Caso base:** Caso base del sistema RTS-96 que no considera integración ERV (ERV 0 %).
- **ERV baja:** Se trabajan con perfiles eólicos y solares y se integran a las tres barras (ERV 5 %).
- **ERV media:** En base al perfil de demanda baja, se aumenta la generación renovable solar, tal que sea 2.5 veces la inicial (ERV 12 %).
- **ERV alta:** En base al perfil de demanda baja, se aumenta la generación renovable solar, tal que sea 5 veces la inicial (ERV 26 %).
- **ERV muy alta:** En base al perfil de demanda baja, se aumenta la generación renovable solar, tal que sea 10 veces la inicial (ERV 50 %).

### 3.3.3.1. Métricas de evaluación de resultados

El análisis de resultados se basa en métricas comparativas de las salidas del problema de planificación. Para ello, inicialmente se realiza la planificación del problema real, con las 8760 horas del año, cuya solución es la referencia para comparar las nuevas soluciones con las técnicas empleadas.

Las métricas de evaluación de resultados utilizadas son:

- Variación respecto a la cantidad de nuevas líneas construidas: Se compara el total de nuevas líneas construidas entre el caso con 8760 horas y el caso de estudio.
- Variación en la Función Objetivo (FO): Se compara el valor final de la función objetivo entre el caso con 8760 horas y el caso de estudio. La función objetivo entrega en estos casos la suma de los costos de inversión en líneas y los costos de operación del sistema.
- MSE y MAE: Error cuadrático medio y error absoluto medio [94]. Estas métricas se utilizan para contrastar los vectores de nuevas líneas construidas. De este modo, se busca contrastar la ubicación de las nuevas líneas construidas para los diferentes casos.
- Correlación: Al igual que en el caso anterior, a través de la correlación entre los vectores solución de las nuevas líneas construidas es posible obtener una idea de cuan similares son las soluciones a comparar. En este caso un valor muy cercano a 1 es una correlación perfecta.
- Variación en la Función Objetivo Recalculada (FO recalculada): Una vez obtenida la solución del TEP con los respectivos *clusters*, esta puede dividirse en dos partes: la solución relativa a la expansión de la transmisión (líneas construidas y costos de expansión) y la solución de operación (costo de despacho de centrales y energía no servida).

Dado que el objetivo principal del trabajo es lograr soluciones de expansión de transmisión lo más próximas a la solución real, esta solución obtenida se mantendrá y se modificará la solución de operación, para obtener los respectivos costos de

operación para las 8760 horas del año. Para ello, se tomarán las soluciones de expansión de transmisión y se impondrán como condiciones iniciales en el OPF para las 8760, con la demanda original del sistema (sin *clusters*). De este modo, la solución de operación será más realista, ya que considerará la operación real del sistema, con lo que el valor de la función objetivo recalculada podrá ser una buena métrica de comparación de soluciones.

### 3.4. Resultados numéricos

Las simulaciones del TEP se realizan para 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 24, 30, 40, 60, 73 y 120 *clusters*, en 11 casos de estudio. El pre-procesamiento de los datos de entrada y la generación de los respectivos *clusters* se realizó a través de R [95], y la respectiva simulación de el problema de expansión de la transmisión fue desarrollado en *PYOMO (Python)* [96, 97], a través de una formulación no lineal, por lo que el *solver* utilizado para la resolución del problema fue *Ipopt* [47]. El computador utilizado para el trabajo en R y la simulación de los casos de estudios con reducción dimensional de la demanda fue un Intel (®)Xeon (®)X5570, 2 procesadores de 2.30 GHz y 64 Gb de RAM.

Relativo al tiempo requerido para la simulación en *PYOMO* con las 8760 horas, este no es del todo comparable con los casos reducidos, debido a que por su gran requerimiento computacional se debió utilizar un servidor externo y diferentes configuraciones para su simulación. Para ello se utilizó el *solver Ipopt + HSL MA86* [98], y el servidor externo utilizado fue un nodo *cluster quartz*<sup>8</sup> (Intel Xeon E5-2695, 344,064 Gb de RAM, CPU Memory/Node de 128 Gb), y bajo estas condiciones los tiempos de simulación promedio fueron de media hora<sup>9</sup>.

En las tablas 3.4.1 y 3.4.2 se representan los tiempos de simulación para dos casos de estudio, de modo tal de comparar el recurso computacional necesario para el pre-procesamiento de la data y la diferencia de tiempo obtenida al emplear las metodologías propuestas. Se aprecia que los pre-procesamientos de datos en R para LDC y k-means son del orden de las décimas de segundo, mientras que el tiempo de PAM crece de manera exponencial con el número de *clusters*.

<sup>8</sup><https://hpc.llnl.gov/hardware/platforms/Quartz>

<sup>9</sup>Esta simulación fue posible gracias al *Lawrence Livermore National Laboratory* de Estados Unidos de América

Tabla 3.4.1: Tiempos de simulación para algunos cluster del caso  $ERV_0$ 

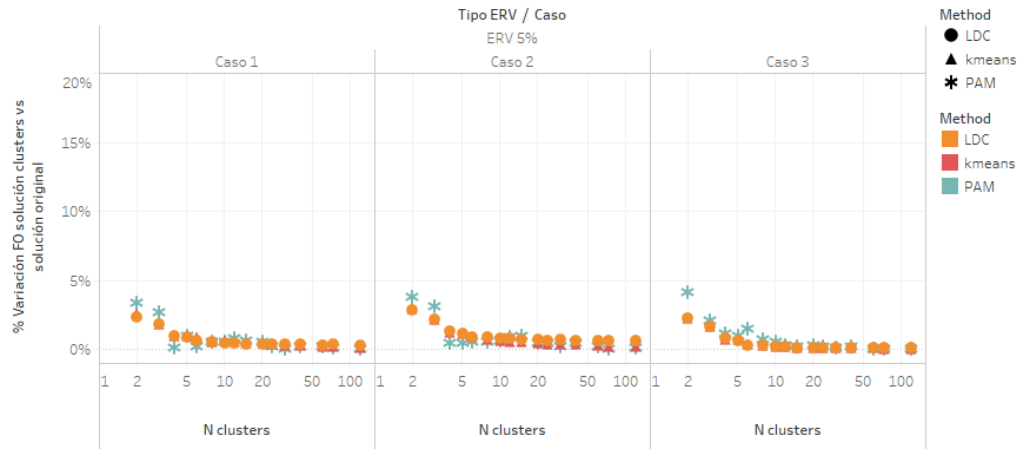
Cluster	Tiempo clusterización (en $R$ )			Tiempo TEP (resolución $PYOMO$ )
	LDC	k-means	PAM	Rango [s]
1	0.1	0.0	3.7	0.2
10	0.1	0.1	903.7	1.8-2.0
20	0.1	0.1	3465.4	3.4-4.3
40	0.1	0.1	6294.1	7.9-9.0
60	0.1	0.1	8425.5	12.3-13.6
120	0.1	0.1	32890.3	26.8-29.7

Tabla 3.4.2: Tiempos de simulación para algunos clusters del caso  $ERV_{50}$ 

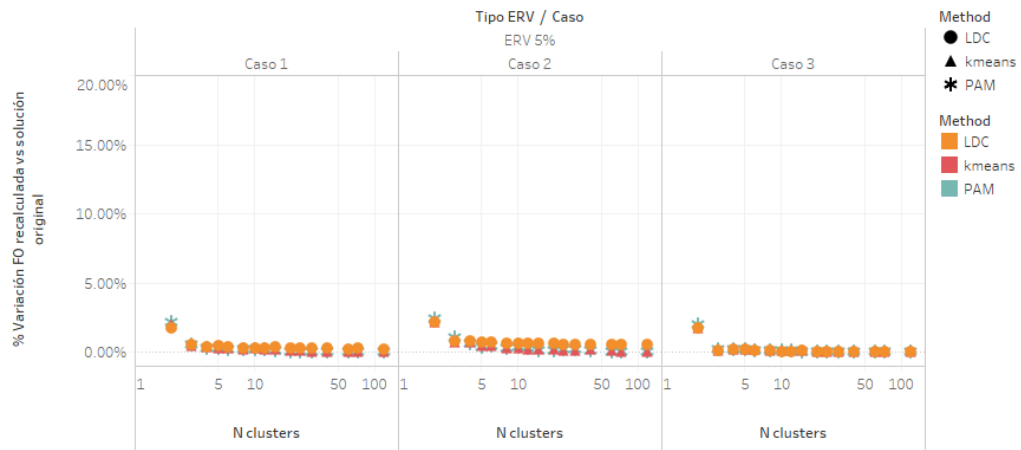
Cluster	Tiempo clusterización (en $R$ )			Tiempo TEP (resolución $PYOMO$ )
	LDC	k-means	PAM	Rango [s]
1	0.1	0.0	3.91	0.1-0.2
10	0.1	0.1	218.7	1.2-1.6
20	0.1	0.1	813.5	2.6-4.0
40	0.1	0.2	2534.9	6.2-7.3
60	0.1	0.4	5028.1	10.4-16.3
120	0.1	0.5	19843.2	26.3-29.4

Tabla 3.4.3: Resultados para 120 *clusters*

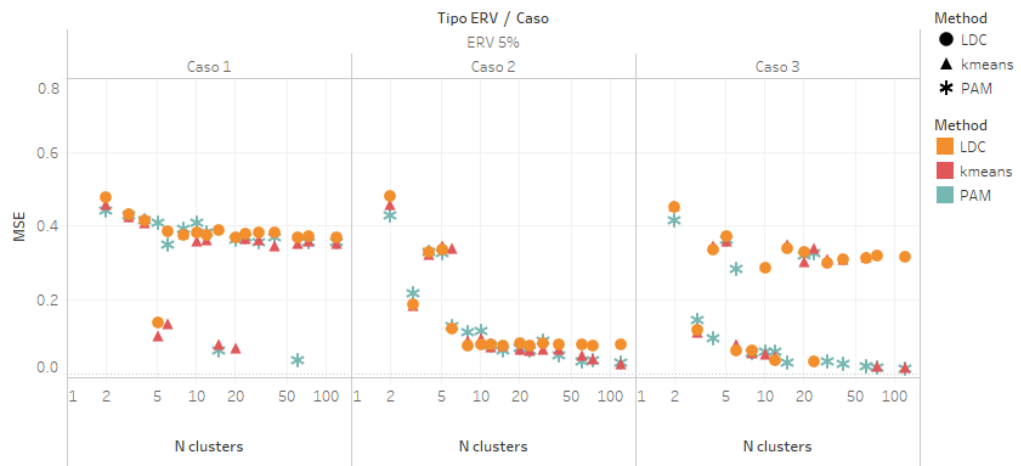
N° Clusters	% ERV	Caso	Método	% Error Nuevas líneas	MSE	MAE	Correlación	% Error FO	% Error FO Recalculada
120	0	-	k-means	0.095 %	0.012	0.006	1.000	0.002 %	0.001 %
			PAM	0.035 %	0.006	0.003	1.000	0.002 %	0.000 %
			LDC	0.050 %	0.004	0.002	1.000	0.003 %	0.000 %
	Caso 1	k-means	0.828 %	0.352	0.137	0.943	0.132 %	0.010 %	
		PAM	0.586 %	0.357	0.143	0.941	0.038 %	0.016 %	
		LDC	2.667 %	0.370	0.163	0.936	0.311 %	0.220 %	
	5	Caso 2	k-means	0.423 %	0.025	0.011	1.000	0.155 %	0.026 %
			PAM	0.408 %	0.031	0.016	1.000	0.138 %	0.025 %
			LDC	2.655 %	0.078	0.034	0.998	0.618 %	0.548 %
	Caso 3	k-means	0.040 %	0.017	0.010	1.000	0.038 %	0.006 %	
		PAM	0.193 %	0.012	0.006	1.000	0.019 %	0.007 %	
		LDC	0.703 %	0.315	0.128	0.957	0.081 %	0.044 %	
	Caso 1	k-means	0.266 %	0.026	0.013	1.000	0.838 %	0.018 %	
		PAM	0.482 %	0.057	0.026	0.999	0.549 %	0.023 %	
		LDC	13.760 %	0.391	0.216	0.944	2.447 %	3.307 %	
	12	Caso 2	k-means	0.698 %	0.040	0.025	0.999	0.624 %	0.030 %
			PAM	0.071 %	0.040	0.024	0.999	0.153 %	0.021 %
			LDC	12.426 %	0.373	0.185	0.948	3.407 %	3.988 %
	Caso 3	k-means	0.344 %	0.023	0.013	1.000	0.158 %	0.013 %	
		PAM	0.098 %	0.015	0.009	1.000	0.184 %	0.005 %	
		LDC	3.716 %	0.111	0.052	0.995	0.638 %	0.594 %	
	Caso 1	k-means	0.441 %	0.061	0.036	0.998	1.618 %	0.109 %	
		PAM	0.410 %	0.058	0.034	0.999	0.896 %	0.079 %	
		LDC	17.622 %	0.353	0.220	0.963	5.764 %	4.243 %	
	25	Caso 2	k-means	1.605 %	0.041	0.022	0.999	1.334 %	0.040 %
			PAM	1.200 %	0.072	0.041	0.998	0.654 %	0.031 %
			LDC	15.108 %	0.405	0.283	0.942	6.155 %	2.389 %
Caso 3	k-means	0.165 %	0.040	0.022	0.999	0.342 %	0.034 %		
	PAM	0.286 %	0.015	0.007	1.000	0.472 %	0.009 %		
	LDC	6.124 %	0.193	0.091	0.986	1.461 %	1.418 %		
50	-	k-means	6.892 %	0.165	0.094	0.981	2.656 %	0.499 %	
		PAM	6.319 %	0.110	0.077	0.993	1.716 %	0.262 %	
		LDC	21.606 %	0.322	0.188	0.944	4.824 %	2.274 %	



(a) Error % FO

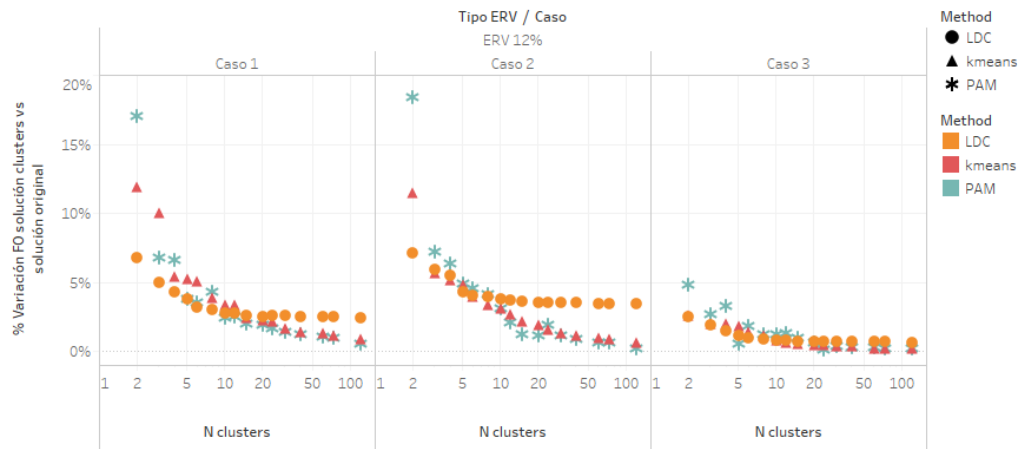


(b) Error % FO recalculada

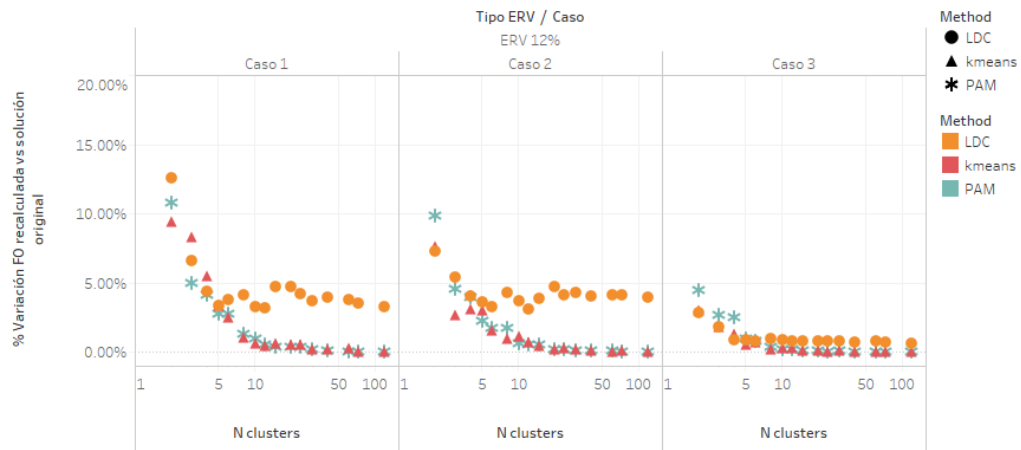


(c) MSE

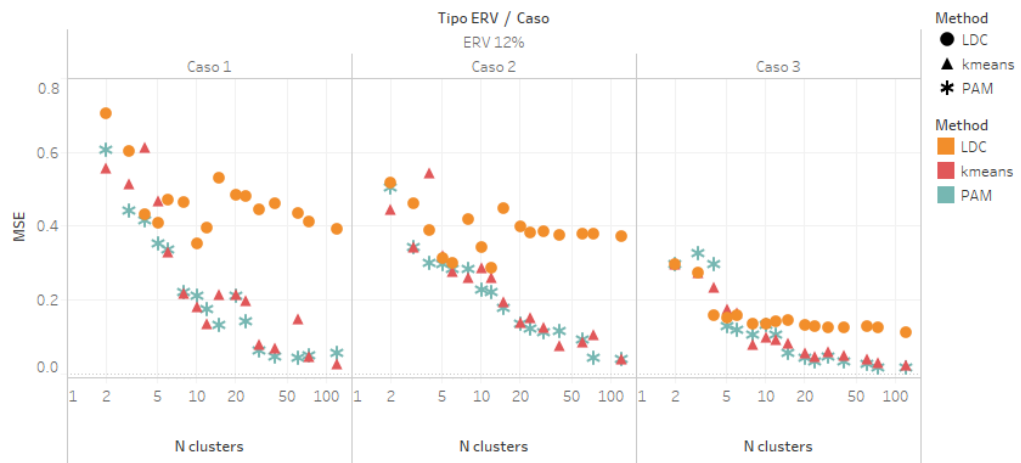
Figura 3.4.1: Métricas de evaluación de casos de estudio para  $ERV_5^1$ ,  $ERV_5^2$  y  $ERV_5^3$ .



(a) Error % FO

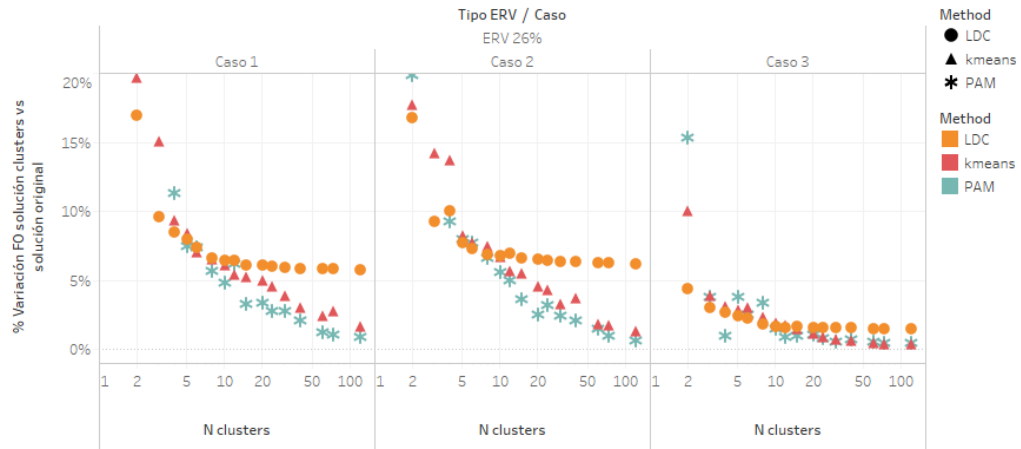


(b) Error % FO recalculada

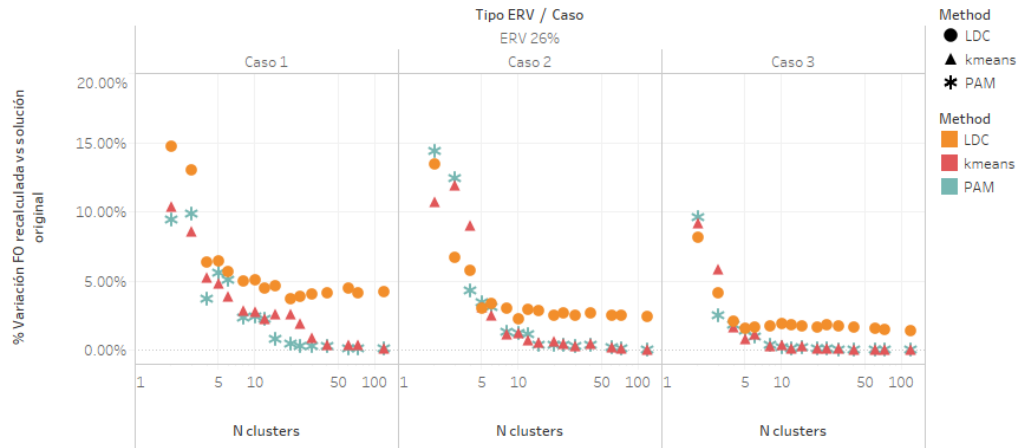


(c) MSE

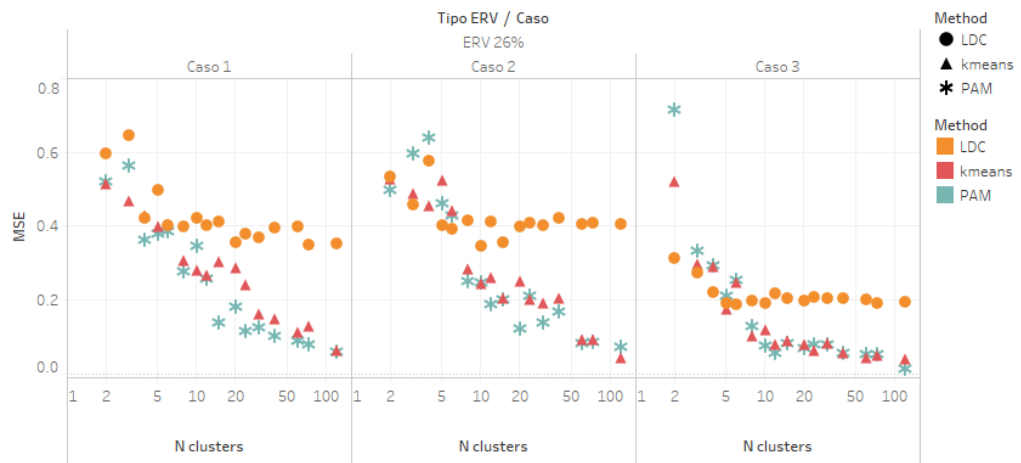
Figura 3.4.2: Métricas de evaluación de casos de estudio para  $ERV_{12}^1$ ,  $ERV_{12}^2$  y  $ERV_{12}^3$ .



(a) Error % FO



(b) Error % FO recalculada



(c) MSE

Figura 3.4.3: Métricas de evaluación de casos de estudio para  $ERV_{26}^1$ ,  $ERV_{26}^2$  y  $ERV_{26}^3$ .

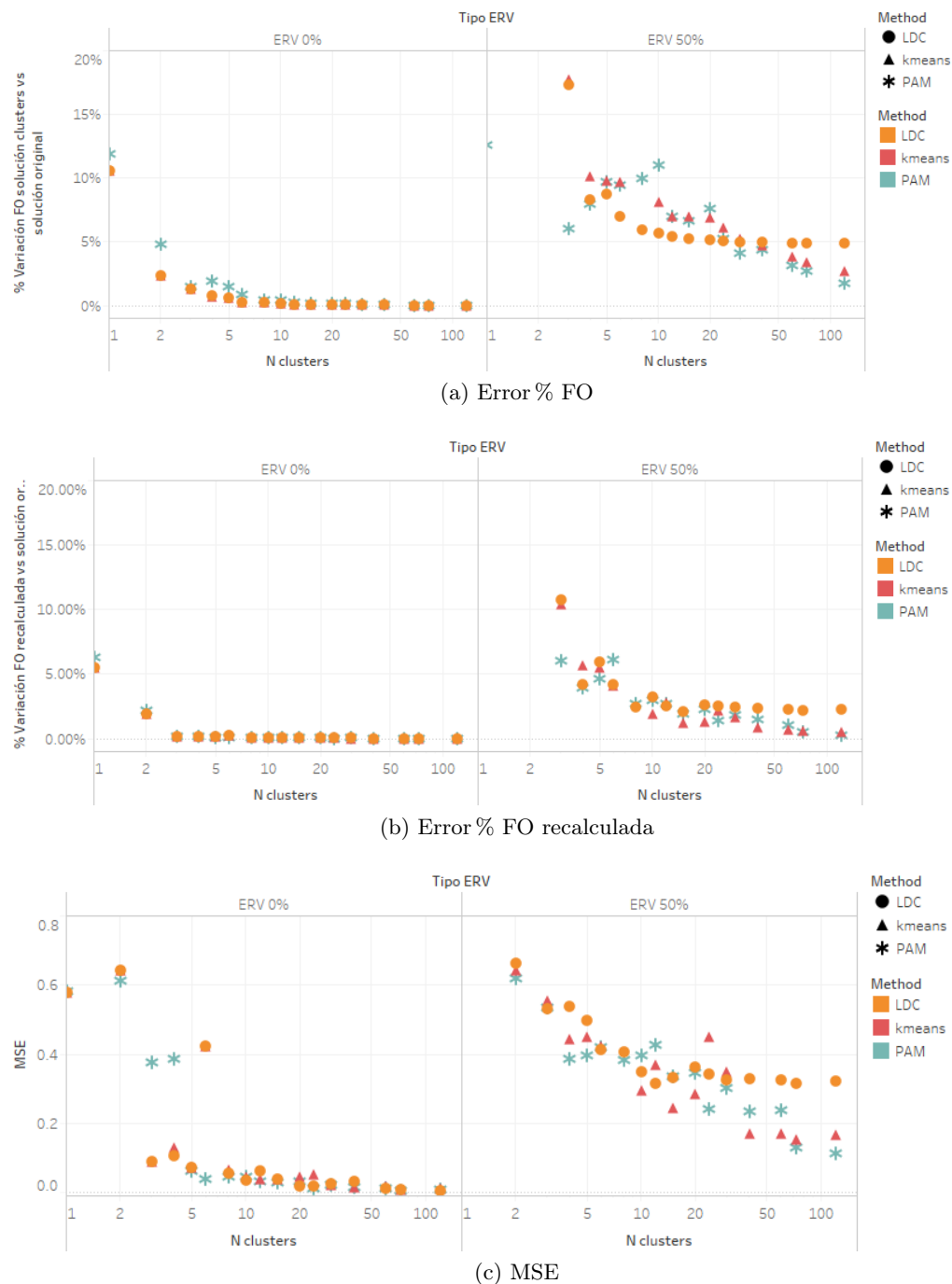


Figura 3.4.4: Métricas de evaluación de casos de estudio para  $ERV_0$  y  $ERV_{50}$ .

En la tabla 3.4.3 se presentan las diferentes métricas de análisis propuestas para el caso de 120 clusters y los diferentes casos de estudio. Además, en las gráficas 3.4.1, 3.4.2, 3.4.3 y 3.4.4 se presentan los indicadores de variación porcentual respecto a la función objetivo, variación respecto a la función objetivo recalculada y el MSE. Los resultados

respecto a variación respecto a las nuevas líneas construidas, MAE y correlación entre vectores de nuevas líneas construidas se anexan en la sección 3.6.4.

Los resultados obtenidos son consistentes que a medida que crece el número de *clusters* la solución cada vez se aproxima de mejor manera a la solución original (8760 horas x 24 barras). Sin embargo, el número ideal de *clusters* para una buena representación depende fuertemente de la variabilidad temporal y espacial de la demanda.

En la figura 3.4.4 se observa que dada una demanda altamente correlacionada entre barras (como en el caso  $ERV_0$ ), los métodos LDC, k-means y PAM entregan resultados similares y con errores bajos. Al contrario, para el caso  $ERV_{50}$  los resultados son muy diferentes, siendo el método basado en la LDC el con peor desempeño.

Respecto al caso  $ERV_{50}$ , para apreciar la calidad de la solución final se debe observar el conjunto de métricas totales, dado que en este caso el método de LDC presenta un buen comportamiento respecto a la función objetivo, pero no así respecto a sus nuevas líneas construidas y su MSE. En este caso, al aumentar el número de *clusters* se aprecia de mejor forma la calidad de las soluciones finales, siendo para 120 *clusters* mejor el método PAM.

Relativo a los casos de estudio con diferente localización espacial de las fuentes renovables ( $ERV_{5,12,26}^{1,2,3}$ ) en las figuras 3.4.1, 3.4.2 y 3.4.3, es de destacar que el comportamiento de las metodologías de reducción de escenarios mejora a medida que la generación se distribuye más alrededor del sistema. Es decir, bajo el mismo nivel de penetración renovable, es preferible la aplicación de las metodologías de reducción de escenarios en casos de generación renovable mayormente distribuida, como se observa en el  $ERV_{5,12,25}^3$ .

Para casos en que se tiene una alta concentración de renovables en algunas barras, o un nivel de generación renovable alto en el sistema en general, se aprecia en las figuras 3.4.1 y 3.4.2 que la mejor reducción de demanda se logra a través del *clustering* con PAM, cuya precisión mejora con el número de *clusters*.

En las figuras 3.4.1 se aprecia el comportamiento del caso  $ERV_5^{1,2,3}$ , en donde se tiene un caso con relativamente poca variabilidad espacial, y por tanto errores de FO y FO recalculada muy bajos. De esta forma se vuelve difícil el dimensionar qué tan buena es la solución obtenida, siendo en este caso el MSE la métrica de comparación adecuada, ya que permite comparar los vectores solución de nuevas líneas construidas y concluir respecto a la representatividad que entregan las reducciones utilizadas en el TEP. De este modo, es de destacar la figura 3.4.1c, en donde se observan diferencias considerables respecto al MSE respecto a las diferentes localizaciones de la demanda.

Un análisis similar es posible de realizar respecto al caso  $ERV_{12}^{1,2,3}$  en la figura 3.4.2. Aquí se comienza a hacer más notoria la falencia de la metodología LDC, ya que en los casos 1 y 2, en donde se tiene la generación renovable concentrada en 3 barras, el desempeño de la metodología es malo, tendiendo a un error constante de un 5% en la FO y FO recalculada. Respecto al método k-means y PAM, es difícil concluir si hay alguno que se comporte mucho mejor que otro, ya que su tendencia es bastante similar.

En el caso  $ERV_{26}^{1,2,3}$ , presentado en 3.4.3, los resultados son similares a los anteriores. Sin embargo, ya se empieza a notar un mal desempeño de LDC en el caso 3, especialmente respecto a la FO recalculada.

En la tabla 3.4.3 se presentan los resultados para 120 *clusters*, en donde se aprecian

con mayor detalle las respectivas métricas de comparación en cada caso. Para  $ERV_0$  con 120 *clusters*, el método LDC se comporta ligeramente mejor que los otros en términos de nuevas líneas construidas y MSE. Sin embargo, para  $ERV_{50}$  se logran visibilizar mejor los beneficios de PAM y k-means, siendo las métricas del método LDC las con peor desempeño, triplicando al resto de los casos en cuanto al error en la función objetivo.

### 3.5. Conclusiones

Debido a la dificultad computacional que implica la resolución del problema de expansión de la transmisión, se proponen metodologías alternativas para realizar una reducción de variables de entrada del sistema, específicamente respecto a la demanda. Además, en base al rápido aumento de la penetración ERV en los SEP, se decidió incorporar esta en los respectivos casos de estudio, y por ende, se agrega una mayor variabilidad espacial y temporal de los perfiles de demanda neta.

En función de lo anterior, se contrasta la comúnmente empleada metodología de discretización de la LDC con metodologías de *clustering* propuestas basadas en K-means y PAM, las cuales se contrastan a través de escenarios de estudio con diferentes niveles de penetración renovable en diferentes localizaciones del sistema.

Las principales conclusiones de este trabajo son:

- \* **Número de clusters:** Se observa que a medida que crece el número de *clusters* la solución cada vez se aproxima de mejor manera a la solución original (8760 horas x 24 barras). Sin embargo, el número ideal de *clusters* para una buena representación depende fuertemente de la variabilidad temporal y espacial de la demanda, siendo suficiente con 10 *clusters* para casos con baja variabilidad, aumentando progresivamente el número a medida que se integran renovables en el sistema.
- \* **Diversidad espacial de ERV:** Los resultados numéricos indican que es posible obtener ganancias significativas en precisión y tiempo de simulación mediante el uso de k-means y PAM, especialmente en los casos con una alta variabilidad espacial en los perfiles de demanda neta por barra. Sin embargo, entre los métodos de *clustering* propuestos no hay evidencia decisiva para preferir uno sobre el otro. Además, en general los métodos de reducción temporal de la demanda suele comportarse mejor para aquellos casos en que la generación renovable está distribuida en el sistema que en los casos con altas concentraciones de renovables en pocas barras de este.
- \* **Diversidad de niveles de ERV:** Relativo a los niveles de penetración renovable del sistema, para los casos con bajo nivel de generación ERV se tiene un alto nivel de correlación espacial entre sus demandas por barras. En estos casos la metodología LDC se vuelve competitiva dado que al no ser tan relevante la espacialidad del sistema, el uso de la demanda neta discretizada puede rescatar una gran cantidad de información relevante del caso de estudio. Sin embargo, al aumentar los niveles de generación ERV en el sistema los métodos propuestos se vuelven más competitivos, siendo evidentemente mejores que la LDC.

- \* **Comparación de requerimientos computacionales:** Respecto a los métodos basados en k-means y PAM, si bien, se obtienen resultados ligeramente mejores a través de PAM, el tiempo que requiere k-means para realizar su proceso de *clustering* es mucho menor, lo cual le entrega una ventaja en la aplicación práctica.
- \* **Métricas de análisis de resultados:** Respecto a la métricas utilizadas para el análisis de resultados, se propuso utilizar una “FO recalculada”, cuyo análisis permite comparar de una forma realista las decisiones de construcción de nuevas líneas en el TEP, dado que solo el análisis de la FO obtenida con los respectivos *clusters* introduce errores de precisión respecto a los costos de operación del sistema, dado que los pesos de los bloques de demanda calculados determinan todo un periodo de operación del sistema, afectando directamente al valor final de la función objetivo. El uso de esta métrica permitió comparar de mejor manera las soluciones obtenidas.

Futuras extensiones de este trabajo pueden incluir nuevas formulaciones del modelo de expansión de transmisión, con la incorporación de restricciones inter-temporales en el problema y posibles aplicaciones a metodologías de selección de días típicos.

## 3.6. Anexos

### 3.6.1. Construcción de bloques de demanda [1].

La construcción de bloques de demanda se realiza usualmente utilizando el método de mínimos cuadrados sobre la demanda horaria total del sistema, es decir, se construyen los bloques de modo tal que las diferencias cuadráticas entre la curva de duración real y su aproximación por tramos sean mínimas. El problema de construcción los bloques se puede formular según 3.6.1.

*Variables del problema*

$J$	Conjunto de bloques de demanda que aproximan la curva de duración
$I_j$	Conjunto de horas en que la demanda real es aproximada por el bloque de demanda $j$
$d_i$	Demanda total del sistema durante la hora $i$
$d_j^{block}$	Demanda en el bloque de demanda

$$\text{mín} \quad \sum_{j \in J} \sum_{i \in I_j} (d_i - d_j^{block})^2 \quad (3.6.1)$$

El problema presentado es factible de resolver computacionalmente, y es bastante sencillo de aproximar, por lo que su formulación se ha logrado masificar. Para ello se requiere definir los conjuntos  $I_j$  mediante alguna técnica, y por propiedades de los mínimos cuadrados se sabe que las demandas  $d_j^{block}$  que minimizan la función objetivo con los conjuntos  $I_j$  definidos, corresponden a los promedios de las demandas  $d_i$  en cada conjunto, es decir, pueden calcularse a través de la relación 3.6.2, donde  $N_{I_j}$  corresponde al número de elementos en el conjunto  $I_j$

$$d_j^{block} = \frac{1}{N_{I_j}} \sum_{i \in I_j} (d_i) \quad (3.6.2)$$

Posteriormente, según el requerimiento del problema, es posible tratar el resultado de distintas maneras. Así, en la aplicación directa al problema de expansión de la transmisión, se requiere de necesariamente el conocimiento de los perfiles de demanda en cada barra del sistema, para lo cual, en el caso de la aplicación de este método, la demanda  $d_{block\ j}$  de cada bloque  $j$  es repartida entre los nodos del sistema a prorrata de los factores de participación de energía de cada uno durante las horas en  $I_j$ .

### 3.6.2. Algoritmos de clustering [1].

---

#### Algoritmo 3.6.1 Algoritmo PAM.

---

**Requiere:**

datos  $\mathbf{x}_j$  ó matriz de disimilitudes  $\mathcal{D}^{\text{diss}}$   
 número de *clusters*,  $k$

**Entrega:** asignación de elementos en *clusters* y  $k$  representantes

Inicialización:  $\hat{\mathcal{S}} = \mathcal{S} \quad \mathcal{M} = \emptyset \quad \check{\mathcal{M}} = \emptyset$

**fase BUILD**

- 1:  $p = \arg \min_{t \in \hat{\mathcal{S}}} \sum_{j \in \hat{\mathcal{S}}} D(\mathbf{x}_j, \mathbf{x}_t)$
- 2:  $\hat{\mathcal{S}} = \hat{\mathcal{S}} - \{p\} \quad \mathcal{M} = \mathcal{M} \cup \{p\}$
- 3: **for**  $i = 2, 3, \dots, k$  **do**
- 4:  $d_j = \min_{\theta \in \mathcal{M}} D(\mathbf{x}_j, \theta), \forall j \in \hat{\mathcal{S}}$   
 (disimilitudes respecto a los medoides ya seleccionados)
- 5:  $p = \arg \max_{t \in \hat{\mathcal{S}}} \sum_{j \in \hat{\mathcal{S}}} \max(d_j - D(\mathbf{x}_j, \mathbf{x}_t), 0)$   
 (selección de elemento que más disminuye suma de disimilitudes)
- 6:  $\hat{\mathcal{S}} = \hat{\mathcal{S}} - \{p\} \quad \mathcal{M} = \mathcal{M} \cup \{p\}$
- 7: **end for**

**fase SWAP**

- 8:  $F = \sum_{p \in \mathcal{M}} \min_{j \in \hat{\mathcal{S}}} D(\mathbf{x}_j, \mathbf{x}_p)$  (suma de disimilitudes tras BUILD)
  - 9: **for all**  $p \in \mathcal{M}$  **do**
  - 10: **for**  $j \in \hat{\mathcal{S}}$  **do**
  - 11:  $\hat{\mathcal{S}}_j = \{\hat{\mathcal{S}} \cup \{p\}\} - \{j\}$   
 (elementos no medoides tras cambiar  $\mathbf{x}_p$  por  $\mathbf{x}_j$ )
  - 12:  $\mathcal{M}_j = \{\mathcal{M} \cup \{j\}\} - \{p\}$   
 (elementos medoides tras cambiar  $\mathbf{x}_p$  por  $\mathbf{x}_j$ )
  - 13:  $F_j^{\text{swap}} = \sum_{l \in \hat{\mathcal{S}}_j} \min_{r \in \mathcal{M}_j} D(\mathbf{x}_l, \mathbf{x}_r)$   
 (suma de disimilitudes tras cambiar  $\mathbf{x}_p$  por  $\mathbf{x}_j$ )
  - 14: **end for**
  - 15:  $\iota = \arg \min_{j \in \hat{\mathcal{S}}} F_j^{\text{swap}}$  (mejor cambio o *swap*)
  - 16: **if**  $F_j^{\text{swap}} < F$  **then**
  - 17:  $\check{\mathcal{M}} = \check{\mathcal{M}} \cup \{\iota\}$
  - 18: **else**
  - 19:  $\check{\mathcal{M}} = \check{\mathcal{M}} \cup \{p\}$
  - 20: **end if**  
 (realizo el mejor cambio si éste disminuye la suma de disimilitudes)
  - 21: **end for**
  - 22:  $\{\theta_i\} = \check{\mathcal{M}}$
  - 23:  $S_i = \{\mathbf{x}_j / \|\mathbf{x}_j - \theta_i\| \leq \|\mathbf{x}_j - \theta_r\|, \forall r \in \{1, 2, \dots, k\}\},$   
 $\forall i \in \{1, 2, \dots, k\}$
-

---

**Algoritmo 3.6.2** Algoritmo *k-means*.
 

---

**Requiere:**

datos  $\mathbf{x}_j$   
 número de *clusters*,  $k$

**Entrega:** asignación de elementos en *clusters* y  $k$  representantesInicialización:  $l = 0$  $\{\boldsymbol{\theta}_i^{(l)}\} \leftarrow \text{aleatorio}(\{\mathbf{x}_j, j = 1, 2, \dots, N\}, k)$ 1: **repeat**2:   **for**  $i = 1, 2, \dots, k$  **do**3:      $S_i^{(l)} = \left\{ j \in \{1, 2, \dots, N\} / \|\mathbf{x}_j - \boldsymbol{\theta}_i^{(l)}\| \leq \|\mathbf{x}_j - \boldsymbol{\theta}_p^{(l)}\|, \forall p \in \{1, 2, \dots, k\} \right\}$     $N_i^{(l)} = \#S_i^{(l)}$    (definición de elementos en cada *cluster*)4:   **end for**5:    $l = l + 1$ 6:   **for**  $i = 1, 2, \dots, k$  **do**7:      $\boldsymbol{\theta}_i^{(l)} = \frac{1}{N_i^{(l)}} \sum_{j \in S_i^{(l)}} \mathbf{x}_j$ 

(actualización de los representantes)

8:   **end for**9: **until**  $\{\boldsymbol{\theta}_i^{(l)}\} = \{\boldsymbol{\theta}_i^{(l-1)}\}$ 


---

### 3.6.3. Datos de entrada adicionales

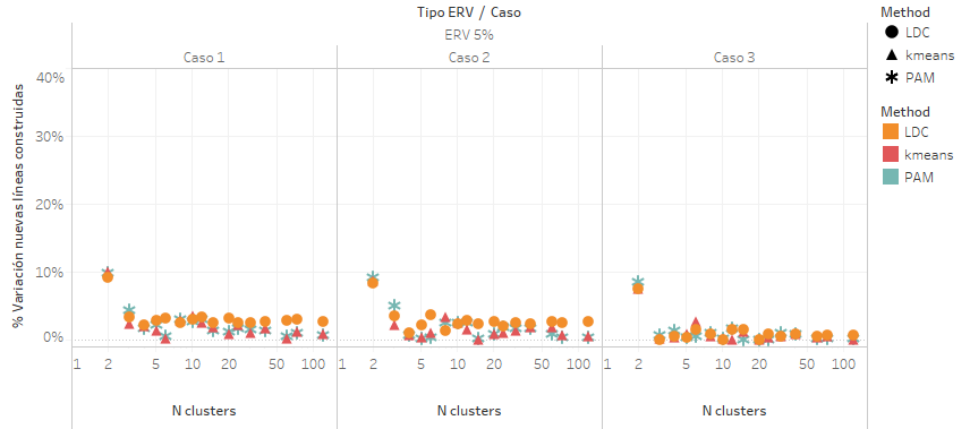
Tabla 3.6.1: Datos de generadores del sistema

$Gen_j$	$Location$	$Gen_{Max}$ [MW]	$Gen_{Min}$	$C_{MW}$ [\$/MW]
1	1	100	0	148.62
2	1	100	0	148.77
3	1	380	0	50.4
4	1	380	0	50.45
5	2	100	0	148.92
6	2	100	0	149.07
7	2	380	0	50.5
8	2	380	0	50.55
9	7	500	0	54
10	7	500	0	54.05
11	7	500	0	54.11
12	13	985	0	27.35
13	13	985	0	27.38
14	13	985	0	27.4
15	15	60	0	198.48
16	15	60	0	198.68
17	15	60	0	198.88
18	15	60	0	199.08
19	15	60	0	199.27
20	15	775	0	26.96
21	16	775	0	26.99
22	18	2000	0	16.32
23	21	2000	0	16.34
24	22	250	0	0.01
25	22	250	0	0.02
26	22	250	0	0.03
27	22	250	0	0.04
28	22	250	0	0.05
29	22	250	0	0.06
30	23	775	0	26.96
31	23	775	0	26.99
32	23	1750	0	3.31

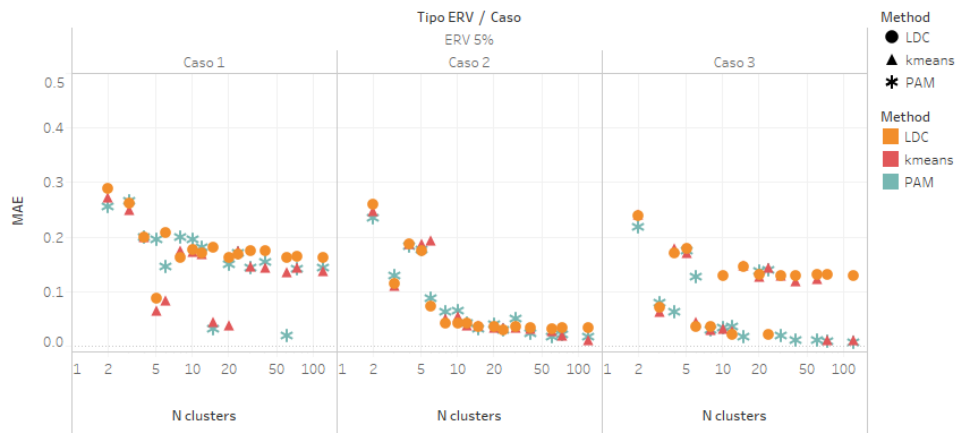
Tabla 3.6.2: Datos de líneas del sistema

$Bus_k$	$Bus_m$	$X_{km}$	$f_{max}$	$n_0$	$c_{km}[\$/MW]$	$n_{max}$
1	2	0.014	175	1	3,000,000	3
1	3	0.211	175	1	55,000,000	3
1	5	0.085	175	1	22,000,000	3
2	4	0.127	175	1	33,000,000	3
2	6	0.132	175	1	50,000,000	3
3	9	0.119	175	1	31,000,000	3
3	24	0.064	400	1	50,000,000	3
4	9	0.104	175	1	27,000,000	3
5	10	0.068	175	1	23,000,000	3
6	10	0.061	175	1	16,000,000	3
7	8	0.061	175	1	16,000,000	3
8	9	0.165	175	1	43,000,000	3
8	10	0.165	175	1	43,000,000	3
9	11	0.084	400	1	50,000,000	3
9	12	0.084	400	1	50,000,000	3
10	11	0.084	400	1	50,000,000	3
10	12	0.084	400	1	50,000,000	3
11	13	0.048	500	1	66,000,000	3
11	14	0.042	500	1	58,000,000	3
12	13	0.048	500	1	66,000,000	3
12	23	0.097	500	1	134,000,000	3
13	23	0.087	500	1	120,000,000	3
14	16	0.069	500	1	54,000,000	3
15	16	0.017	500	1	24,000,000	3
15	21	0.049	500	2	68,000,000	3
15	24	0.052	500	1	72,000,000	3
16	17	0.026	500	1	36,000,000	3
16	19	0.023	500	1	32,000,000	3
17	18	0.014	500	1	20,000,000	3
17	22	0.105	500	1	146,000,000	3
18	21	0.026	500	2	36,000,000	3
19	20	0.04	500	2	55,000,000	3
20	23	0.022	500	2	30,000,000	3
21	22	0.068	500	1	94,000,000	3

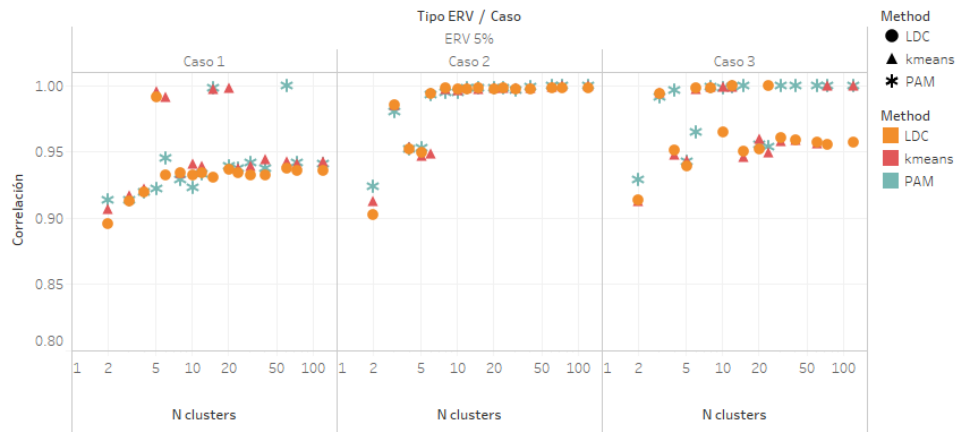
3.6.4. Resultados adicionales



(a) % Error nuevas líneas construidas

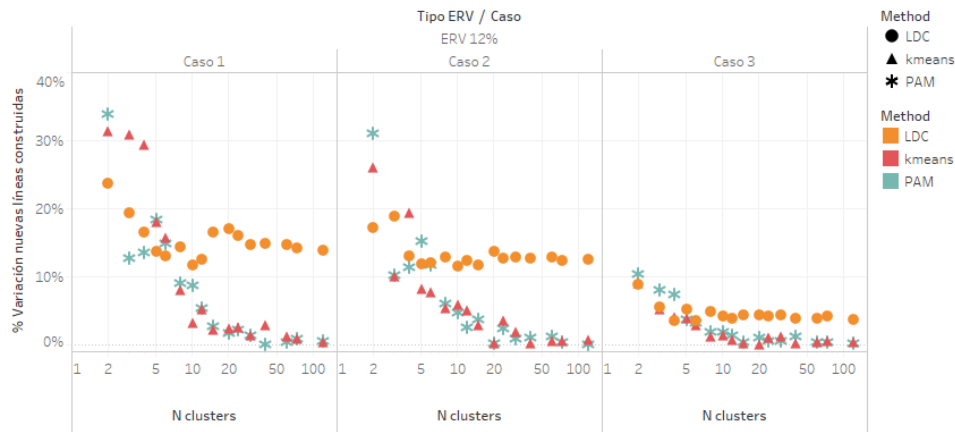


(b) MAE

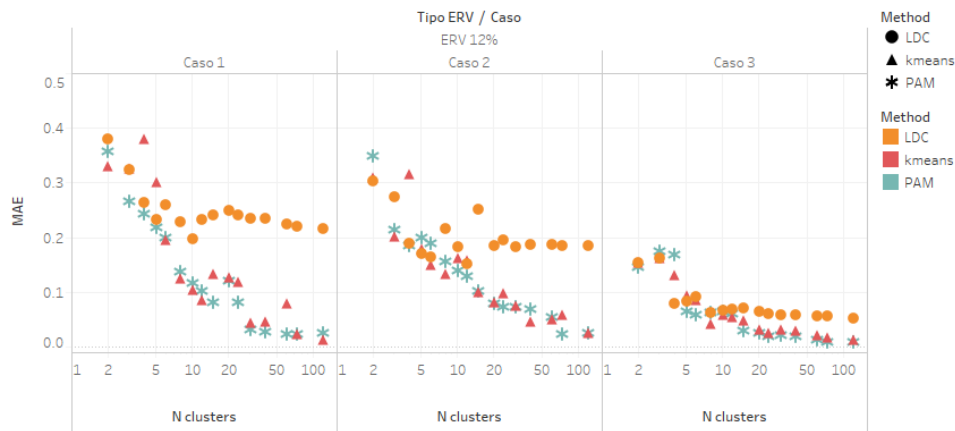


(c) Correlación

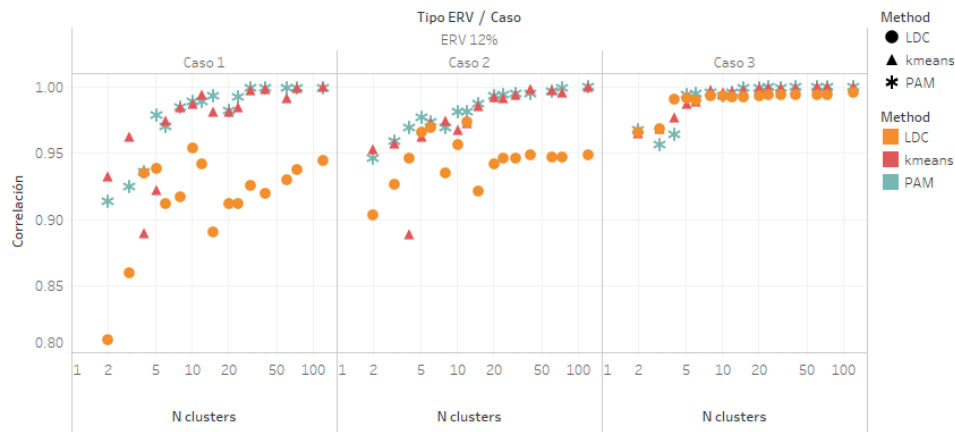
Figura 3.6.1: Métricas de evaluación de casos de estudio para  $ERV_{\frac{1}{5}}$ ,  $ERV_{\frac{2}{5}}$  y  $ERV_{\frac{3}{5}}$ .



(a) % Error nuevas líneas construidas

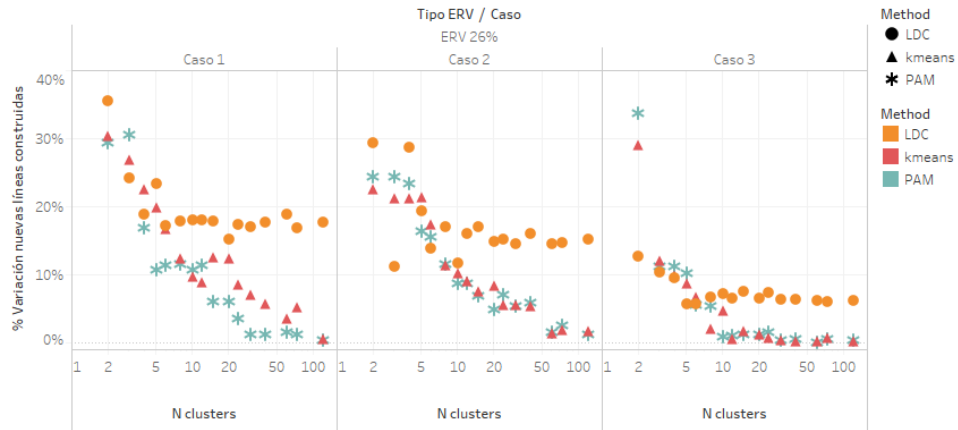


(b) MAE

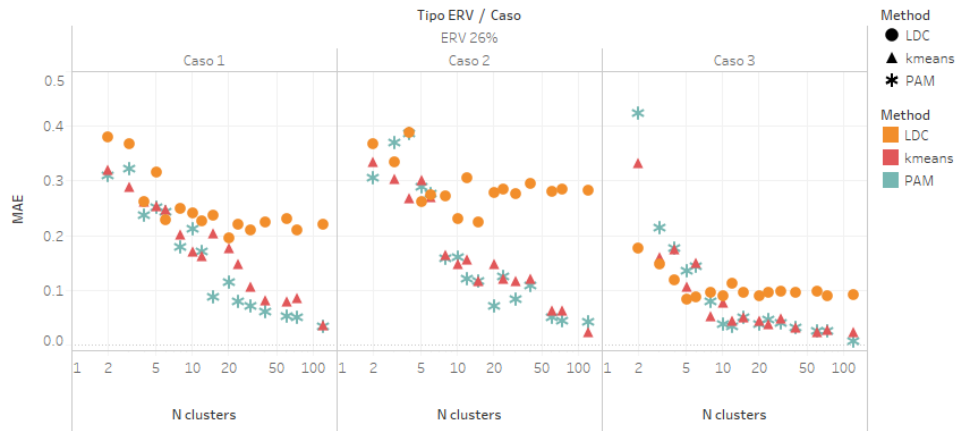


(c) Correlación

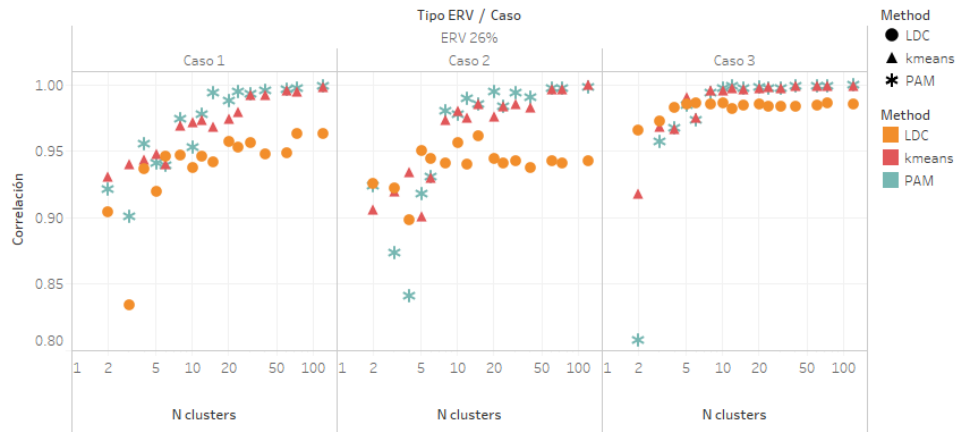
Figura 3.6.2: Métricas de evaluación de casos de estudio para  $ERV_{12}^1$ ,  $ERV_{12}^2$  y  $ERV_{12}^3$ .



(a) % Error nuevas líneas construidas

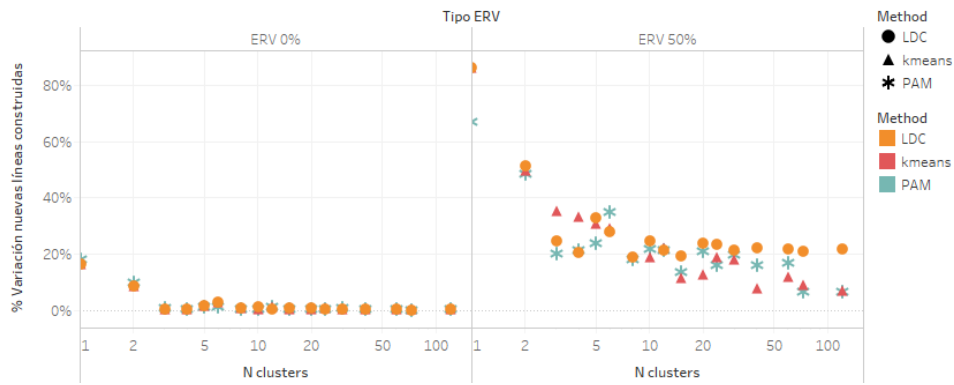


(b) MAE

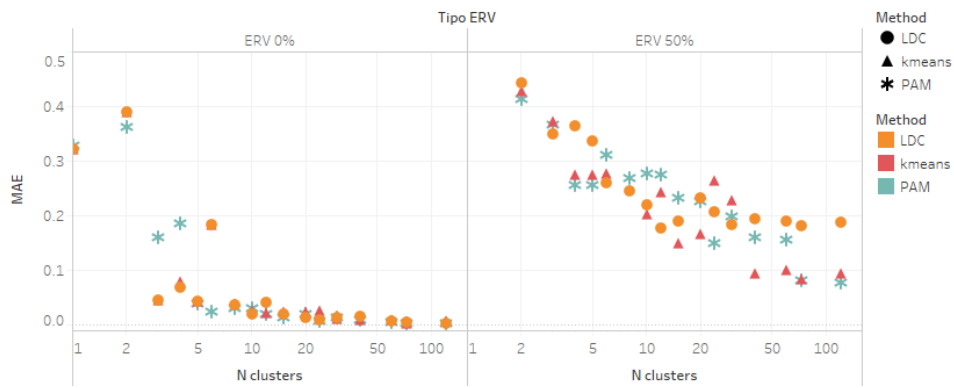


(c) Correlación

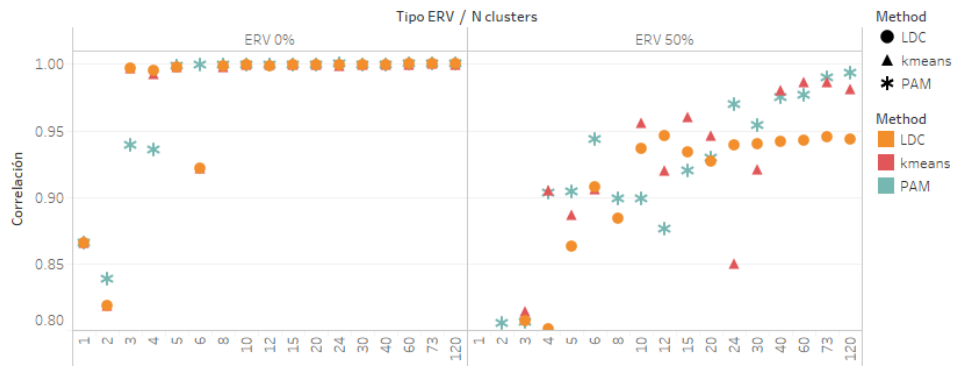
Figura 3.6.3: Métricas de evaluación de casos de estudio para  $ERV_{26}^1$ ,  $ERV_{26}^2$  y  $ERV_{26}^3$ .



(a) % Error nuevas líneas construidas



(b) MAE



(c) Correlación

Figura 3.6.4: Métricas de evaluación de casos de estudio para  $ERV_0$  y  $ERV_{50}$ .

## Capítulo 4

# Reducción Dimensional de la Demanda en Base a Condiciones Operacionales de un Análisis Post-óptimo

**RESUMEN:** El presente capítulo presenta una metodología alternativa a la reducción de escenarios de demanda, basando su *clustering* en un análisis post-óptimo de variables de interés para el problema de expansión de la transmisión. La metodología empleada consiste en el pre-procesamiento de datos de entrada a través de un flujo óptimo de potencia, cuyas salidas relativas a costos de operación y flujos por las líneas serán procesadas y clusterizadas para rescatar grupos de demanda representativos. Las metodologías propuestas realizan su proceso de *clustering* basadas en las técnicas de k-means, PAM y la conocida LDC (demanda neta), siendo PAM la técnica con mayor requerimiento computacional asociado. De este modo, se agrupan los respectivos indicadores horarios de las variables de interés y a través de estos realizar los respectivos grupos de demanda del sistema. De manera adicional, se presenta una variable relativa a los flujos del sistema que considera información de interés respecto a los flujos de líneas con y sin límites de capacidad de transmisión. Las metodologías propuestas se contrastan con la metodología propuesta en el capítulo anterior, que consiste en el *clustering* directo en la demanda. El método propuesto se prueba en el sistema de prueba IEEE de 24 barras para diferentes niveles de penetración renovable (0 %, 5 %, 12 %, 26 % y 50 %), ubicando parques eólicos y solares en diferentes localizaciones para mostrar los efectos de la variabilidad espacial y temporal en los perfiles de demanda neta. Los resultados numéricos indican que la selección de una metodología adecuada depende directamente del nivel de variabilidad del sistema. De manera general, el agrupar a partir de la demanda y flujos de potencia sin límites de capacidad son las metodologías con mejor desempeño cuando se utiliza k-means y PAM en sus *clusters*. Así, para los casos de estudio con baja variabilidad el estas dos tienen un rendimiento similar, sin embargo en los casos de alta variabilidad el agrupamiento basado en PAM tiene

el mejor desempeño.

## 4.1. Introducción

El crecimiento exponencial de la generación de Energías Renovables (**ER**) en los Sistemas Eléctricos de Potencia (**SEP**) [3] ha alcanzado en el mundo niveles del orden del 25 % de la producción eléctrica a nivel mundial [83]. Relativo a este tipo de energías, la generación de Energías Renovables Variables (**ERV**)<sup>1</sup> se modela generalmente como una reducción en la demanda del sistema, trabajando en estos casos con una "**Demanda Neta**", es decir, a la curva de demanda real se le resta esta generación de ERV y se trabaja con el perfil de demanda resultante.

De este modo, junto con el aumento de generación de ERV, los mercados se han visto en la necesidad de generar incentivos económicos que busquen la inserción de recursos que permitan aumentar la flexibilidad del sistema, y de este modo integrar nuevas restricciones operativas al problema de planificación como lo son el almacenamiento en base a baterías para Control Primario de Frecuencia (CPF), máquinas con alta tasa de toma de carga para Control Secundario de Frecuencia (CSF), mínimos técnicos, tiempos mínimos de encendido y apagado, entre otras; generando de este modo nuevos desafíos para los operadores de los sistemas eléctricos actuales. En el caso particular de la presente tesis, se trabajará enfocados en las ERV, tales como la energía solar y eólica.

Otros factores de relevancia son las nuevas necesidades de los sistemas, como aquellas relativas a interconexiones de sistemas eléctricos (integración energética entre sistemas), el desarrollo de redes inteligentes y nuevas tecnologías. Estos factores conllevan un aumento y variaciones en los perfiles de demanda y generación, afectando a variables temporales y espaciales del sistema de potencia. Es así como el problema de optimización relativo al SEP ha aumentado su nivel de complejidad, haciendo cada vez más difícil el tratamiento de sus variables.

El presente capítulo presenta una metodología alternativa de representación reducida de la demanda y generación ERV aplicada al problema de expansión de la transmisión, en donde se pretende realizar una representación adecuada de los bloques de demanda a través de información relevante al TEP. Así, se considera en este caso un análisis post-óptimo de los datos de entrada del problema, los cuales ingresarán previamente a un subproblema con características similares al TEP (se utiliza un OPF), para luego rescatar información relativa a sus costos de operación y flujos de potencia para realizar los respectivos *clusters* de demanda.

El método propuesto se prueba en el sistema de prueba IEEE de 24 barras para diferentes niveles de penetración renovable (0 %, 5 %, 12 %, 26 % y 50 %), ubicando parques eólicos y solares en diferentes localizaciones para mostrar los efectos de la variabilidad espacial y temporal en los perfiles de demanda neta.

Respecto al análisis de los beneficios relativos a la metodología propuesta se utilizan diferentes métricas de comparación de los resultados obtenidos respecto a la solución

---

<sup>1</sup>Se considerarán la energía solar y eólica.

del problema sin reducción de escenarios. Las métricas utilizadas son: MSE, MAE y correlaciones para análisis de los vectores de nuevas líneas construidas, comparaciones porcentuales de variación respecto a función objetivo y nuevas líneas, y además, se introduce una métrica que se denominará “FO recalculada”, la cual consiste en comparar el valor de la función objetivo de la solución real (8760 horas) con la solución de las nuevas líneas construidas obtenida aplicando la reducción de escenarios, y evaluando esta en un OPF, de modo tal de obtener la operación bajo la escala temporal original.

Finalmente, los resultados se contrastan con los relativos al capítulo anterior, para poder concluir respecto al aporte del desarrollo de esta nueva metodología.

#### 4.1.1. Estado del Arte

El problema de la expansión de la transmisión ha sido estudiado desde numerosas perspectivas a través del tiempo [36, 45], siendo el enfoque principal de este trabajo la reducción de la dimensionalidad de sus datos de entrada con alta variabilidad, como lo son la demanda y la generación ERV.

Debido a la gran cantidad de variables que considera el problema, es necesario reducir de manera óptima su representación, procurando perder la menor cantidad de información relevante en el proceso. De este modo, en sus inicios las primeras representaciones utilizadas fueron a través de la LDC [67, 66]. Sin embargo, dados los nuevos requerimientos de los sistemas respecto a la consideración de variabilidad de demanda e integración renovable, su implementación se ha ido adaptando según los diferentes casos de estudios. De este modo, la representación de la LDC se ha visto limitada a la representación de perfiles con baja variabilidad y alta correlación espacial, dado que su implementación considera la reducción espacial de la demanda. La evolución de esta aproximación a través del tiempo se desarrolla en [72, 70, 71, 69, 68]. Algunos casos de aplicación de esta metodología se presentan en [30, 32, 62].

Otro tipo de representación alternativa es la presentada en [59], en donde se definen estados del sistema, y en torno a estos se define una matriz de transición que ayuda capturar de mejor manera su cronología. Si bien este tipo de representaciones son más precisas, el planteamiento requerido al problema de expansión de la transmisión es diferente, dado el gran espacio temporal considerado, y por ende, el complejo manejo computacional de este tipo de modelos.

De este modo, el problema de modelamiento de la demanda se ha convertido en un desafío en los sistemas eléctricos, definiendo sus niveles de precisión acorde al tipo de problema, modelo y espacio temporal de simulación requerido. Es así como, respecto a la planificación de los sistemas de potencia, existen desarrollos relativos a la representación de demanda en la expansión de capacidad del sistema [29], como también desarrollos relativos a la transmisión cuyos enfoques de representatividad deben rescatar las variables de interés según el modelo.

Además, el estudio de reducción de variables puede ser considerado desde diferentes perspectivas, como lo son la espacialidad, temporalidad y estocasticidad. De manera particular, el presente trabajo se enfoca en la reducción temporal de la demanda aplicada en el TEP.

Trabajos similares se presentan en [85, 86], en donde se plantea una metodología de trabajo similar a la desarrollada. Sin embargo el planteamiento de la reducción de escenarios es respecto a la variable estocástica, tema en el cual se han desarrollado numerosos trabajos respecto al tratamiento de la variabilidad e incertidumbre en el problema de planificación de la transmisión.

Relativos a la reducción de la espacialidad del problema se tienen trabajos como [87, 88] en donde consideran agrupamientos por zonas.

Finalmente, el trabajo más reciente respecto a esta temática se presenta en [89], en donde se desarrollan tres nuevas técnicas para reducir el tamaño del problema TEP en sus tres dimensiones principales: la representación espacial [88], la representación operacional (representación temporal) [52] y el número de elementos de la red a considerar. Si bien el desarrollo de esta tesis posee aspectos similares a los desarrollados en [89] en cuanto a la reducción temporal, las técnicas de reducción de escenarios propuestas son distintas, intentando replicar algunos casos comparativos para poder rescatar los beneficios de las nuevas técnicas propuestas. Además se propone un mayor número de casos de estudio y un modelamiento diferente del problema de expansión de la transmisión considerando nuevas métricas de comparación de resultados.

Relativos a la metodología de trabajo empleada, trabajos similares al presente estudio son [85, 52], en donde se analizan de manera post-óptima las variables de entrada del problema de transmisión, realizando una clusterización basada en un post procesamiento de los datos, que han sido tratados previamente en un problema más pequeño, como por ejemplo un flujo óptimo de potencia (OPF). La gran diferencia con el trabajo desarrollado en [85] es que ese estudio es respecto a la representación estocástica de la demanda, a diferencia del presente estudio que realiza una reducción respecto a la variable temporal. Respecto al trabajo desarrollado en [52], las metodologías de agrupamiento de demanda propuestas son diferentes, debido a que el autor basa la selección de casos representativos en los potenciales beneficios de construcción de la líneas, a diferencia del presente trabajo, en donde la agrupación de los datos es basada en rescatar el efecto de la demanda en los costos operacionales y flujos de potencia del sistema.

#### 4.1.2. Contribuciones y Organización del Capítulo

Las principales contribuciones de este trabajo son:

1. Una metodología de representación reducida de la demanda para su aplicación en el problema de planificación de la expansión de la transmisión, basada en variables relativas al sistema de transmisión.
2. La aplicación de la metodología de representación reducida en diferentes perfiles de demanda y localizaciones, resaltando la importancia de considerar la espacialidad en este tipo de estudios.
3. El uso y definición de métricas de evaluación de resultados acorde al problema de expansión de la transmisión.

4. Comparación de diferentes metodologías de representación reducida de la demanda y análisis según caso de estudio.

El presente capítulo se organiza según:

- **Introducción:** se plantea la problemática abarcada.
- **Estado del Arte:** se hace un breve resumen respecto a la literatura actual referente al tema y a los principales trabajos utilizados como referentes para el desarrollo de esta tesis, y se resaltan las diferencias respecto a las contribuciones de estos.
- **Metodología:** se realiza una visión general respecto a las variables y simplificaciones del problema de expansión de transmisión consideradas en este trabajo. Además, se describen las metodologías utilizadas para la reducción de dimensionalidad a través del análisis post-óptimo de las variables obtenidas en el OPF. Se describen las variables de estudio utilizadas y las principales hipótesis formuladas. Finalmente se realiza una descripción general de la totalidad de casos estudiados, y los diferentes métodos de agrupación utilizados (LDC, k-means y PAM).
- **Casos de estudio:** de manera general se describen los casos de estudio seleccionados con sus respectivos datos considerados. Se presentan además las métricas de evaluación de resultados utilizadas en el trabajo.
- **Resultados numéricos:** se presentan los principales resultados numéricos a través de gráficas representativas y tablas resumen de resultados principales.
- **Conclusiones:** se indican las principales conclusiones de este trabajo y las posibles extensiones futuras.
- **Anexos:** se adjunta una sección anexa con resultados relevantes del presente estudio.

## 4.2. Metodología

### 4.2.1. El problema de la expansión de la Transmisión (TEP)

Una adecuada planificación y expansión de la transmisión de un sistema eléctrico en un mercado competitivo debe permitir su desarrollo a un mínimo costo, y proveer de señales que incentiven la eficiencia económica y técnica de éste, adaptando los requerimientos de generadores y consumidores, remunerando adecuadamente a sus propietarios [35].

El TEP busca encontrar respuestas al dónde, cuándo y cómo nuevas líneas deberían ser instaladas en el sistemas para ser capaz de abastecer su demanda, sujeto a restricciones físicas y económicas [24]. De este modo, en su planteamiento matemático el TEP es un problema de programación no lineal entero mixto con variables de decisión binarias relativas a la construcción de nuevas líneas o su expansión.

Para el presente trabajo se utiliza un modelo de planificación estática, es decir, se realiza la planificación de un año entero, en donde en el tiempo  $t_0$  se realiza inversión en la transmisión para el resto del año, para posteriormente llevar a valor presente los respectivos costos de operación del sistema, y de este modo, proyectar la inversión a tiempo futuro.

Respecto al modelo de planificación, se ha escogido un modelo de planificación no lineal, cuyas variables de decisión respecto a cantidad de nuevas líneas son enteras, sin embargo, dada la naturaleza entera del problema, se ha escogido además realizar una relajación a este problema. Para ello se han considerando las variables variables de decisión de nuevas líneas construidas como continuas, haciendo de este modo el problema más tratable computacionalmente.

Es importante notar que según el tipo de modelo de planificación y sus restricciones, es posible que a medida que se considera un mayor número de *clusters* el problema aumente su cantidad de variables enteras de decisión, y por ende aumenten de manera exponencial sus tiempos de simulación, por lo que encontrar la solución para el horizonte temporal total se hace altamente costoso computacionalmente. Sin embargo, dado el modelo de planificación estático escogido, se han utilizado restricciones que mantienen fijo el número de variables enteras<sup>2</sup>, y que por tanto simplifican esta problemática.

Además, al considerar las variables de decisión de construcción de líneas como continuas se tendrán menos requerimientos computacionales para la determinación de sus soluciones y un mejor contraste respecto a la solución real del problema, ya que pequeñas variaciones en el vector solución podrán ser percibidas por las métricas utilizadas. Más detalles respecto al modelo de transmisión utilizado se presentan en la sección 2.1.2.3.

#### 4.2.2. Curva de duración de la demanda

Como se mencionó anteriormente, históricamente se ha implementado en los problemas de planificación eléctrica una representación reducida de la demanda a través de su llamada curva de duración.

El objetivo es aproximar la curva de demanda de un sistema en  $L$  bloques, donde cada bloque de demanda define un nivel y duración. Para lograr este objetivo, se resuelve un problema de optimización que divide la curva de carga continua en  $L$  grupos. Esta partición de datos busca minimizar la suma, sobre todos los grupos, de las distancias punto a punto entre el grupo y un “centro”<sup>3</sup>. El resultado de este proceso produce una representación de la demanda discreta que parece una función escalonada.

Relativo al presente trabajo, para la construcción de la LDC se emplea el esquema presentado en la figura 4.2.1, en donde inicialmente se tienen los perfiles de demanda del sistema separados de forma temporal (horas) y espacial (barras). Luego se agrupan espacialmente a través de su demanda neta, perdiendo la información espacial relativa a su localización.

<sup>2</sup>Las decisiones de inversión solo se toman en el tiempo  $t_0$ , por lo que son fijas para el resto del horizonte temporal.

<sup>3</sup>Al existir diferentes metodologías de agrupamiento de datos o *clustering*, el tipo de centro se escogerá según la técnica empleada. La metodología más comúnmente empleada para este tipo de casos es el método de mínimos cuadrados.

Posteriormente, se realiza el agrupamiento unidimensional, a través de alguna técnica de *clustering*, en este caso *k-means*, el cual realiza un agrupamiento de datos basado en la minimización de la distancia al centro de los *clusters*, en donde se rescatará su índice horario para poder asignarlos posteriormente a las respectivas barras desagregadas.

Así, el tercer paso es desagregar la demanda por barras, a través de sus respectivos índices horarios asignados por *clusters*, para luego, a través del promedio de cada uno, obtener su centroide representativo. De este modo se obtiene finalmente una matriz reducida temporalmente, pero que conserva la dimensión espacial.

Las principales falencias de esta metodología son relativas a la pérdida de información espacial de la demanda al realizar la respectiva agregación a través de la demanda neta.

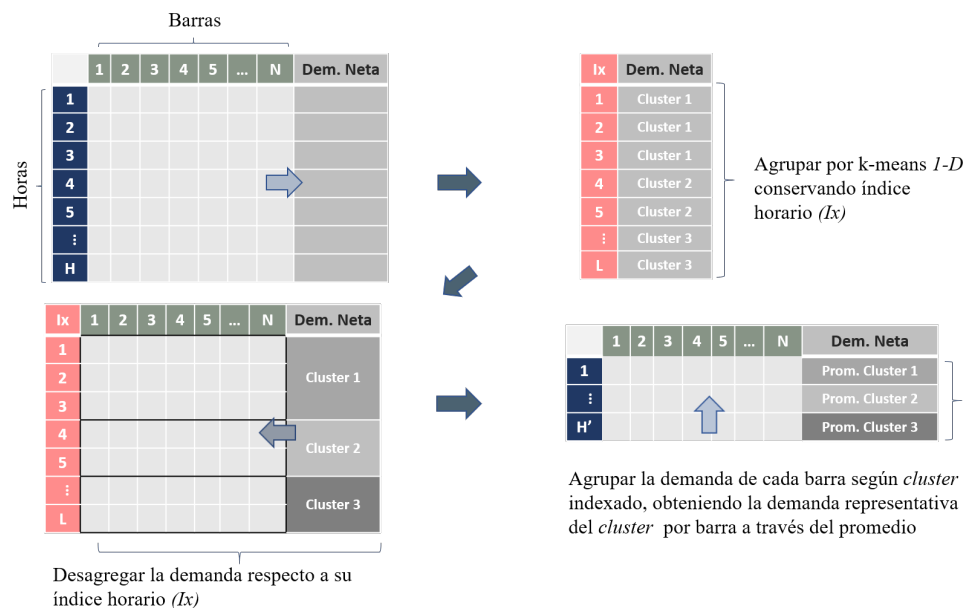


Figura 4.2.1: Esquema típico de reducción de la LDC.

#### 4.2.2.1. Técnicas de Agrupamiento de Datos: *Clustering*

A continuación se presentan las principales técnicas *Clustering* empleadas en este trabajo y sus respectivas ventajas y desventajas.

*K-means* [82]:

- *Ventajas de k-means:*

1. Si se tiene un gran número de variables, *k-means* la mayor parte del tiempo es más rápido que los *Hierarchical Clustering Methods*.
2. *K-Means* produce *clusters* más estrictos que los *Hierarchical clustering Methods*.

- *Desventajas de k-means:*

1. Su valor  $\mathbf{K}$  (o número de particiones) es difícil de predecir.
2. Al iniciar aleatoriamente los centros del *cluster* puede producir resultados muy diferentes en diferentes ejecuciones del algoritmo, haciendo que los resultados a veces carezcan de consistencia y lo hace difícil de comparar versus otros *clusters*.

*PAM (Partitioning Around Medoids)* [91]:

■ *Ventajas de PAM:*

1. Entrega resultados más estables y robustos que k-means
2. Las distancias entre los elementos puede calcularse en forma previa y una sola vez, lo que permite usar cualquier tipo de distancia.
3. Asignación determinista de elementos.
4. Es simple de entender y fácil de implementar.
5. Tiene rápida convergencia en un número fijo de pasos.
6. Es menos sensible a valores atípicos que otros algoritmos de partición.

■ *Desventajas de PAM:*

1. Es más costoso que k-means dada su complejidad temporal.
2. No funciona bien para grandes cantidades de datos.
3. Sus resultados y su tiempo de simulación dependen de la partición inicial.

#### 4.2.2.2. Implementación de k-means y PAM para reducción directa de la demanda

Relativo al presente trabajo, la reducción de dimensionalidad de la demanda a través de *clustering* se empleará el esquema presentado en la figura 4.2.2, en donde inicialmente se tienen los perfiles de demanda del sistema separados de forma temporal (horas) y espacial (barras). Luego se agrupan multidimensionalmente a través del algoritmo de *clustering* seleccionado, relacionando de esta forma la espacialidad y temporalidad del problema y obteniendo como salida los respectivos índices (Ix) de los respectivos grupos de *clusters*.

Posteriormente, a través de los índices horarios que generan los nuevos *clusters* se asignarán los nuevos grupos representados a través de las respectivas demandas netas horarias. Así, el tercer paso es desagregar la demanda por barras, a través de sus respectivos índices horarios asignados, para luego, a través del promedio de cada uno, obtener su centroide representativo. De este modo se obtiene finalmente una matriz reducida temporalmente, pero que conserva la dimensión espacial.

A diferencia del esquema de la LDC aplicado directamente en la demanda, en el caso de los métodos de k-means y PAM la selección de los índices de los *clusters* representativos se realizar observando las variables temporales y espaciales del problema,

conservando información espacial relevante en la selección de los respectivos grupos de demanda.

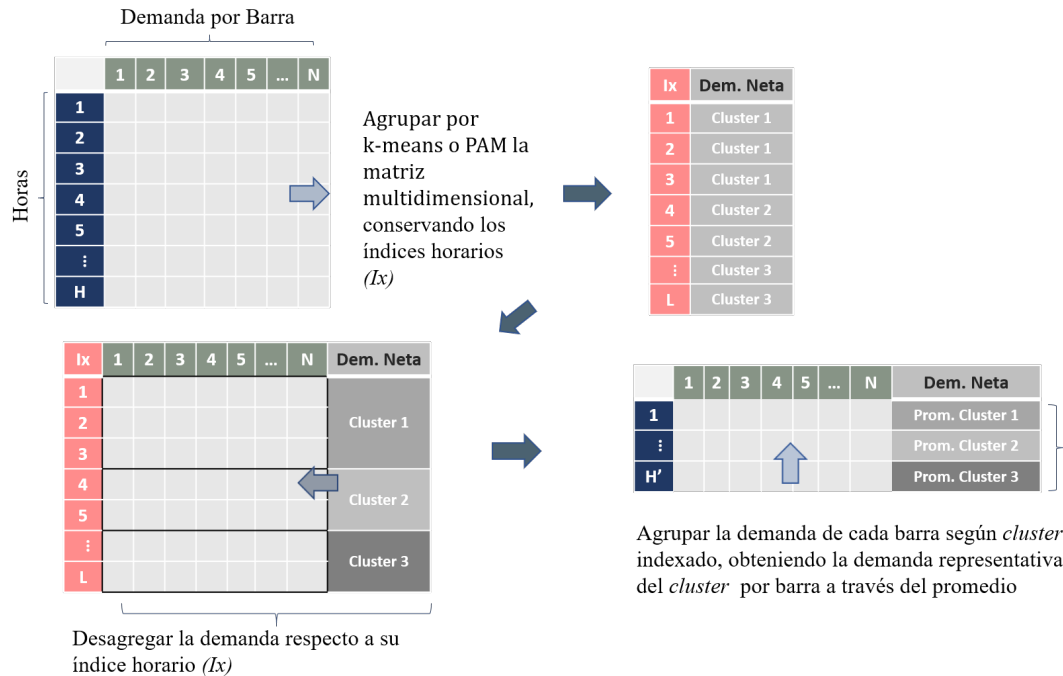


Figura 4.2.2: Esquema de reducción para k-means y PAM.

### 4.2.3. Metodología de análisis post-óptimo

La metodología general utilizada para el análisis post-óptimo consiste realizar un *clustering* respecto a alguna de las salidas obtenidas en un OPF<sup>4</sup> preliminar, realizado con los respectivos datos de demanda según el caso de estudio correspondiente. Para ello se agrupa el vector o conjunto de vectores de salida y se “agrega” temporalmente según alguna técnica de agrupamiento o *clustering* (LDC, PAM o k-means). En este proceso se busca rescatar los respectivos índices horarios del grupo, para posteriormente agrupar la demanda original según estos índices. Luego, se utiliza el promedio de los datos de cada *cluster* como su dato representativo.

#### 4.2.3.1. Tipos de datos de entrada según metodología de clustering:

A continuación se presentan las diferentes variables de *clustering* utilizadas para el respectivo análisis post-óptimo de los datos de entrada del problema y su respectiva nomenclatura utilizada. En cada caso el *clustering* se basará en:

- **COp (con):** Se considera el vector de costos de operación de OPF realizado con restricciones de capacidad de líneas.

<sup>4</sup>La formulación del OPF utilizada se presenta en la sección 2.1.1.1.

- **COp (sin):** Se consideran el vector de costos de operación de OPF realizado sin restricciones de capacidad de líneas.
- **Flujos (con):** Se considera el conjunto de vectores de flujos por las líneas del OPF realizado con restricciones de capacidad de líneas.
- **Flujos (sin):** Se considera el conjunto de vectores de flujos por las líneas del OPF realizado sin restricciones de capacidad de líneas.
- **Alt.:** Se considera un conjunto de vectores de métricas de congestión ( $fl_{alt}$ ), el cual consiste en un agrupamiento respecto a flujos con ( $fl_{con}$ ) y sin ( $fl_{sin}$ ) límites de capacidad en las líneas. De este modo, considerando la reducción por unidad de los diferentes flujos por las líneas, se define  $fl_{alt}$  según:

$$fl_{alt} = \begin{cases} 0, & \text{si } -1 < fl_{sin} < 1 \\ fl_{sin} - fl_{con}, & \text{en otro caso.} \end{cases}$$

Está formulación busca agrupar previamente los casos de las líneas no saturadas asignándoles un valor cero, y de este modo dar un mayor énfasis a los casos de líneas mayormente estresadas.

- **Demanda:** Se considera el conjunto de vectores originales de demanda, y se realiza el *clustering* aplicando directamente el método de agrupamiento en la demanda. Esta metodología es la señalada en las figuras 4.2.1 y 4.2.2 y tal como se realizó en el Capítulo 3.

A continuación se presenta la metodología utilizada para la reducción de demanda a partir de costos de operación y los respectivos flujos por las líneas.

**Implementación de reducción considerando Costos de Operación:** El esquema para reducción de escenarios basados en costos de operación se presenta en la figura 4.2.3, en donde inicialmente se tienen los costos de operación horarios del sistema representados a través de una columna única. En este caso la metodología de agrupamiento será similar para LDC, k-means y PAM, realizando el *cluster* sobre este vector unidimensional y rescatando su índice horario ( $I_x$ ). Cada índice horario pertenece a un respectivo *cluster*, en donde posteriormente el valor de costo de operación será reemplazado por la respectiva demanda neta de esas horas.

Una vez obtenidos los respectivos grupos de demanda el siguiente paso es desagregarlos espacialmente (por barras), a través de sus respectivos índices horarios asignados, para luego, a través del promedio de cada uno, obtener su centroide representativo. De este modo se obtiene finalmente una matriz reducida temporalmente, pero que conserva la dimensión espacial.

Este esquema de trabajo se utiliza para los casos COp (con) y COp (sin).

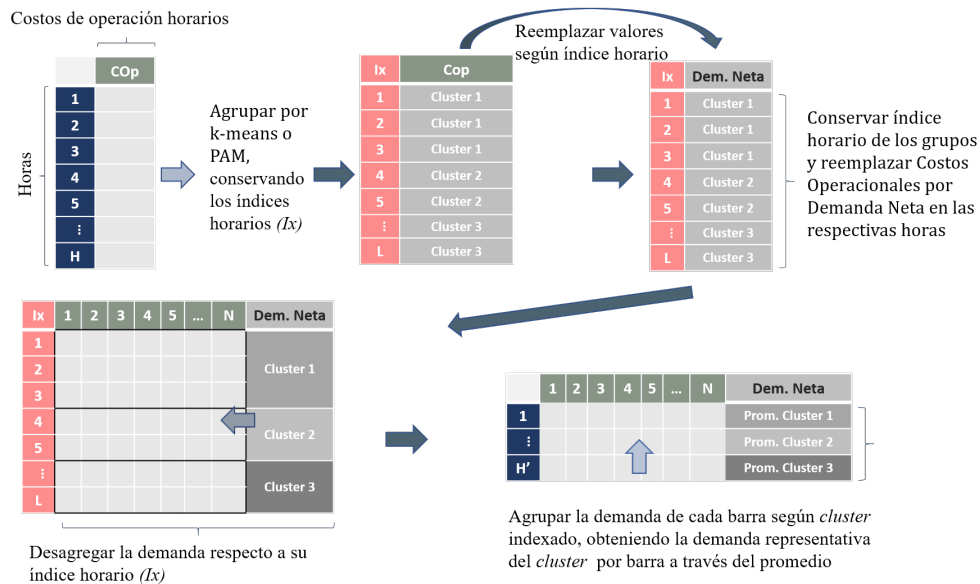


Figura 4.2.3: Esquema de reducción según Costos de Operación a través de LDC, k-means y PAM.

#### 4.2.3.2. Implementación de reducción considerando Flujos

La metodología para la reducción de escenarios basados en los flujos de potencia por las líneas se presenta dos esquemas representativos en las figuras 4.2.4 y 4.2.5. En la figura 4.2.4 se desarrolla el esquema de reducción a través del método LDC, el cual se ha basado en replicar la metodología implementada en el resto de los casos. Inicialmente, se tiene una matriz de valores de flujos a nivel horario y por línea.

Continuando con el esquema de trabajo implementado para los casos anteriores, se crea un vector de “Flujo neto horario”<sup>5</sup>, el cual consiste en la suma de todos los flujos por las líneas a nivel horario. Luego este vector resultante es agrupado de forma unidimensional a través de k-means, procurando conservar su respectivo índice horario ( $I_x$ ). Una vez obtenidos los respectivos *clusters* basados en sus indicadores horarios, se reemplaza la variable de flujo por su respectiva demanda neta horaria, para luego desagregarlos espacialmente (por barras), a través de sus respectivos índices horarios asignados. Finalmente, a través del promedio de cada *cluster* se obtiene su centroide representativo de cada *cluster*, agregándolos de manera temporal y obteniendo la reducción dimensional de la demanda.

Para el agrupamiento de datos a través de k-means y PAM se utiliza el esquema de la figura 4.2.5, cuya principal diferencia es la aplicación del agrupamiento de datos de manera directa en la respectiva matriz de flujos. Una vez obtenidos los *cluster* para cada índice horario, al igual que en el caso anterior se varía la variable de flujo por las

<sup>5</sup>Si bien la creación de este vector no posee un sentido físico, su representación se ha desarrollado para poder contrastar las metodologías utilizadas

líneas por la demanda neta del sistema, procediendo luego desagregando espacialmente la demanda, y agregándola de manera espacial a través del promedio de los *clusters*.

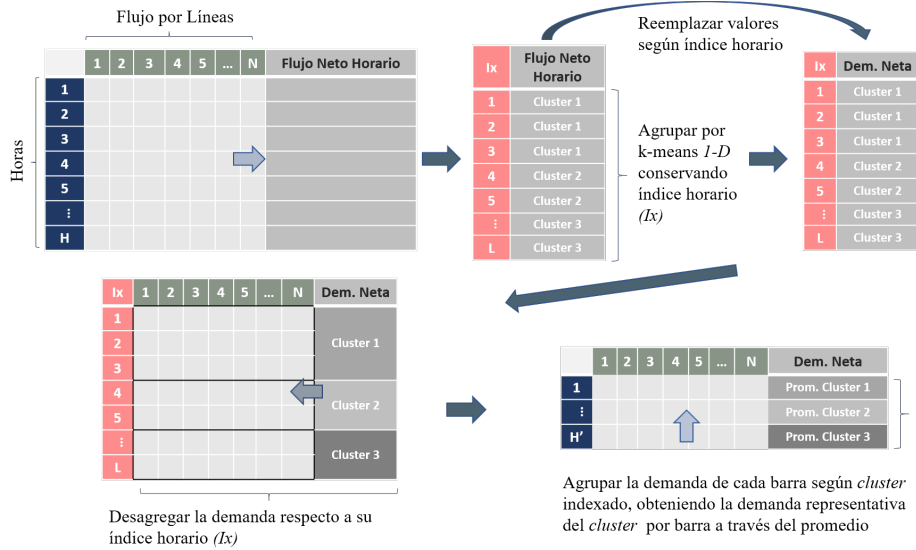


Figura 4.2.4: Esquema de reducción según Flujos por las líneas a través de LDC.

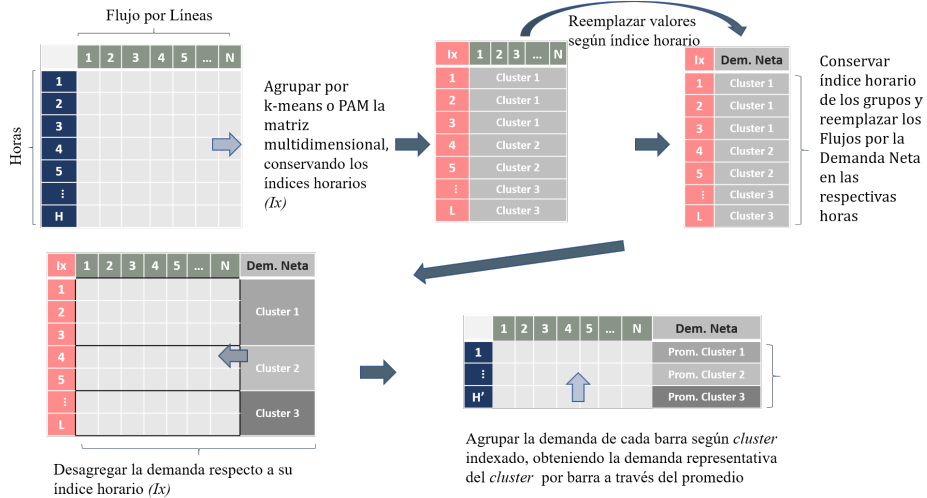


Figura 4.2.5: Esquema de reducción según Flujos por las líneas a través de k-means y PAM.

Estas metodologías de reducción espacial a través de los flujos serán aplicadas para los casos flujos (con), flujos (sin) y flujo métricas de congestión.

## 4.2.3.3. Resumen de características de las variables de estudio escogidas

Clustering basado en	Abreviatura	Comentario general	Método	Dimensionalidad del Cluster			Expectativa de representatividad espacial
				1-D	2-D	Baja Média Alta	
Costos de operación OPF con límites de capacidad de líneas	COP (con)	Se espera que los costos de operación con límites de capacidad de línea entreguen mayor información respecto a costos elevados asociados a saturaciones de las líneas del sistema	LDC	x		x	
			K-means	x		x	
			PAM	x		x	
Costos de operación OPF sin límites de capacidad de líneas	COP (sin)	Se espera una no muy variada representación de los costos del sistema debido a la no existencia de saturación en el sistema	LDC	x		x	
			K-means	x		x	
			PAM	x		x	
Flujos por la líneas OPF con límites de capacidad de líneas	Flujos (con)	Se espera una no muy variada representación a través de los flujos, debido al rango acotado de valores que pueden tomar los flujos [1;1]	LDC	x		x	
			K-means	x		x	
			PAM	x		x	
Flujos por la líneas OPF sin límites de capacidad de líneas	Flujos (con)	Se espera una mejor representación dado que los flujos por las líneas pueden tomar valores mayores a 1, y por tanto se puede rescatar de mejor forma la localización de las líneas que requieren ser reforzadas en el sistema.	LDC	x		x	
			K-means	x		x	
			PAM	x		x	
Métricas de congestión (o método alternativo) -Alt		Esta técnica es una mezcla del caso flujos (con) y flujos (sin), por lo cual se dá un mayor énfasis en las líneas saturadas del sistema y se rescatan sus valores. En los casos de líneas no saturadas su valor se hace cero, entregando una señal de que son un grupo relativamente homogéneo respecto a la demanda. De este modo, este caso de estudio debiese tener mejor rendimiento que los anteriores.	LDC	x		x	
			K-means	x		x	
			PAM	x		x	
Demanda	Demanda	La realización del clustering de manera directa sobre la demanda es el caso base para comparar cuán necesario es el desarrollo de estas nuevas técnicas de post-procesamiento de datos.	LDC	x		x	
			K-means	x		x	
			PAM	x		x	

Tabla 4.2.1: Tabla resumen características principales de variables de estudio escogidas

### 4.3. Casos de estudio

En la figura 4.3.1 se presenta un esquema de las metodologías empleadas y la forma de reducción de escenarios utilizada en este trabajo. Esta metodología es empleada en 5 escenarios de generación ERV (0%, 5%, 12%, 26% y 50%), en los cuales los escenarios con 5%, 12% y 26% de ERV se subdividen en 3 casos de estudio con diferentes localizaciones de los generadores renovables. Para hacer referencia a estos casos se utiliza la siguiente notación  $ERV_{0,5,12,25,50}^{1,2,3}$  descrita en la sección Abreviaturas.

En la figura 4.3.1 se presentan las divisiones de los diferentes casos de estudio, según nivel de generación renovable, y según localización de estas.

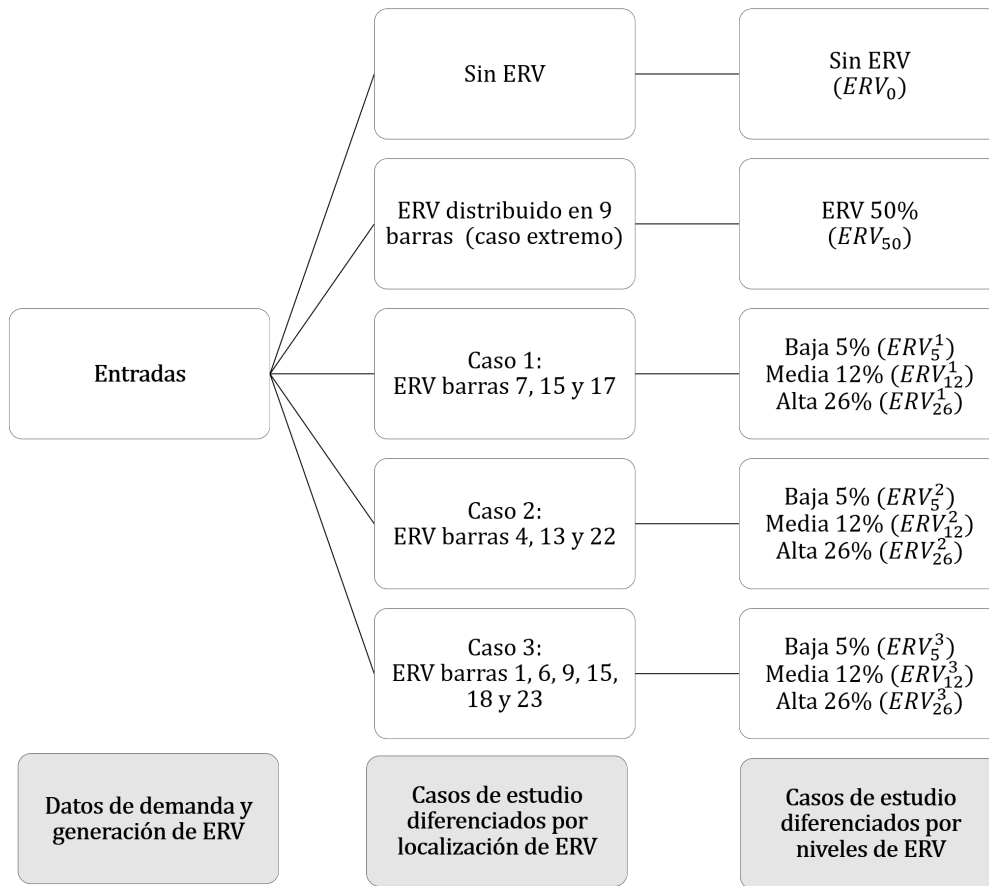


Figura 4.3.1: Subdivisión de casos de estudio por localización y nivel de generación ERV.

En las figuras 4.3.2 y 4.3.3 se presenta una especificación de las metodologías de *clustering* empleadas en cada caso de estudio, en donde se presenta el método base, con el cual serán contrastados los escenarios reducidos. Inicialmente se determinará la solución relativa a un caso base, en donde se trabajará con los bloques horarios originales (8760 horas), y las soluciones de este caso serán contrastadas posteriormente con los casos

de dimensionalidad reducida.

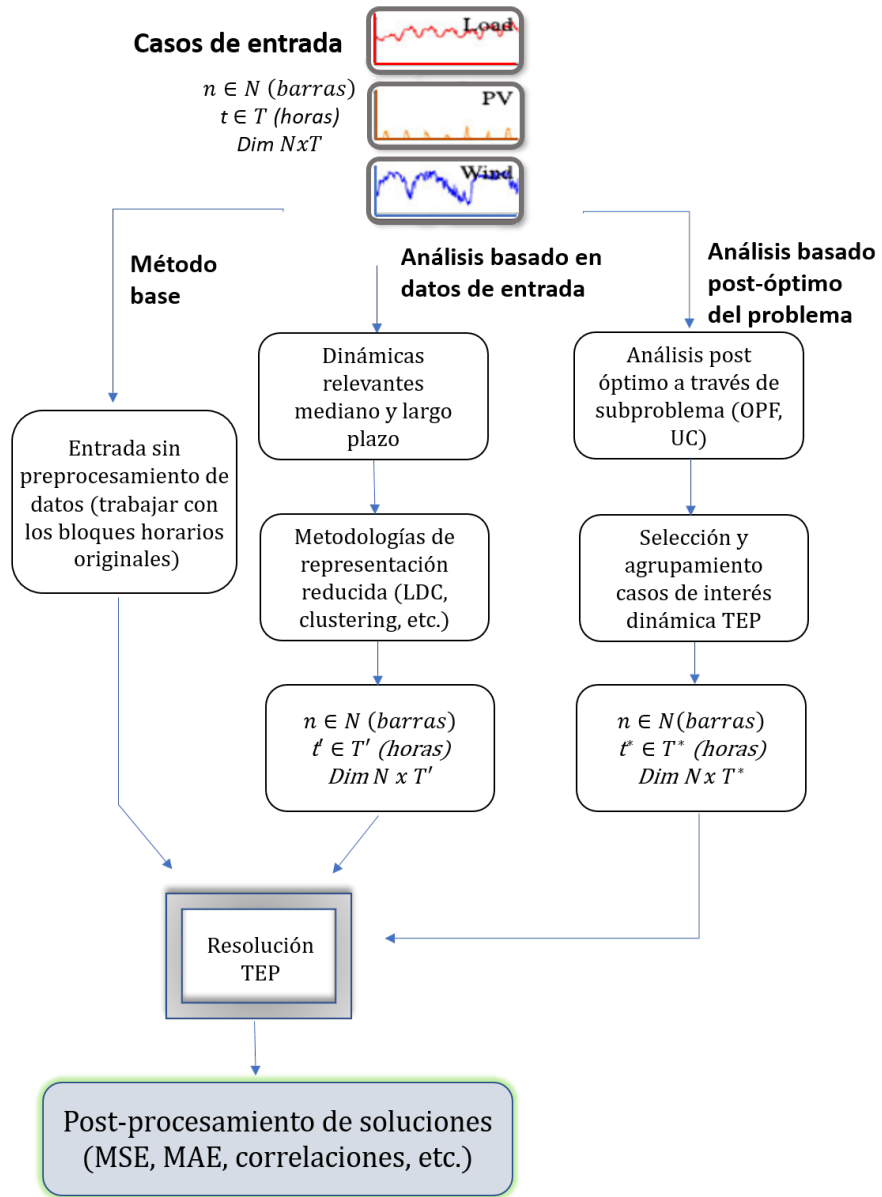


Figura 4.3.2: Esquema del trabajo propuesto.

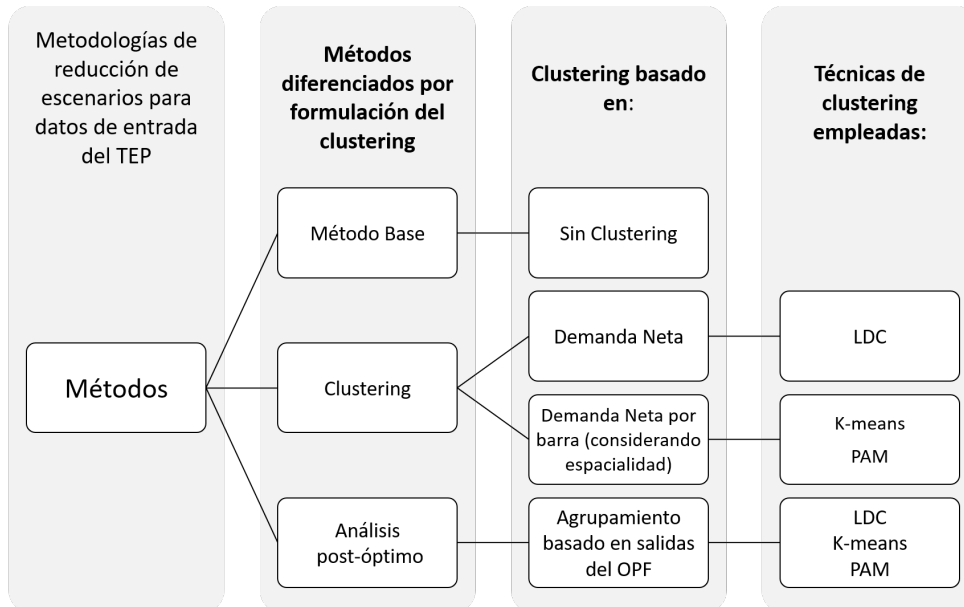


Figura 4.3.3: Esquema del trabajo propuesto.

#### 4.3.1. Datos

Se utiliza la data del sistema RTS-96, área A (24 barras) [64], considerando:

- Generadores: 32
- Barras: 24
- Tramos de líneas: 34
- Nuevos tramos por línea: 3 por línea
- VoLL: 2000 [\$/MW]
- Sbase: 100 [MVA]
- Tasa de descuento: 7 %

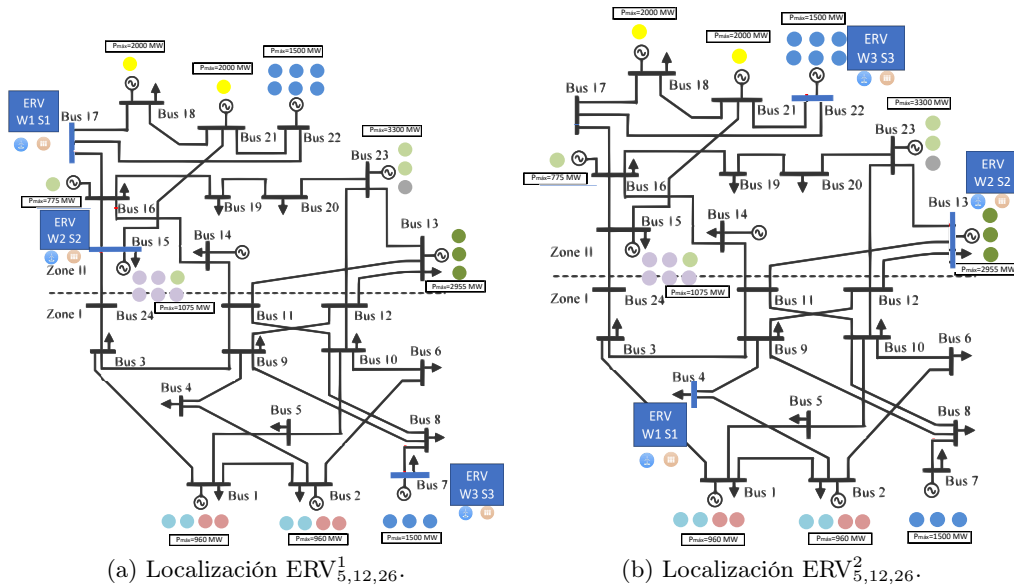
Relativo a la demanda y generación convencional, se utilizaron los datos del sistema *RTS-96*, y se realizaron las siguientes modificaciones:

- **Demanda:** Perfiles zona A *RTS-96*, multiplicados por 4.
- **Generación convencional:** Generación máxima sistema *RTS-96* multiplicada por 5.

Estas modificaciones al sistema se realizan debido a que se estresar el sistema, para que de este modo el problema de planificación de la transmisión posea un mayor número de opciones de construcción, y por ende, tener un mayor número de soluciones para

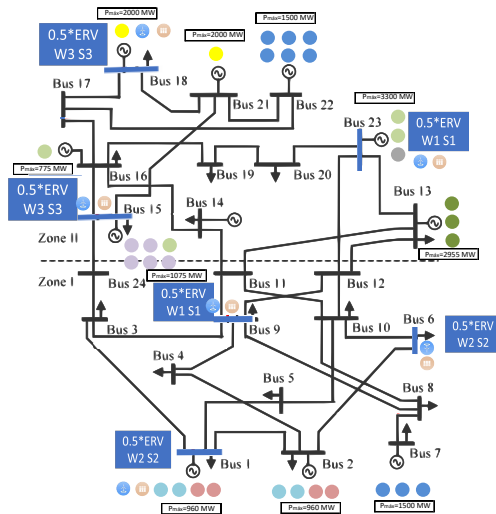
contrastar. Relativo a los costos de generación del sistema, estos se presentan en la figura 4.3.4d. En la subsección anexa 3.6.3 se presentan los datos de la respectivas líneas y generadores del sistema de estudio.

En la figura 4.3.4 se presenta el sistema de estudio, y las diferentes localizaciones de generadores ERV en estudio.



(a) Localización ERV<sup>1</sup><sub>5,12,26</sub>.

(b) Localización ERV<sup>2</sup><sub>5,12,26</sub>.



(c) Localización ERV<sup>3</sup><sub>5,12,26</sub>.

Unit type	Name	Cost \$/MWh	Color
Oil/Steam	U12	198,5	●
Oil/CT	U20	148,6	●
Hydro	U50	0,00	●
Coal/Steam	U76	50,4	●
Oil/Steam	U100	54,0	●
Coal/Steam	U155	27,0	●
Oil/Steam	U197	27,3	●
Coal/Steam	U350	3,3	●
Nuclear	U400	16,3	●
Wind	Wind	0,00	●
Solar	Solar	0,00	●

(d) Costos según tipo de generación

Figura 4.3.4: Localización de generadores ERV en el sistema y costos de operación de centrales.

4.3.1.1. Perfiles de demanda y generación ERV

En la figura 4.3.5 se presentan las correlaciones entre los diferentes perfiles de demanda según los casos de estudio. La generación renovable se considera restando su generación en la demanda de la barra en la cual está ubicada. De este modo, los perfiles de demanda por barra aumentan su variabilidad espacial y temporal. Adicionalmente, en las figuras 4.3.6 y 4.3.7 se presentan gráficamente los perfiles de demanda promedio a nivel horario, para los diferentes casos de estudio.

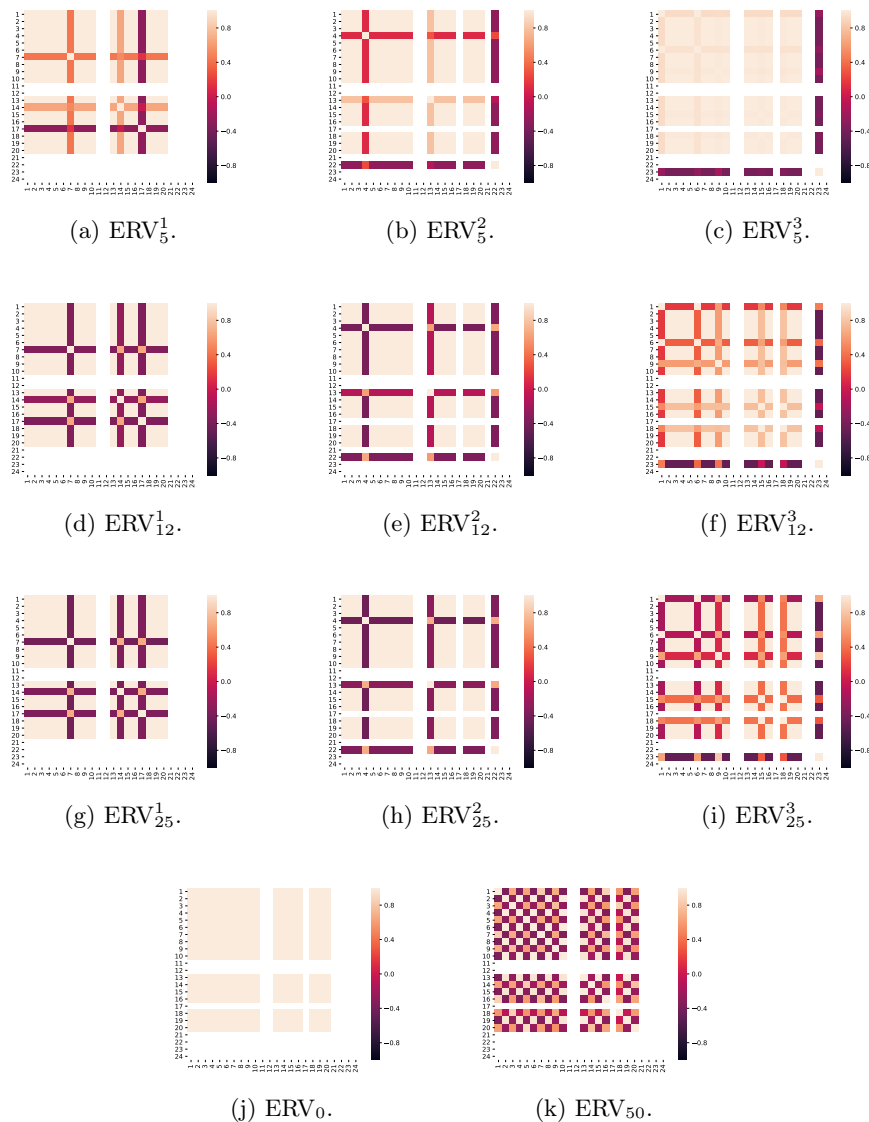
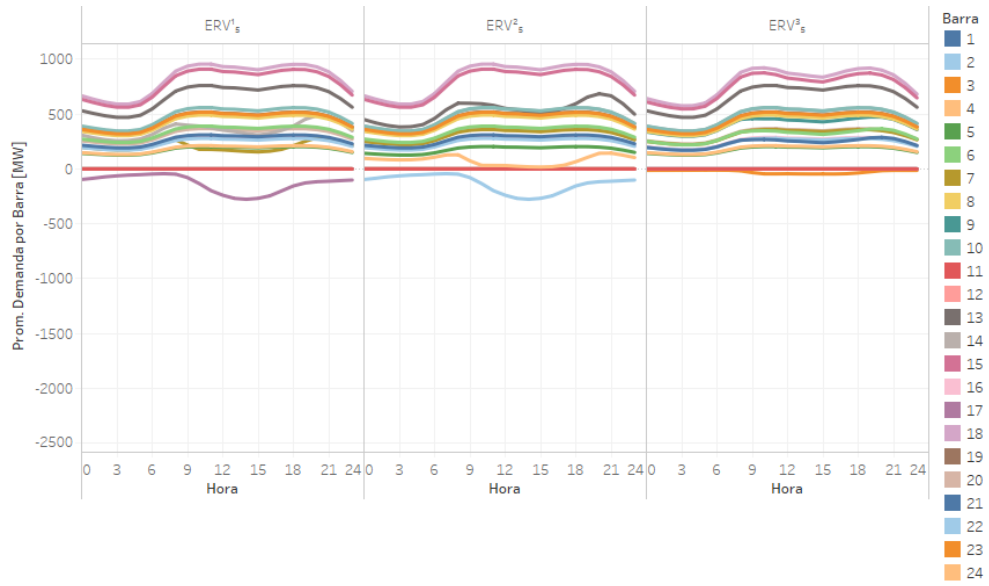
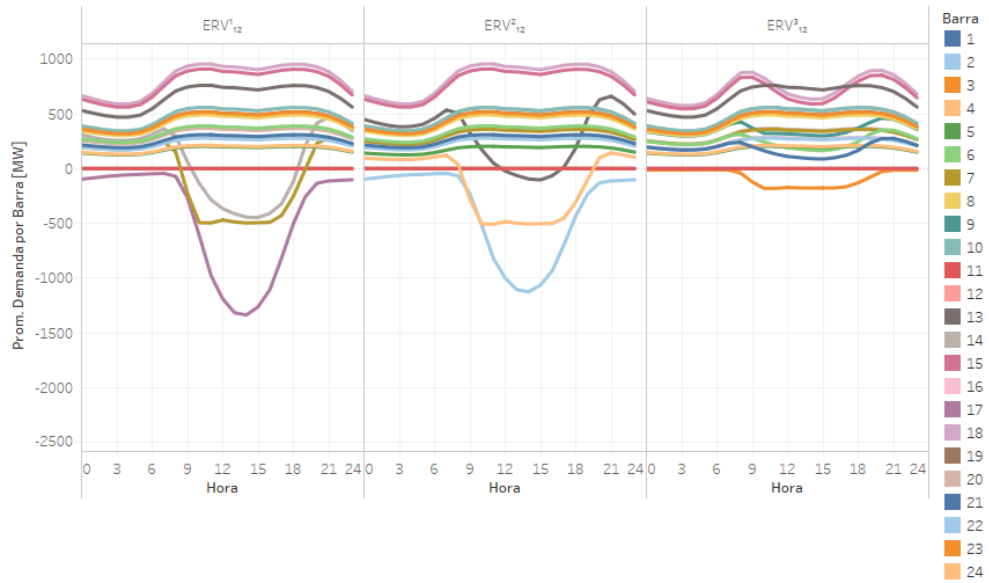


Figura 4.3.5: Gráficas de correlación espacial entre perfiles de demanda por barra.

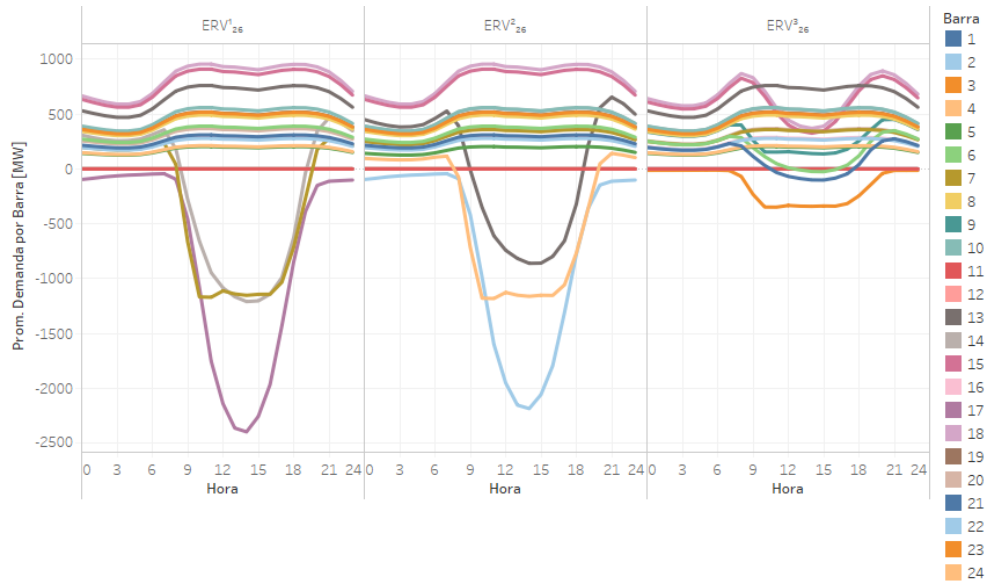


(a)  $ERV_5^{1,2,3}$ .

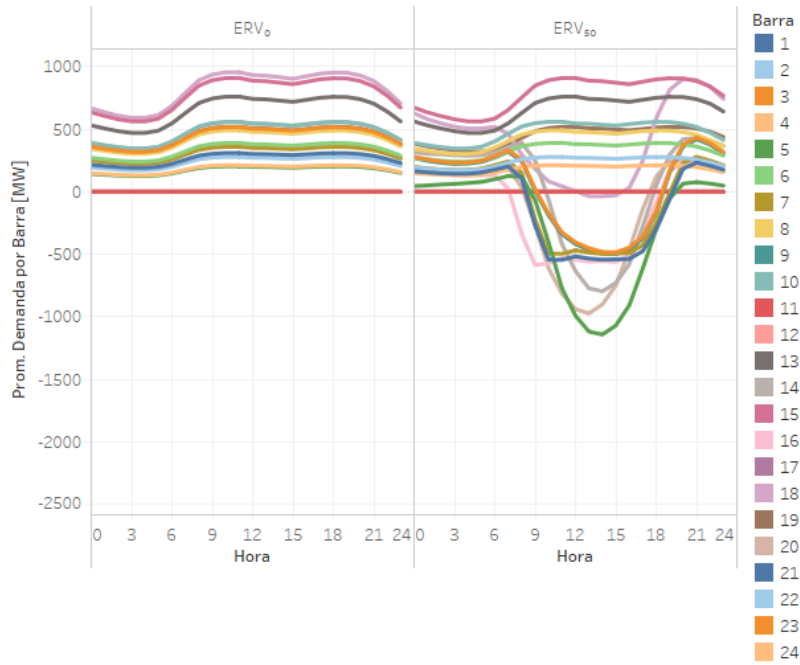


(b)  $ERV_{12}^{1,2,3}$ .

Figura 4.3.6: Gráficas de perfiles de demanda promedio por barra.



(a)  $ERV_{26}^{1,2,3}$ .



(b)  $ERV_0$  y  $ERV_{50}$ .

Figura 4.3.7: Gráficas de perfiles de demanda promedio por barra.

### 4.3.2. Selección de escenarios de estudio

- **Demanda:** Para el caso base de estudio se utilizan los perfiles de demanda dados por el sistema *RTS-96*. Luego estos son modificados de modo tal que el sistema en estudio requiera de considerable expansión de la transmisión en la planificación. Según esto, se determinó utilizar la demanda multiplicada por 4.
- **Generación convencional:** Para lograr abastecer la mayor demanda planteada en el sistema, se trabaja con los niveles de demanda dados por el sistema *RTS-96* y estos se multiplican por 5, como se observa en la tabla de la sección anexo 3.6.1.
- **Generación ERV:** Se tomaron perfiles de generación de ERV eólica y solar, y estos se adaptaron según los diferentes casos de estudio, de modo tal de proveer diferentes perfiles de demanda en el sistema, cuya variabilidad espacial fuese en aumento y se pudieran observar los fenómenos buscados en el estudio.
- **Ubicación generación ERV:** Para determinar el primer caso de ubicación de las centrales de generación ERV, primero se realiza una pequeña prueba del sistema a través de un flujo de potencia sin límite en las líneas. Además, por medio de esta información se detectan las direcciones promedio de los flujos y sus respectivos gradientes en el sistema, dimensionando también los tiempos en que se supera el límite real de capacidad de cada línea.

De este modo se observó que las direcciones de los flujos con mayor saturación eran las dadas entre las barras 17-16, 16-14, 24-15, 3-24, 3-9, 9-8 y 7-8. Así, según la dirección de los flujos obtenida se determinó una metodología de localización estratégica de fuentes renovables con el objetivo de congestionar la red, incentivando de este modo la construcción de nuevas líneas. Este consiste básicamente en posicionar la generación renovable en las barras 17, 15 y 7, como se observa en la figura 3.3.4a.

Respecto al segundo caso de ubicación de renovables, se utilizó la misma que en [89], de modo tal de poder contrastar resultados (ver figura 3.3.4b). Finalmente, el tercer caso de ubicación de renovables fue planteado buscando repartir en un mayor número de barras la generación renovable, eligiendo de este modo 9 barras (ver figura 3.3.4c).

- **Escenarios de estudio:** Dado que el objetivo de este estudio es observar el comportamiento de las diferentes técnicas de *clustering* seleccionadas sobre la demanda, se requerirá además contar con escenarios de demanda lo suficientemente variados para observar los efectos buscados. De este modo, será relevante considerar la correlación espacial entre los perfiles de demanda y sus barras, y de este modo poder analizar la importancia de esta en la selección de una técnica de *clustering* adecuada. Se considera además que la integración de energías renovables en el sistema será trabajada como una *demanda negativa*<sup>6</sup>. Así, durante el desarrollo de esta tesis se definieron inicialmente 4 perfiles de integración renovable en las 3 barras mencionadas anteriormente, estos son:

<sup>6</sup>Existen diversas formas de tratar esta información, pero resulta más práctico para el *clustering* esta simplificación.

- **Caso base:** Caso base del sistema RTS-96 que no considera integración ERV (ERV 0%).
- **ERV baja:** Se trabajan con perfiles eólicos y solares y se integran a las tres barras (ERV 5%).
- **ERV media:** En base al perfil de demanda baja, se aumenta la generación renovable solar, tal que sea 2.5 veces la inicial (ERV 12%).
- **ERV alta:** En base al perfil de demanda baja, se aumenta la generación renovable solar, tal que sea 5 veces la inicial (ERV 26%).
- **ERV muy alta:** En base al perfil de demanda baja, se aumenta la generación renovable solar, tal que sea 10 veces la inicial (ERV 50%).

#### 4.3.2.1. Métricas de evaluación de resultados

El análisis de resultados se basa en métricas comparativas de las salidas del problema de planificación. Para ello, inicialmente se realiza la planificación del problema real, con las 8760 horas del año, cuya solución es la referencia para comparar las nuevas soluciones con las técnicas empleadas.

Las métricas de evaluación de resultados utilizadas son:

- Variación respecto a la cantidad de nuevas líneas construidas: Se compara el total de nuevas líneas construidas entre el caso con 8760 horas y el caso de estudio.
- Variación en la Función Objetivo (FO): Se compara el valor final de la función objetivo entre el caso con 8760 horas y el caso de estudio. La función objetivo entrega en estos casos la suma de los costos de inversión en líneas y los costos de operación del sistema.
- MSE y MAE: Error cuadrático medio y error absoluto medio [94]. Estas métricas se utilizan para contrastar los vectores de nuevas líneas construidas. De este modo, se busca contrastar la ubicación de las nuevas líneas construidas para los diferentes casos.
- Correlación: Al igual que en el caso anterior, a través de la correlación entre los vectores solución de las nuevas líneas construidas es posible obtener una idea de cuan similares son las soluciones a comparar. En este caso un valor muy cercano a 1 es una correlación perfecta.
- Variación en la Función Objetivo Recalculada (FO recalculada): Una vez obtenida la solución del TEP con los respectivos *clusters*, esta puede dividirse en dos partes: la solución relativa a la expansión de la transmisión (líneas construidas y costos de expansión) y la solución de operación (costo de despacho de centrales y energía no servida).

Dado que, el planteamiento principal del trabajo es lograr soluciones de expansión de transmisión lo más próximas a la solución real, esta solución obtenida se mantendrá y se modificará la solución de operación, para obtener los respectivos

---

costos de operación para las 8760 horas del año. Para ello, se tomarán las solución de expansión de transmisión y se impondrán como condiciones iniciales en el OPF para las 8760 horas, con la demanda original del sistema (sin *cluster*). De este modo, la solución de operación será más realista, ya que considerará la operación real del sistema, con lo que el valor de la función objetivo podrá ser una buena métrica de comparación de soluciones.

### 4.4. Resultados numéricos

Las simulaciones del TEP se realizan para 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 24, 30, 40, 60, 73 y 120 *clusters*, en 11 casos de estudio. En cada caso de estudio se analizó la opción de seleccionar los *clusters* de demanda a partir de 6 diferentes variables de estudio, a través de un pre-procesamiento de datos de entrada, las cuales a su vez fueron agregadas temporalmente a través de 3 técnicas de *clustering* distintas: LDC, k-means y PAM, con el objetivo de comparar resultados.

El pre-procesamiento de los datos de entrada y la generación de los respectivos *clusters* es desarrollado a través de *R* [95], y la respectiva simulación del problema de expansión de la transmisión se desarrollada en *PYOMO (Python)* [96, 97], a través de una formulación no lineal, por lo que el *solver* utilizado para la resolución del problema fue *Ipopt* [47]. El computador utilizado para el trabajo en *R* y la simulación de los casos de estudios con reducción dimensional de la demanda fue un Intel  $\text{\textcircled{R}}$ Xeon  $\text{\textcircled{R}}$ X5570, 2 procesadores de 2.30 GHz y 64 Gb de RAM.

Relativo al tiempo requerido para la simulación en PYOMO con las 8760 horas, este no es del todo comparable con los casos reducidos, debido a que por su gran requerimiento computacional se debió utilizar un servidor externo y diferentes configuraciones para su simulación. Para ello se utilizó el solver *Ipopt + HSL MA86* [98], y el servidor externo utilizado fue un nodo *cluster quartz*<sup>7</sup> (Intel  $\text{\textcircled{R}}$ Xeon  $\text{\textcircled{R}}$ E5-2695, 344,064 Gb de RAM, CPU Memory/Node de 128 Gb), y bajo estas condiciones los tiempos de simulación promedio fueron de media hora<sup>8</sup>.

Tabla 4.4.1: Tiempos de simulación para algunos *clusters* del caso  $ERV_0$

Método	<i>N</i> <sup>o</sup> <i>clusters</i>	Tiempo de <i>clustering</i> en <i>R</i>					Tiempo simulación TEP por cluster (Resolución PYOMO)	
		Cop (con)	Cop (sin)	Flujos (con)	Flujos (sin)	Alternativo	Promedio cluster	
LDC	1	0.08	0.07	0.07	0.07	0.07	0.20	
	10	0.09	0.06	0.08	0.08	0.08	1.63	
	20	0.11	0.10	0.08	0.06	0.07	3.14	
	40	0.09	0.07	0.09	0.08	0.09	8.31	
	60	0.11	0.09	0.11	0.10	0.09	12.99	
	120	0.12	0.13	0.14	0.14	0.12	29.14	
k-means	1	0.06	0.06	0.08	0.06	0.08	0.17	
	10	0.07	0.06	0.08	0.10	0.12	1.62	
	20	0.10	0.09	0.11	0.11	0.12	2.98	
	40	0.09	0.09	0.16	0.10	0.11	8.16	
	60	0.11	0.09	0.18	0.14	0.19	13.42	
	120	0.11	0.13	0.20	0.19	0.22	33.32	
PAM	1	1.56	1.62	4.63	4.67	4.59	0.18	
	10	265.69	333.43	238.36	260.12	407.61	1.64	
	20	1,742.52	1,896.31	1,818.72	795.75	1,372.87	3.22	
	40	3,546.71	8,885.56	2,908.35	4,546.21	8,140.35	8.03	
	60	9,389.22	9,220.26	11,795.17	10,082.88	10,999.44	13.43	
	120	26,012.41	36,330.47	33,400.26	30,798.83	37,190.21	30.71	

<sup>7</sup><https://hpc.llnl.gov/hardware/platforms/Quartz>

<sup>8</sup>Esta simulación fue posible gracias al Lawrence Livermore National Laboratory de Estados Unidos de América

Tabla 4.4.2: Tiempos de simulación para algunos *clusters* del caso ERV<sub>50</sub>

Método	N° <i>clusters</i>	Tiempo de <i>clustering</i> en R					Tiempo simulación TEP por cluster (Resolución PYOMO)	
		Cop (con)	Cop (sin)	Flujos (con)	Flujos (sin)	Alternativo	Promedio cluster	
LDC	1	0.06	0.06	0.06	0.08	0.05	0.15	
	10	0.09	0.06	0.06	0.07	0.09	1.29	
	20	0.11	0.07	0.10	0.10	0.08	2.97	
	40	0.13	0.09	0.13	0.14	0.10	7.95	
	60	0.15	0.11	0.11	0.11	0.13	12.77	
	120	0.13	0.12	0.14	0.16	0.14	29.75	
k-means	1	0.05	0.06	0.06	0.06	0.08	0.14	
	10	0.07	0.08	0.13	0.11	0.16	1.29	
	20	0.12	0.11	0.18	0.17	0.20	2.75	
	40	0.18	0.13	0.24	0.24	0.32	6.57	
	60	0.12	0.14	0.30	0.25	0.40	10.26	
	120	0.15	0.14	0.48	0.45	0.75	26.11	
PAM	1	1.55	1.51	4.53	4.59	4.49	0.14	
	10	619.18	381.46	146.51	126.35	126.21	1.39	
	20	1,385.93	2,413.99	422.56	407.41	532.87	2.47	
	40	5,463.91	6,176.14	2,998.08	2,391.64	2,412.73	7.16	
	60	21,708.55	14,001.48	6,213.72	5,546.39	3,688.79	11.55	
	120	56,027.53	39,359.05	22,571.72	26,209.14	21,227.64	25.99	

En las tablas 4.4.1 y 4.4.2 se representan los tiempos de simulación para dos casos de estudio, de modo tal de contrastar el recurso computacional necesario para el pre-procesamiento de la data y la diferencia de tiempo obtenida al emplear las metodologías propuestas.

De manera general se observa que para LDC y k-means los tiempos de agregación temporal a través del *clustering* en R son del orden de las décimas de segundo, al contrario del PAM, en donde el tiempo de *clustering* crece de manera exponencial, llegando al orden de las horas para más de 40 *clusters*. Respecto a los tiempos de simulación en *PYOMO*, estos son muy bajos en casi todos los casos de estudio, reduciendo de manera considerable el requerimiento computacional versus el caso de estudio con las 8760 horas.

A continuación, se presentan gráficas de las soluciones obtenidas respecto a su métrica de evaluación seleccionada y tablas resumen respecto a los resultados para 120 *clusters*. En ellas se presenta la variación porcentual respecto a la función objetivo, el MSE y la función objetivo recalculada. Adicionalmente, en la sección anexos se presenta el resto de las gráficas y tablas respecto a la variación porcentual de nuevas líneas construidas, MAE y correlación entre vectores de nuevas líneas construidas.

Posteriormente, en base a los análisis de las gráficas, se presentan de manera comparativa los casos de estudio con mejor comportamiento para su comparación de manera más detallada y una tabla resumen del *ranking* de las mejores técnicas desarrolladas.

4.4.1. Análisis respecto a la Función Objetivo (FO)

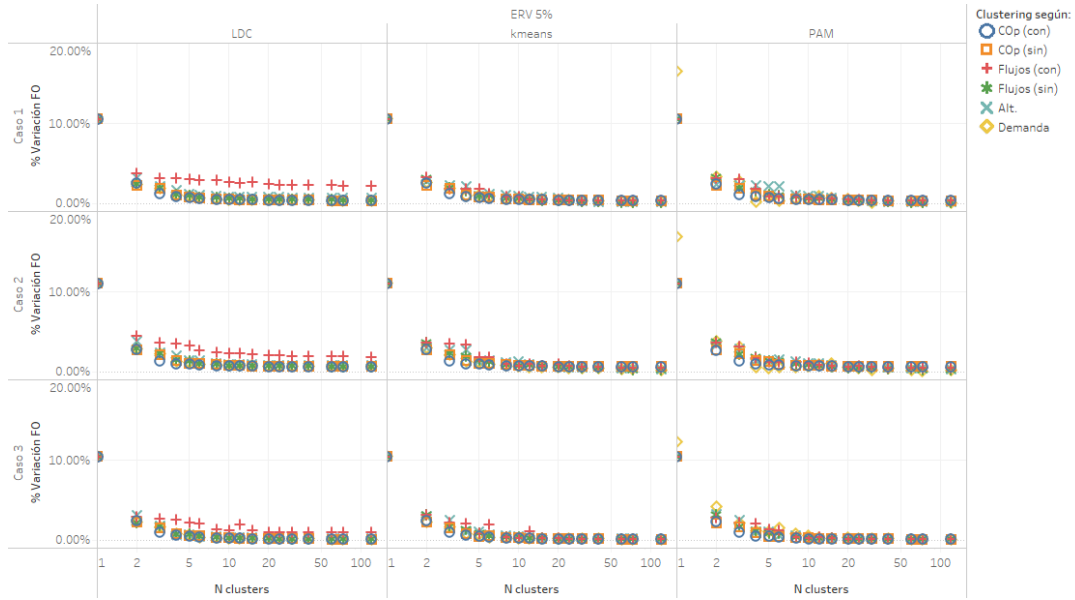


Figura 4.4.1: Error % FO para  $ERV_5^1$ ,  $ERV_5^2$  y  $ERV_5^3$

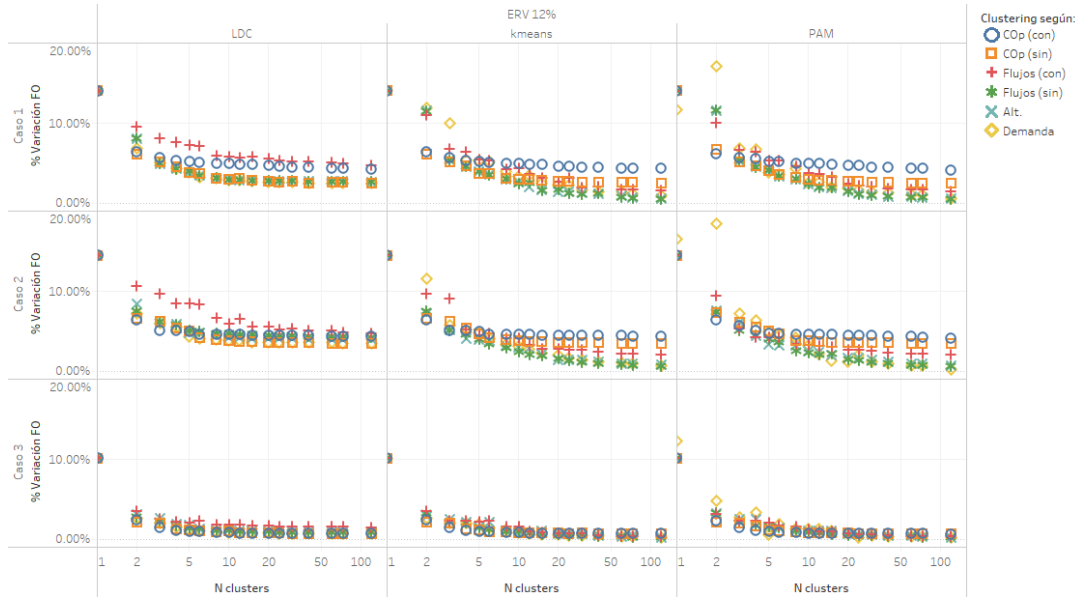


Figura 4.4.2: Error % FO para  $ERV_{12}^1$ ,  $ERV_{12}^2$  y  $ERV_{12}^3$

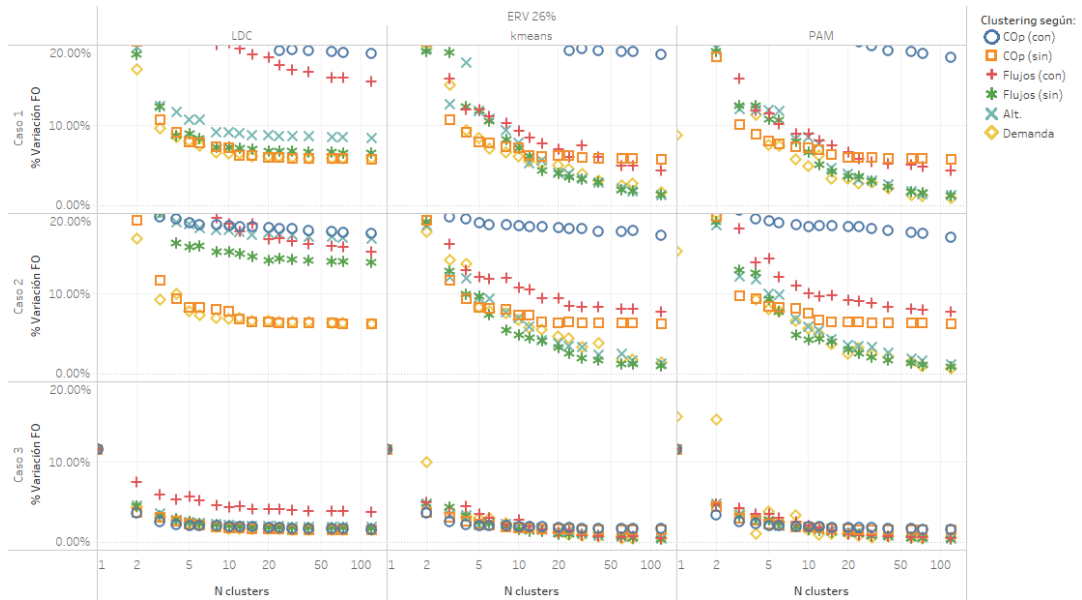


Figura 4.4.3: Error % FO para  $ERV_{26}^1$ ,  $ERV_{26}^2$  y  $ERV_{26}^3$

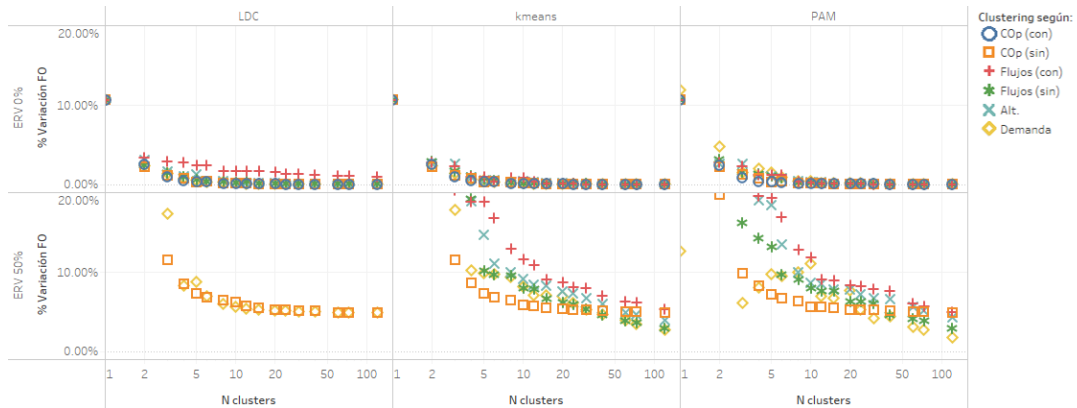


Figura 4.4.4: Error % FO para  $ERV_0$  y  $ERV_{50}$

El análisis de los gráficos respecto a la función objetivo indica que el comportamiento de las metodologías es relativamente bueno en todos los casos para muy baja variabilidad espacial, como lo es en el caso  $ERV_0$ . No así, la principal diferencia de resultados se aprecia en el caso de alta variabilidad  $ERV_{50}$ , en donde las soluciones no tienden a converger a un punto común, presentando porcentajes de error muy altos versus el resto de los casos de estudio.

Acorde a lo esperado, para el agrupamiento a partir de los flujos con límites de capacidad de líneas se tienen resultados con altos porcentajes de error en casi todos los casos de estudio.

Para el caso  $ERV_{12}^{1,2,3}$ , se aprecia que los mejores resultados son para los casos de K-means y PAM, agrupando los datos a partir de la demanda y de los flujos sin límites

Tabla 4.4.3: Variación porcentual de la FO para 120 *clusters*

$N^\circ$ <i>clusters</i>	% ERV	Caso	Método	Clustering Cop (con)	Clustering COp (sin)	Clustering Flujos (con)	Clustering Flujos (sin)	Clustering Alternativo	Clustering según demanda
120	0%	-	LDC	0.00%	0.00%	0.99%	0.04%	0.05%	0.00%
			K-means	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
			PAM	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	Caso 1	LDC	0.34%	0.31%	2.16%	0.38%	0.65%	0.31%	
		K-means	0.34%	0.32%	0.27%	0.10%	0.11%	0.13%	
		PAM	0.34%	0.31%	0.27%	0.10%	0.09%	0.04%	
	5%	Caso 2	LDC	0.58%	0.63%	1.80%	0.61%	0.75%	0.62%
			K-means	0.59%	0.61%	0.41%	0.20%	0.26%	0.16%
			PAM	0.58%	0.63%	0.41%	0.21%	0.26%	0.14%
	Caso 3	LDC	0.08%	0.08%	0.88%	0.12%	0.16%	0.08%	
		K-means	0.08%	0.08%	0.05%	0.04%	0.05%	0.04%	
		PAM	0.08%	0.08%	0.05%	0.03%	0.04%	0.02%	
	Caso 1	LDC	4.22%	2.45%	4.70%	2.60%	2.60%	2.45%	
		K-means	4.25%	2.43%	1.50%	0.50%	0.52%	0.84%	
		PAM	4.08%	2.43%	1.42%	0.46%	0.46%	0.55%	
	12%	Caso 2	LDC	4.26%	3.42%	4.65%	4.20%	3.83%	3.41%
			K-means	4.27%	3.40%	2.05%	0.58%	0.71%	0.62%
			PAM	4.13%	3.40%	2.02%	0.54%	0.61%	0.15%
	Caso 3	LDC	0.65%	0.65%	1.47%	0.79%	0.85%	0.64%	
		K-means	0.64%	0.65%	0.32%	0.17%	0.21%	0.16%	
		PAM	0.62%	0.64%	0.27%	0.18%	0.20%	0.18%	
	Caso 1	LDC	18.85%	5.76%	15.42%	6.46%	8.38%	5.76%	
		K-means	18.79%	5.75%	4.27%	1.35%	1.29%	1.62%	
		PAM	18.42%	5.74%	4.30%	1.14%	1.18%	0.90%	
26%	Caso 2	LDC	17.47%	6.12%	15.21%	13.80%	16.83%	6.16%	
		K-means	17.27%	6.15%	7.73%	0.90%	1.26%	1.33%	
		PAM	16.92%	6.16%	7.68%	0.82%	1.09%	0.65%	
Caso 3	LDC	1.57%	1.46%	3.73%	1.70%	1.80%	1.46%		
	K-means	1.61%	1.47%	0.44%	0.37%	0.43%	0.34%		
	PAM	1.51%	1.45%	0.42%	0.35%	0.43%	0.47%		
50%	-	LDC	36.96%	4.81%	35.24%	23.12%	27.85%	4.82%	
		K-means	37.26%	4.80%	5.30%	2.87%	3.87%	2.66%	
		PAM	36.75%	4.90%	4.93%	2.88%	4.30%	1.72%	

de capacidad de líneas. Luego, para el caso  $ERV_{26}^{1,2,3}$ , se observa una mayor variabilidad respecto a las soluciones obtenidas, siendo en este caso bueno el agrupamiento por demanda, flujo sin límite de capacidad y el flujo alternativo.

Finalmente, para el caso  $ERV_{50}$ , se aprecia que el menor error absoluto se obtiene para el *clustering* según demanda, a través de PAM.

4.4.2. Análisis respecto al MSE

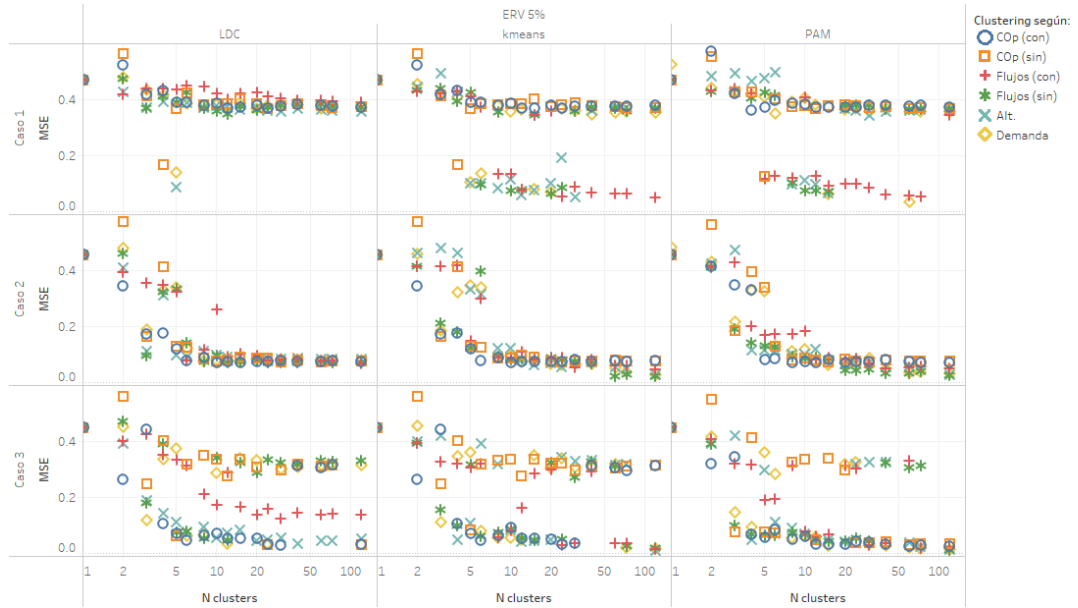


Figura 4.4.5: MSE para  $ERV_5^1$ ,  $ERV_5^2$  y  $ERV_5^3$

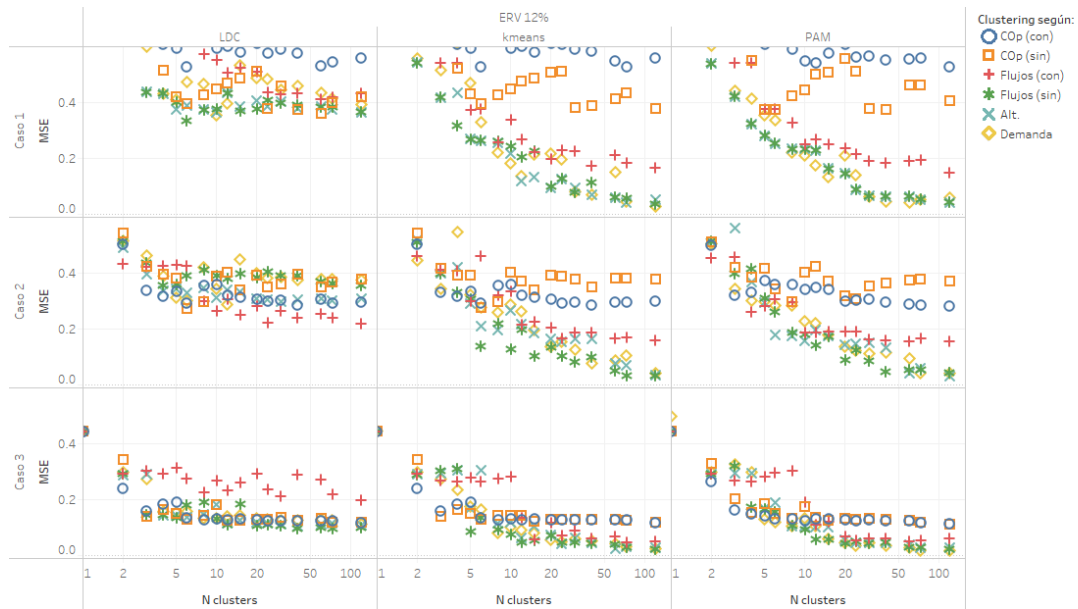


Figura 4.4.6: MSE para  $ERV_{12}^1$ ,  $ERV_{12}^2$  y  $ERV_{12}^3$

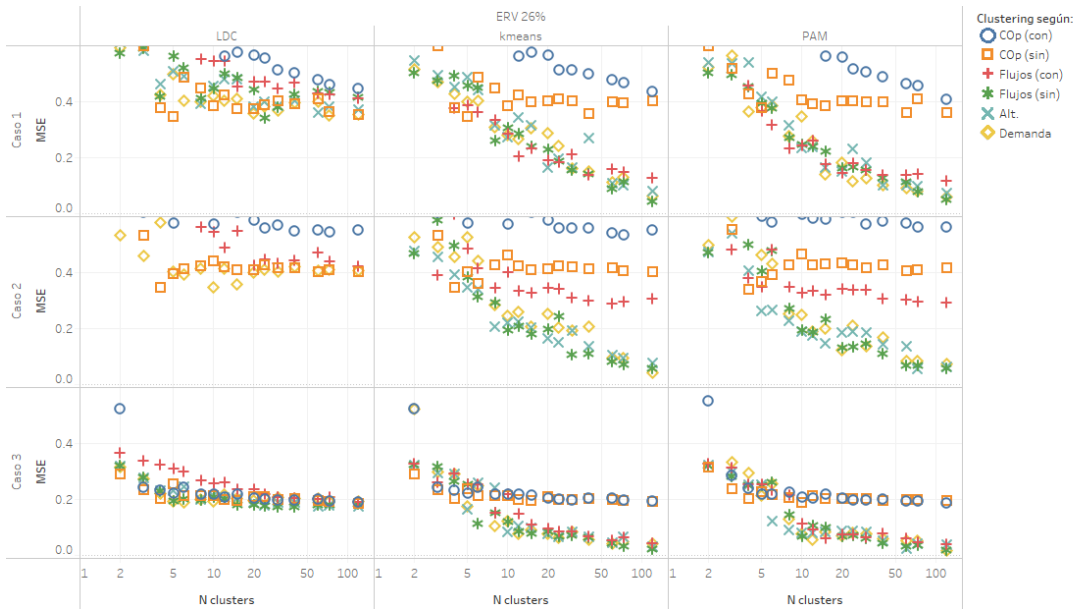


Figura 4.4.7: MSE para  $ERV_{26}^1$ ,  $ERV_{26}^2$  y  $ERV_{26}^3$

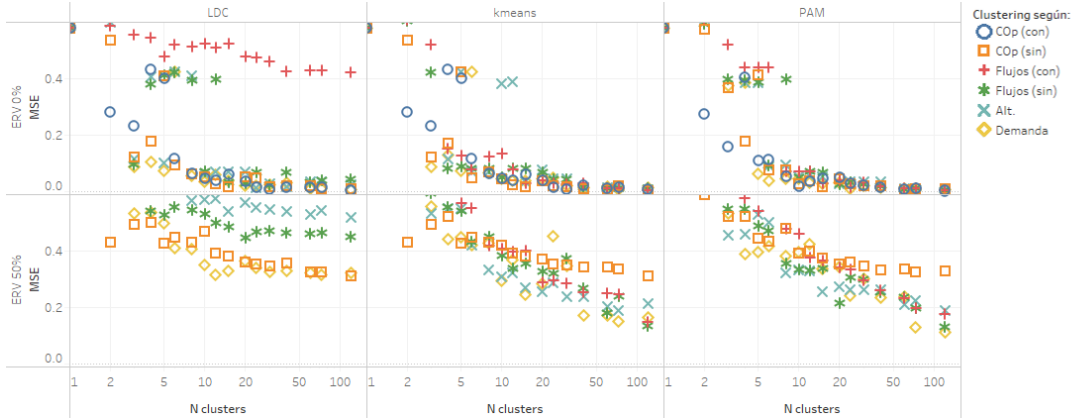


Figura 4.4.8: MSE para  $ERV_0$  y  $ERV_{50}$

Continuando con el análisis respecto al MSE, es aquí más visible la variación de soluciones de los vectores de nuevas líneas construidas, y por tanto, refleja directamente las localizaciones de las decisiones de inversión en transmisión del sistema.

En los casos  $ERV_0$  y  $ERV_5^{1,2,3}$  se aprecian variados resultados respecto a los vectores solución. Esto es debido a que la baja variabilidad del sistema conduce a una mayor gama de opciones de expansión, a diferencia de los casos con mayor nivel de penetración renovable, en donde las saturaciones de líneas son más localizadas, y por tanto los requerimientos del sistemas son más enfocados en esas líneas.

Para el resto de los casos de estudio, se observa que el agrupamiento basado en flujos alternativos, flujos sin límite de capacidad y demanda son los más representativos.

Tabla 4.4.4: MSE para 120 clusters

N° clusters	% ERV	Caso	Método	Clustering Cop (con)	Clustering COp (sin)	Clustering Flujos (con)	Clustering Flujos (sin)	Clustering Alternativo	Clustering según demanda	
120	0%	-	LDC	0.0050	0.0103	0.4214	0.0435	0.0222	0.0044	
			K-means	0.0065	0.0037	0.0048	0.0095	0.0066	0.0119	
			PAM	0.0032	0.0077	0.0044	0.0052	0.0074	0.0056	
		Caso 1	LDC	0.3728	0.3740	0.3911	0.3738	0.3580	0.3703	
			K-means	0.3778	0.3691	0.0495	0.3685	0.3669	0.3520	
			PAM	0.3725	0.3661	0.3440	0.3640	0.3657	0.3571	
		5%	Caso 2	LDC	0.0739	0.0766	0.0684	0.0696	0.0844	0.0783
				K-means	0.0779	0.0743	0.0473	0.0205	0.0252	0.0253
				PAM	0.0721	0.0754	0.0510	0.0250	0.0276	0.0312
	Caso 3	LDC	0.0307	0.0320	0.1389	0.3305	0.0521	0.3147		
		K-means	0.3110	0.3154	0.0157	0.0165	0.0101	0.0173		
		PAM	0.0266	0.0322	0.0200	0.0103	0.0136	0.0119		
	12%	Caso 1	LDC	0.5566	0.4210	0.4317	0.3650	0.3637	0.3909	
			K-means	0.5564	0.3786	0.1661	0.0344	0.0510	0.0256	
			PAM	0.5252	0.4070	0.1477	0.0416	0.0416	0.0573	
	Caso 2	LDC	0.2967	0.3791	0.2171	0.3533	0.3080	0.3735		
		K-means	0.2996	0.3765	0.1585	0.0307	0.0355	0.0403		
		PAM	0.2820	0.3712	0.1554	0.0412	0.0317	0.0402		
	Caso 3	LDC	0.1178	0.1194	0.1977	0.1003	0.1076	0.1109		
		K-means	0.1178	0.1164	0.0510	0.0228	0.0266	0.0225		
		PAM	0.1119	0.1102	0.0586	0.0230	0.0280	0.0152		
	26%	Caso 1	LDC	0.4461	0.3532	0.4101	0.4149	0.3688	0.3533	
			K-means	0.4340	0.4026	0.1252	0.0438	0.0803	0.0614	
			PAM	0.4058	0.3603	0.1155	0.0496	0.0712	0.0582	
	Caso 2	LDC	0.5497	0.4036	0.4199	0.6815	0.6449	0.4054		
		K-means	0.5515	0.4024	0.3069	0.0563	0.0754	0.0411		
		PAM	0.5612	0.4161	0.2906	0.0562	0.0621	0.0721		
	Caso 3	LDC	0.1898	0.1845	0.1898	0.1829	0.1740	0.1928		
		K-means	0.1937	0.1911	0.0431	0.0210	0.0314	0.0400		
		PAM	0.1856	0.1956	0.0399	0.0168	0.0389	0.0146		
50%	-	LDC	0.6369	0.3102	0.6248	0.4479	0.5174	0.3216		
		K-means	0.6308	0.3124	0.1474	0.1355	0.2121	0.1648		
		PAM	0.6098	0.3272	0.1753	0.1296	0.1872	0.1099		

4.4.3. Análisis respecto a la Función Objetivo Recalculada



Figura 4.4.9: Error % FO recalculada para  $ERV_{\frac{1}{5}}$ ,  $ERV_{\frac{2}{5}}$  y  $ERV_{\frac{3}{5}}$

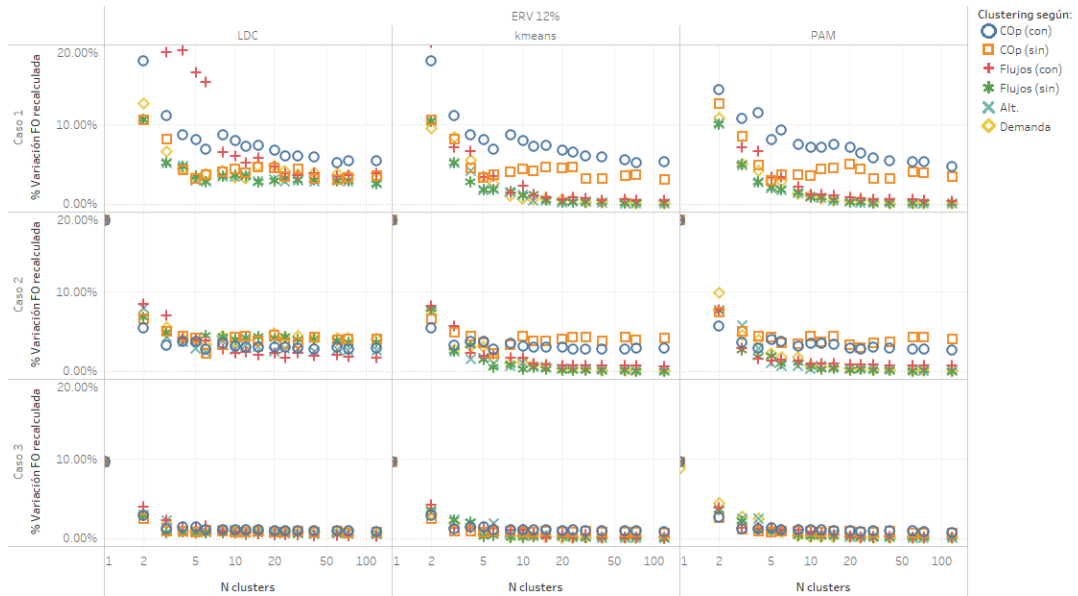


Figura 4.4.10: Error % FO recalculada para  $ERV_{12}^1$ ,  $ERV_{12}^2$  y  $ERV_{12}^3$

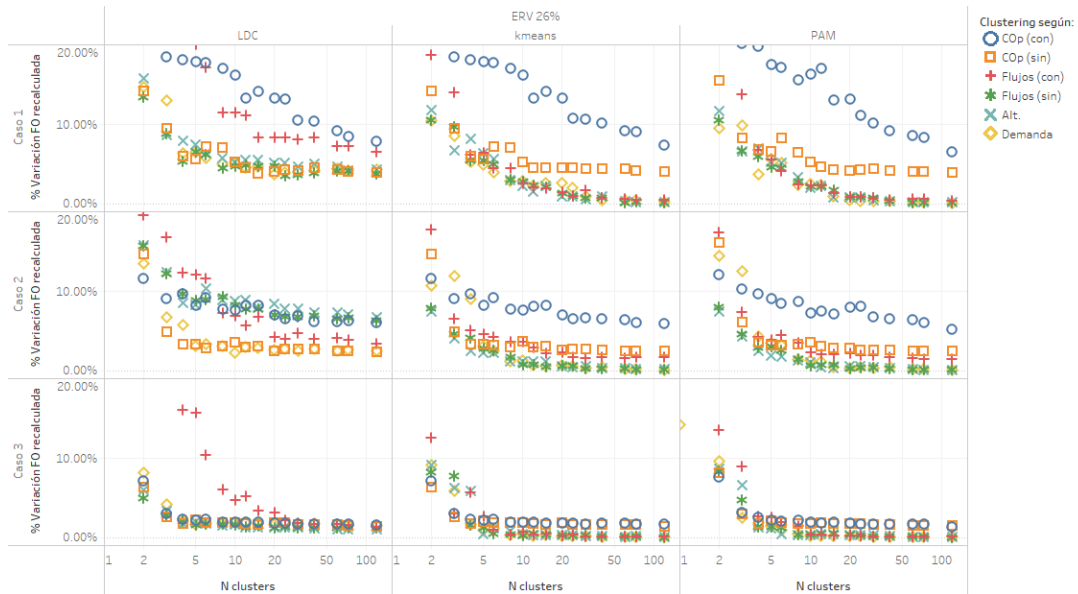


Figura 4.4.11: Error % FO recalculada para  $ERV_{26}^1$ ,  $ERV_{26}^2$  y  $ERV_{26}^3$

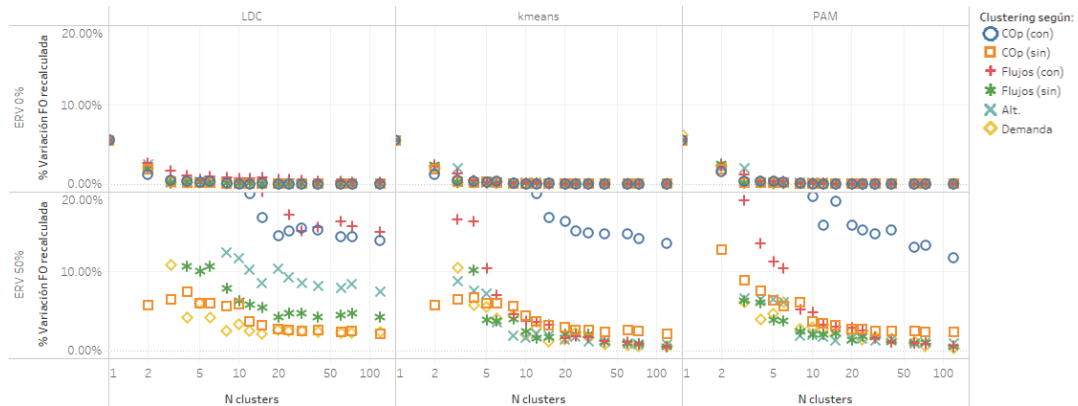


Figura 4.4.12: Error % FO recalculada para  $ERV_0$  y  $ERV_{50}$

Tabla 4.4.5: Variación porcentual de FO recalculada para 120 clusters

$N^\circ$ clusters	% ERV	Caso	Método	Clustering Cop (con)	Clustering COp (sin)	Clustering Flujos (con)	Clustering Flujos (sin)	Clustering Alternativo	Clustering según demanda
120	0%	-	LDC	0.00%	0.00%	0.23%	0.00%	0.01%	0.00%
			K-means	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
			PAM	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	5%	Caso 1	LDC	0.23%	0.22%	0.83%	0.19%	0.24%	0.22%
			K-means	0.23%	0.24%	0.07%	0.04%	0.03%	0.01%
			PAM	0.21%	0.21%	0.07%	0.03%	0.03%	0.02%
	5%	Caso 2	LDC	0.49%	0.53%	0.30%	0.45%	0.43%	0.55%
			K-means	0.50%	0.52%	0.17%	0.02%	0.04%	0.03%
			PAM	0.46%	0.54%	0.18%	0.02%	0.05%	0.03%
	5%	Caso 3	LDC	0.04%	0.03%	0.25%	0.05%	0.07%	0.04%
			K-means	0.04%	0.04%	0.01%	0.01%	0.01%	0.01%
			PAM	0.03%	0.03%	0.01%	0.01%	0.01%	0.01%
	12%	Caso 1	LDC	5.34%	3.47%	3.79%	2.54%	2.51%	3.31%
			K-means	5.29%	3.01%	0.46%	0.04%	0.06%	0.02%
			PAM	4.73%	3.36%	0.35%	0.03%	0.03%	0.02%
	12%	Caso 2	LDC	2.82%	4.05%	1.67%	3.60%	2.61%	3.99%
			K-means	2.89%	4.16%	0.60%	0.02%	0.02%	0.03%
			PAM	2.67%	4.03%	0.64%	0.02%	0.02%	0.02%
12%	Caso 3	LDC	0.84%	0.69%	0.24%	0.36%	0.56%	0.59%	
		K-means	0.81%	0.65%	0.07%	0.02%	0.04%	0.01%	
		PAM	0.65%	0.58%	0.05%	0.03%	0.03%	0.00%	
26%	Caso 1	LDC	7.78%	3.86%	6.48%	3.72%	4.27%	4.24%	
		K-means	7.34%	4.04%	0.42%	0.06%	0.09%	0.11%	
		PAM	6.49%	3.89%	0.39%	0.05%	0.12%	0.08%	
26%	Caso 2	LDC	6.00%	2.35%	3.30%	6.29%	6.62%	2.39%	
		K-means	5.94%	2.43%	1.60%	0.05%	0.10%	0.04%	
		PAM	5.11%	2.43%	1.44%	0.05%	0.07%	0.03%	
26%	Caso 3	LDC	1.58%	1.30%	1.28%	1.16%	0.96%	1.42%	
		K-means	1.65%	1.38%	0.03%	0.02%	0.04%	0.03%	
		PAM	1.35%	1.51%	0.05%	0.02%	0.05%	0.01%	
50%	-	LDC	13.88%	2.09%	14.88%	4.20%	7.39%	2.27%	
		K-means	13.46%	2.07%	0.45%	0.47%	0.78%	0.50%	
		PAM	11.72%	2.28%	0.57%	0.45%	0.91%	0.26%	

Como se mencionó en el capítulo anterior, la métrica “FO recalculada” entrega una comparación más realista respecto a los beneficios de utilizar directamente la función

objetivo, dado que es posible detectar de mejor manera las pequeñas variaciones en cuanto a las decisiones de inversión en el sistema. Su principal característica es el lograr comparar las soluciones bajo un nivel de detalle temporal común, siendo con ello más comparables los costos asociados a la expansión de transmisión en la nueva función objetivo.

La principal diferencia respecto al cálculo inicial de la FO, es que se aprecia desde un menor número de *clusters* la diferencia real de soluciones entre metodologías. Así, para 10 *clusters* se observa de forma general que las soluciones alcanzan su tendencia final.

Relativo a las metodologías en las que se detectaron bajos errores relativos, estas están en un *ranking* similar al caso FO, sin embargo la caída de la línea de tendencia de las gráficas respecto a los *clusters* es más marcada, dada la representación temporal lograda.

#### 4.4.4. Revisión detallada de casos particulares seleccionados

En base a los resultados presentados anteriormente, se seleccionan los casos k-means y PAM para el *clustering* basado en la demanda, flujos sin límites de capacidad de línea y el *clustering* alt. (alternativo), dado que son los casos que a priori presentan menor error según las diferentes métricas. Se busca analizar de una forma más detallada las soluciones y contrastar respecto a sus ventajas y desventajas.

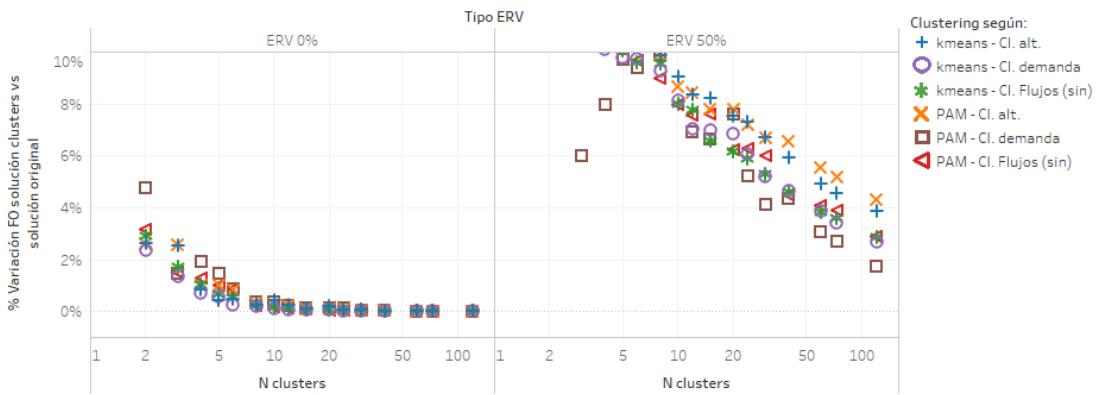


Figura 4.4.13: Error % FO para ERV<sub>0</sub> y ERV<sub>50</sub>

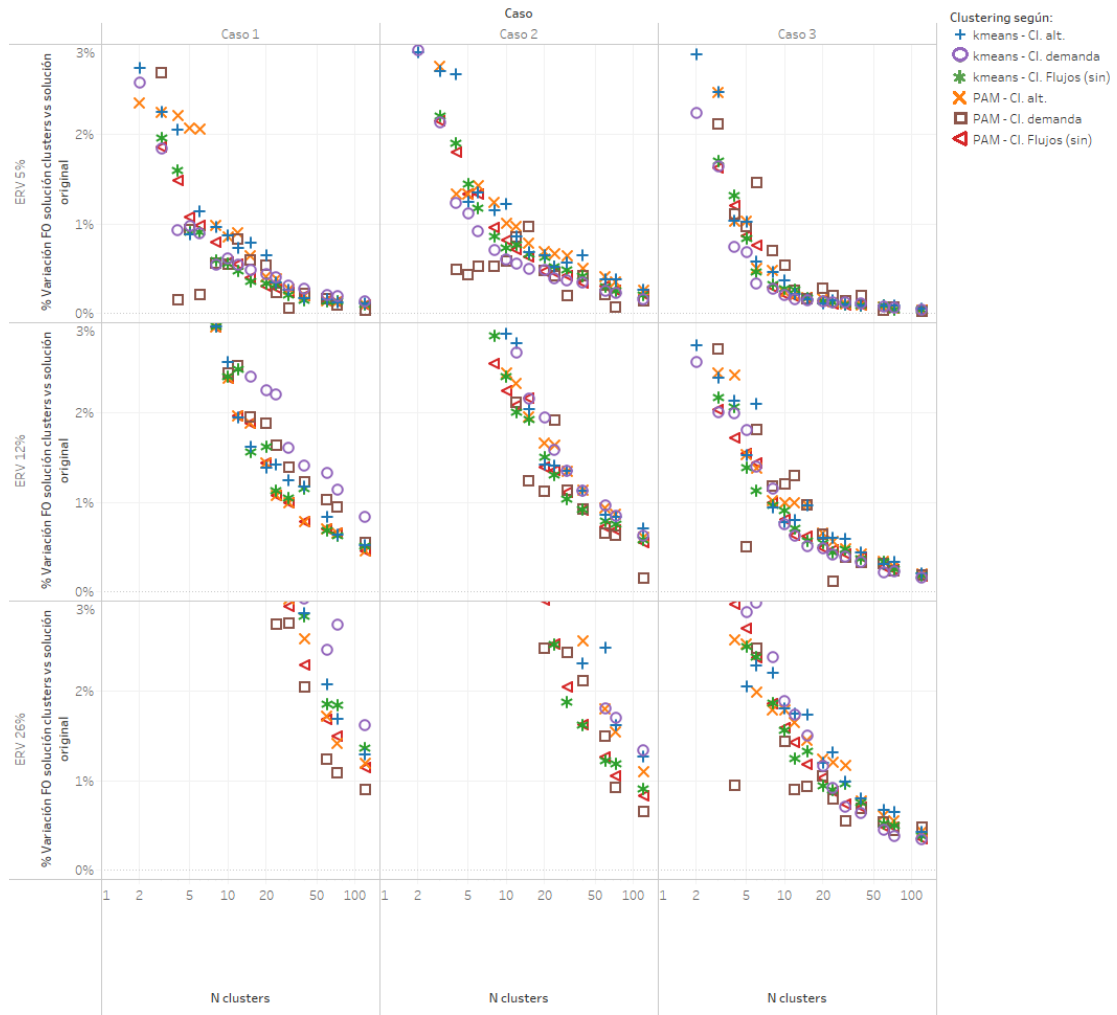


Figura 4.4.14: Error % FO para  $ERV_{5,12,25}^{1,2,3}$

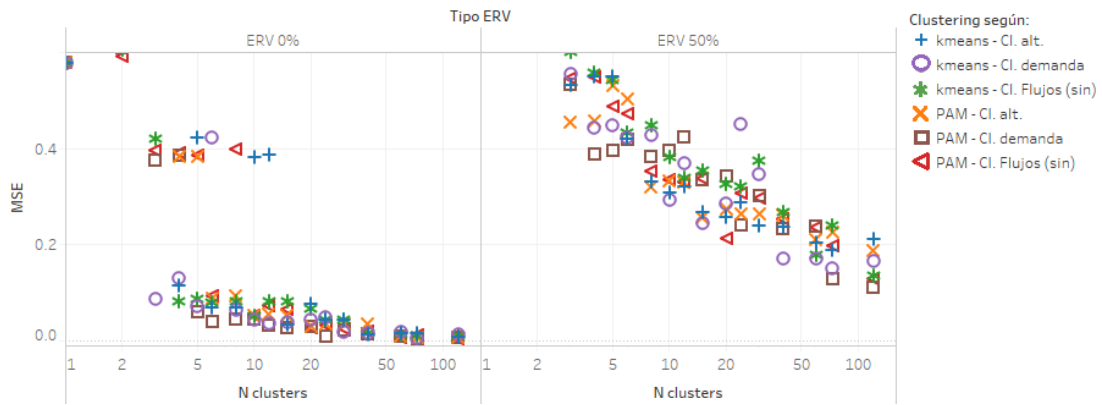


Figura 4.4.15: MSE para  $ERV_0$  y  $ERV_{50}$

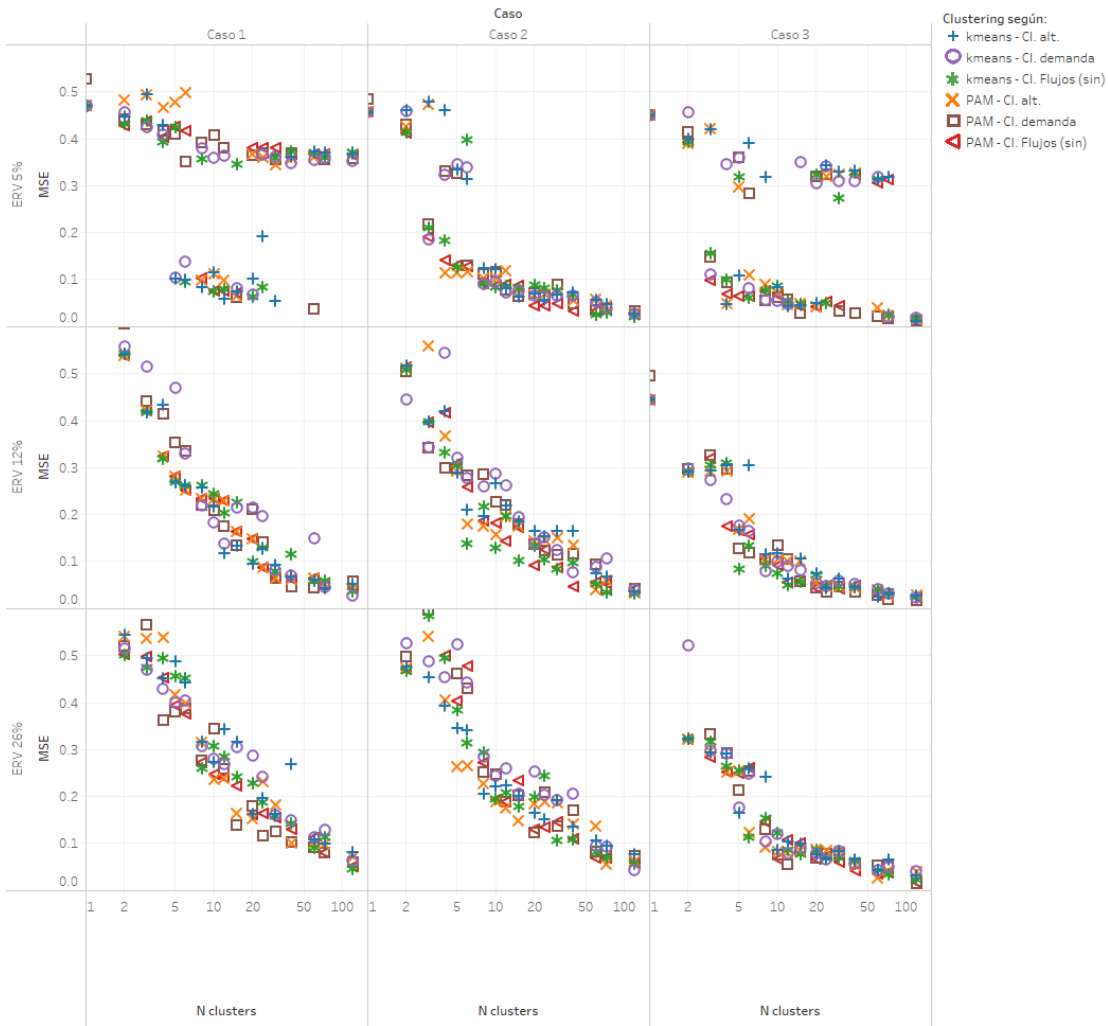


Figura 4.4.16: MSE para  $ERV_{5,12,25}^{1,2,3}$

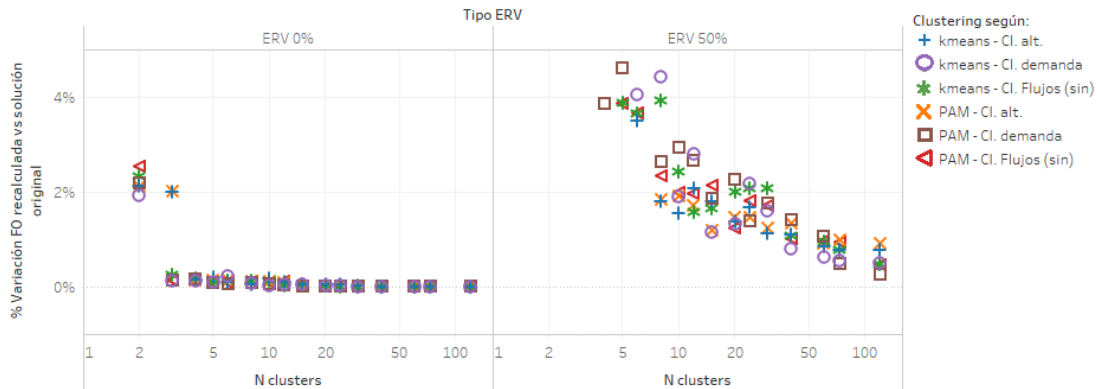


Figura 4.4.17: Error % FO recalculada para  $ERV_0$  y  $ERV_{50}$

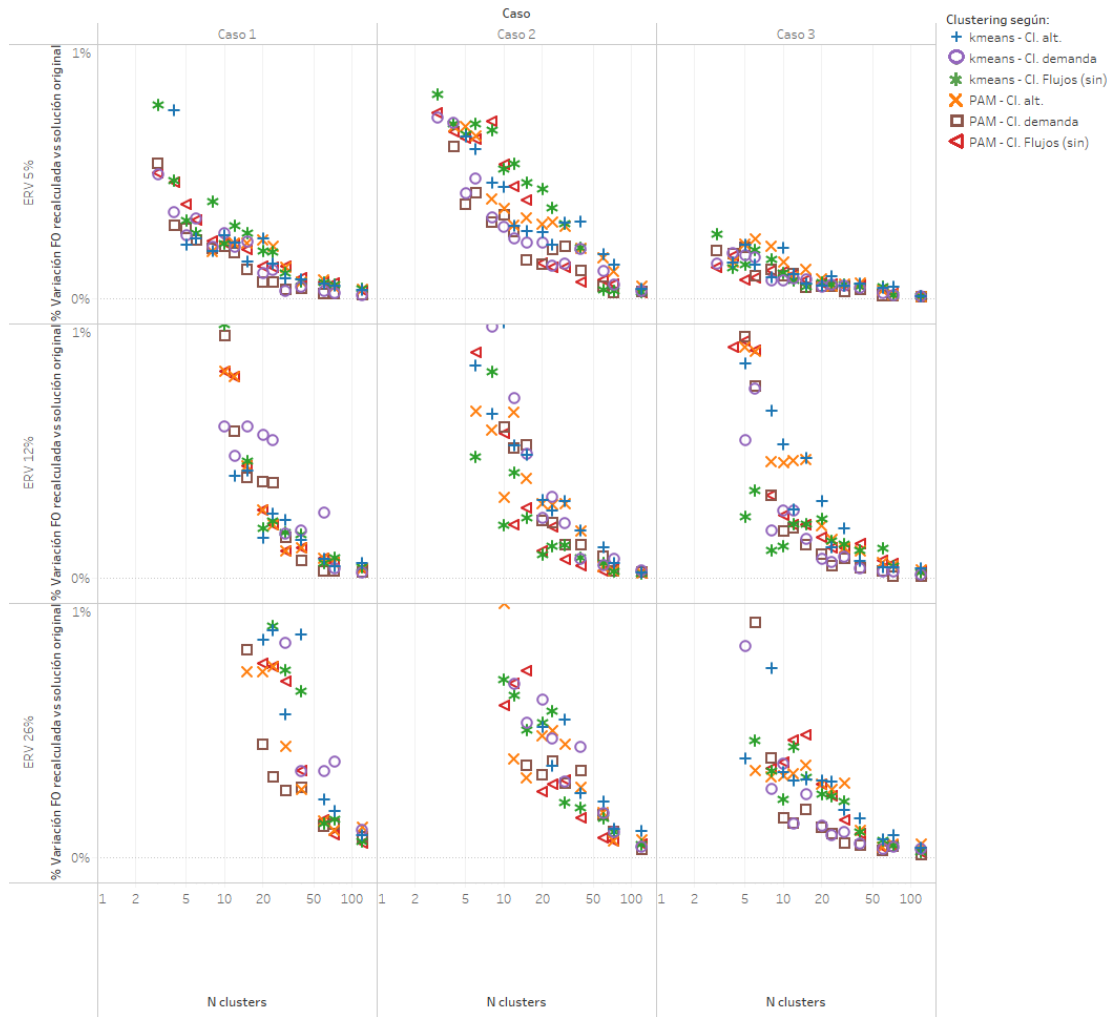


Figura 4.4.18: Error % FO recalculada para  $ERV_{5,12,25}^{1,2,3}$

Tabla 4.4.6: Ranking según FO

Clustering basado en:	Caso de estudio										
	$ERV_0$	$ERV_5^1$	$ERV_5^2$	$ERV_5^3$	$ERV_{12}^1$	$ERV_{12}^2$	$ERV_{12}^3$	$ERV_{26}^1$	$ERV_{26}^2$	$ERV_{26}^3$	$ERV_{50}$
Demanda/k-means	1	1	1	1	6	5	1	3	6	5	2
Flujos (sin)/k-means	2	2	3	3	3	2	3	5	2	3	4
Alternativo/k-means	3	4	5	4	4	3	4	6	5	4	5
Demanda/PAM	6	6	1	6	5	6	6	1	1	6	1
Flujos (sin)/PAM	4	3	6	2	2	1	2	2	3	1	3
Alternativo/PAM	5	5	4	5	1	4	5	4	4	2	6

Tabla 4.4.7: Ranking según MSE

Caso de estudio											
Clustering basado en:	ERV <sub>0</sub>	ERV <sub>5</sub> <sup>1</sup>	ERV <sub>5</sub> <sup>2</sup>	ERV <sub>5</sub> <sup>3</sup>	ERV <sub>12</sub> <sup>1</sup>	ERV <sub>12</sub> <sup>2</sup>	ERV <sub>12</sub> <sup>3</sup>	ERV <sub>26</sub> <sup>1</sup>	ERV <sub>26</sub> <sup>2</sup>	ERV <sub>26</sub> <sup>3</sup>	ERV <sub>50</sub>
Demanda/k-means	1	3	5	5	6	6	3	5	6	5	4
Flujos (sin)/k-means	3	2	3	4	4	1	2	4	3	2	6
Alternativo/k-means	6	1	6	6	1	5	6	6	2	4	3
Demanda/PAM	2	6	4	2	5	4	4	1	5	6	1
Flujos (sin)/PAM	5	4	1	1	3	2	1	2	4	3	5
Alternativo/PAM	4	5	2	3	2	3	5	3	1	1	2

Tabla 4.4.8: Ranking según FO recalculada

Caso de estudio											
Clustering basado en:	ERV <sub>0</sub>	ERV <sub>5</sub> <sup>1</sup>	ERV <sub>5</sub> <sup>2</sup>	ERV <sub>5</sub> <sup>3</sup>	ERV <sub>12</sub> <sup>1</sup>	ERV <sub>12</sub> <sup>2</sup>	ERV <sub>12</sub> <sup>3</sup>	ERV <sub>26</sub> <sup>1</sup>	ERV <sub>26</sub> <sup>2</sup>	ERV <sub>26</sub> <sup>3</sup>	ERV <sub>50</sub>
Demanda/k-means	1	1	1	1	6	5	1	4	5	3	4
Flujos (sin)/k-means	2	3	3	4	4	2	2	5	4	5	6
Alternativo/k-means	6	4	5	5	5	3	5	6	2	6	3
Demanda/PAM	4	6	6	2	1	6	4	1	6	1	1
Flujos (sin)/PAM	3	2	2	3	3	1	3	2	3	2	2
Alternativo/PAM	5	5	4	6	2	4	6	3	1	4	5

En base a las gráficas 4.4.13, 4.4.14, 4.4.15, 4.4.16, 4.4.17 y 4.4.18 se comparan las metodologías seleccionadas y se genera los *rankings* presentados en las tablas 4.4.6 4.4.7 y 4.4.8. En estos se ordenan desde uno a seis las respectivas soluciones obtenidas, siendo el uno la con menor diferencia respecto a la solución original.

Respecto al análisis realizado a través de los diferentes *rankings*, se aprecia que para los casos de baja variabilidad, la técnica con mejores resultados es generalmente k-means agrupando a partir de la demanda. Sin embargo, a medida que aumenta la variabilidad del sistema se obtienen mejores resultados con el agrupamiento respecto a la demanda y los flujos sin límites de capacidad considerando el *clustering* a través de PAM.

Finalmente, en la tabla 4.4.9 se realiza un análisis de los casos escogidos respecto a la cantidad mínima de *clusters* que requieren para llegar a un 1% de error respecto a la FO recalculada<sup>9</sup>. En ella se aprecia que el caso que basa su *clustering* en la demanda a través de k-means es el que requiere un menor número de *clusters* para el caso con mayor variabilidad. Sin embargo, de manera general los casos de agrupamiento de demanda y flujos sin límite de capacidad a través de PAM tienden a ser más robustos en el general de los casos.

Tabla 4.4.9: N° de *clusters* mínimo para un 1% de error en FO recalculada

Caso de estudio											
Clustering basado en:	ERV <sub>0</sub>	ERV <sub>5</sub> <sup>1</sup>	ERV <sub>5</sub> <sup>2</sup>	ERV <sub>5</sub> <sup>3</sup>	ERV <sub>12</sub> <sup>1</sup>	ERV <sub>12</sub> <sup>2</sup>	ERV <sub>12</sub> <sup>3</sup>	ERV <sub>26</sub> <sup>1</sup>	ERV <sub>26</sub> <sup>2</sup>	ERV <sub>26</sub> <sup>3</sup>	ERV <sub>50</sub>
Demanda/k-means	3	3	3	3	10	8	5	30	12	5	40
Flujos (sin)/k-means	3	3	3	3	10	6	5	24	10	6	60
Alternativo/k-means	4	4	5	4	12	6	5	20	20	5	60
Demanda/PAM	3	3	4	3	10	10	5	15	15	6	73
Flujos (sin)/PAM	3	3	3	3	10	6	4	20	10	8	60
Alternativo/PAM	4	8	4	4	10	6	5	15	10	6	60

<sup>9</sup>Se escoge la FO recalculada dado que es la métrica que entrega información más comparable de los casos de estudio.

---

De este modo, al igual que en el caso de la LDC, se puede apreciar que la técnica de k-means es mejor en casos con alta correlación espacial de la demanda, pero a niveles de variabilidad de demanda elevados, se obtienen mejores soluciones a través de PAM. Otra observación importante es que, si bien las técnicas de agrupamiento a partir de los flujos sin límites de capacidad de líneas no fueron las mejores, sus resultados fueron más robustos que los demás, manteniéndose de forma constante en un rango de soluciones aceptables para el problema, independiente de la variabilidad espacial y temporal.

## 4.5. Conclusiones

El presente capítulo presenta una metodología alternativa a la reducción de escenarios de demanda, basando su *clustering* en un análisis post-óptimo de variables de interés para el problema de expansión de la transmisión. La propuesta general es el agrupamiento de la data a partir de los datos obtenidos del pre-procesamiento de datos de entrada del TEP en un flujo óptimo de potencia. De este modo las salidas relativas a costos de operación y flujos por la líneas serán procesadas y clusterizadas para rescatar grupos de demanda representativos. Las principales conclusiones de este trabajo son:

- Al aumentar el número de *clusters* tiende a mejorar la calidad de la solución, independientemente de la métrica escogida. Dicha mejora no es siempre monótona, es decir, a veces un subconjunto de condiciones operacionales de menor cardinalidad se comporta mejor que uno de mayor cardinalidad. También se observa que el número de *clusters* para los cuales se alcanza un determinado nivel de calidad de la solución es dependiente de la diversidad espacial de la demanda neta (o la penetración renovable). Es decir, a mayor diversidad, se necesita un mayor número de *clusters* para un determinado nivel de calidad de la solución.
- Considerar la diversidad espacial de la demanda neta al momento de clusterizar condiciones operacionales permite obtener agrupamiento de mejor calidad, como lo demuestran las comparaciones respecto al agrupamiento en base a LDC (sin considerar diversidad espacial). El agrupamiento en base a LDC es competitivo solo en casos de alta correlación entre las demandas por barra.
- Los resultados numéricos indican que la selección de una metodología adecuada depende directamente del nivel de variabilidad del sistema, siendo generalmente mejor para casos de estudio con baja variabilidad el agrupar a partir de la demanda y flujos de potencia sin límites de capacidad, realizando el *cluster* con las técnicas de k-means y PAM. Sin embargo, para casos de alta variabilidad el agrupamiento basado en PAM destaca mayormente, siendo su mayor restricción su requerimiento computacional asociado.
- A partir del análisis de la FO recalculada, se establece que un número adecuado de *clusters* es 10, ya que en ese punto las soluciones se comienzan a estabilizar.
- Al igual que en el caso de la LDC, se puede apreciar que la técnica de k-means es mejor en casos con alta correlación espacial de la demanda, pero a niveles de variabilidad de demanda elevados, mejores soluciones se obtienen a través de PAM.
- Otra observación importante es que, si bien las técnicas de agrupamiento a partir de los flujos sin límites de capacidad de líneas no fueron las mejores, sus resultados fueron más robustos que los demás, manteniéndose de forma constante en un rango de soluciones aceptables para el problema, independiente de la variabilidad espacial y temporal.
- Relativo al análisis realizado a través de los diferentes ranking de soluciones obtenidas según casos de estudio, se aprecia que para los casos de baja variabilidad, la

técnica con mejores resultados es generalmente k-means agrupando a partir de la demanda. Sin embargo, a medida que aumenta la variabilidad espacial del sistema, el agrupamiento a través de PAM respecto a la demanda y los flujos sin límites por las líneas se vuelven cada vez mejores y más representativas.

- De manera general, PAM es mejor que k-means para casos con mayor diversidad espacial de la demanda.
- Agrupar en base a demanda neta por barra considerando su espacialidad da resultados de similar calidad que considerando un análisis post-óptimo al hacer un análisis transversal de los distintos casos de estudio. Es decir, no se observan mejoras significativas al hacer agrupamiento en base a resultados post-óptimos.

Futuras extensiones de este trabajo son el desarrollar metodologías que representen de mejor manera el valor representativo del *cluster* seleccionado, ya que, en el presente trabajo se utilizaba el promedio. De este modo, adaptando la técnica de *clustering* y estudiando las características del sistema se puede desarrollar alguna metodología de entrada-salida que logre captar de mejor manera su valor representativo. Otra aplicación de esta metodología es respecto a la selección de días representativos, y de este modo aplicar estas técnicas en modelos con restricciones inter-temporales.

## 4.6. Anexos

### 4.6.1. Resultados anexos

#### 4.6.1.1. Nuevas Líneas Construidas

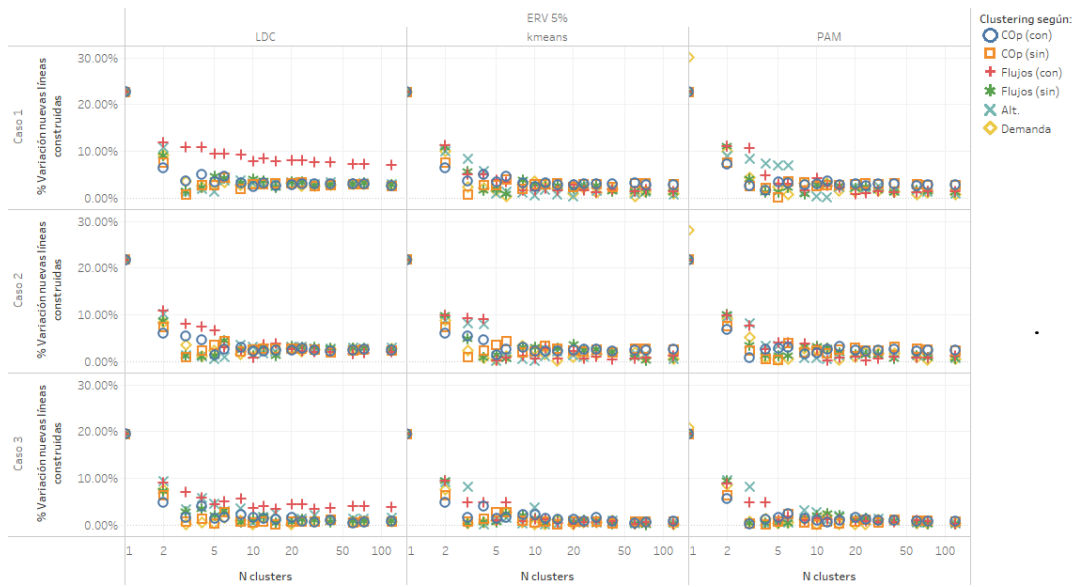


Figura 4.6.1: Error % nuevas líneas para  $ERV_5^1$ ,  $ERV_5^2$  y  $ERV_5^3$



Figura 4.6.2: Error % nuevas líneas para  $ERV_{12}^1$ ,  $ERV_{12}^2$  y  $ERV_{12}^3$



Figura 4.6.3: Error % nuevas líneas para  $ERV_{26}^1$ ,  $ERV_{26}^2$  y  $ERV_{26}^3$

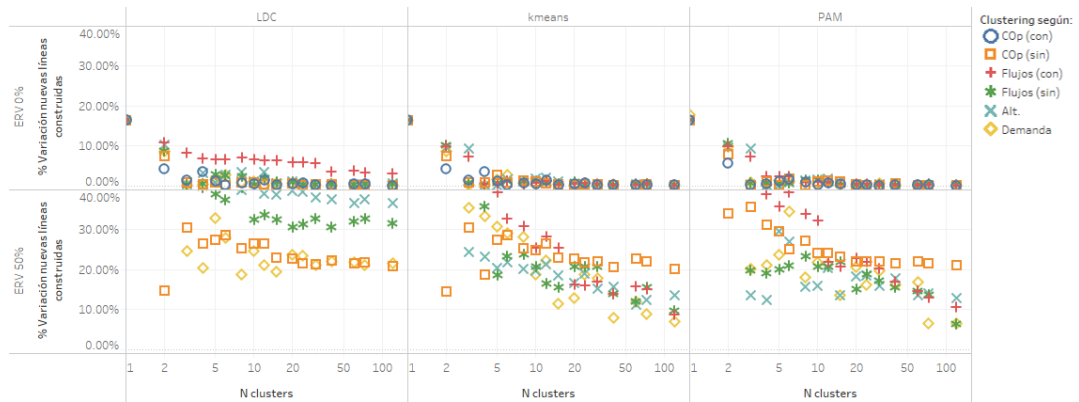


Figura 4.6.4: Error % nuevas líneas para  $ERV_0$  y  $ERV_{50}$

CAPÍTULO 4. Reducción Dimensional de la Demanda en Base a Condiciones Operacionales de un Análisis Post-óptimo

122

Tabla 4.6.1: Error % Nuevas líneas construidas para 120 clusters

N° clusters	% ERV	Caso	Método	Clustering Cop (con)	Clustering COp (sin)	Clustering Flujos (con)	Clustering Flujos (sin)	Clustering Alternativo	Clustering según demanda
120	0%	-	LDC	0.01%	0.01%	3.11%	0.32%	0.44%	0.05%
			K-means	0.16%	0.02%	0.02%	0.12%	0.04%	0.09%
			PAM	0.02%	0.02%	0.04%	0.07%	0.03%	0.03%
	Caso 1	LDC	2.55%	2.52%	6.87%	2.81%	2.89%	2.67%	
		K-means	2.67%	2.84%	1.25%	1.05%	0.55%	0.83%	
		PAM	2.70%	2.73%	1.32%	1.03%	0.79%	0.59%	
	5%	Caso 2	LDC	2.40%	2.25%	1.95%	2.58%	2.77%	2.65%
			K-means	2.58%	2.34%	1.17%	0.64%	0.40%	0.42%
			PAM	2.35%	2.27%	1.09%	0.53%	0.60%	0.41%
	Caso 3	LDC	0.80%	0.63%	3.73%	0.74%	1.48%	0.70%	
		K-means	0.68%	0.60%	0.11%	0.05%	0.44%	0.04%	
		PAM	0.69%	0.67%	0.17%	0.17%	0.29%	0.19%	
	Caso 1	LDC	18.49%	14.35%	16.45%	8.70%	8.78%	13.76%	
		K-means	18.72%	12.90%	4.75%	0.58%	1.31%	0.27%	
		PAM	17.39%	14.26%	4.23%	1.03%	1.03%	0.48%	
	12%	Caso 2	LDC	10.41%	12.47%	10.51%	10.28%	9.52%	12.43%
			K-means	10.29%	12.73%	3.87%	0.37%	0.46%	0.70%
			PAM	10.31%	12.36%	3.53%	0.04%	0.37%	0.07%
	Caso 3	LDC	3.83%	3.80%	1.39%	2.90%	3.20%	3.72%	
		K-means	3.73%	3.78%	1.13%	0.64%	0.62%	0.34%	
		PAM	3.44%	3.52%	0.61%	0.78%	0.74%	0.10%	
	Caso 1	LDC	23.48%	16.55%	22.93%	15.99%	17.03%	17.62%	
		K-means	23.73%	17.35%	4.60%	1.51%	1.69%	0.44%	
		PAM	22.22%	17.40%	4.55%	1.02%	1.59%	0.41%	
	26%	Caso 2	LDC	21.34%	14.74%	18.84%	22.41%	20.06%	15.11%
			K-means	20.92%	15.01%	7.72%	0.65%	2.01%	1.61%
			PAM	21.41%	15.26%	7.53%	0.92%	0.22%	1.20%
Caso 3	LDC	6.53%	6.01%	7.95%	5.85%	4.80%	6.12%		
	K-means	7.11%	5.99%	0.43%	0.32%	0.71%	0.17%		
	PAM	6.28%	6.14%	0.24%	0.46%	0.41%	0.29%		
50%	-	LDC	49.99%	20.80%	45.09%	31.42%	36.47%	21.61%	
		K-means	49.19%	20.20%	8.73%	9.63%	13.54%	6.89%	
		PAM	47.54%	21.12%	10.47%	6.28%	12.72%	6.32%	

4.6.1.2. MAE

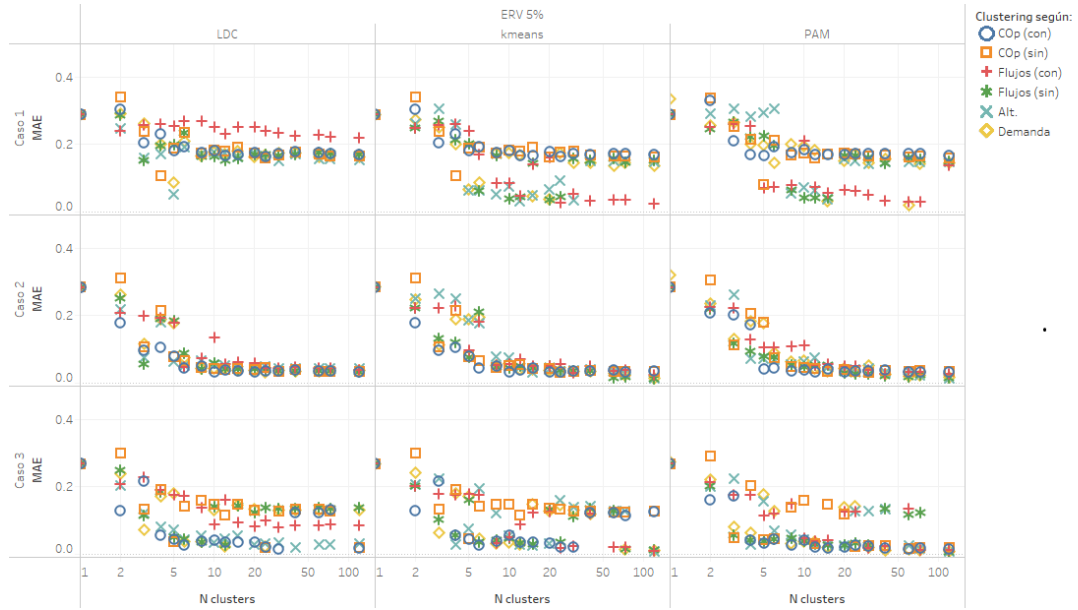


Figura 4.6.5: MAE para  $ERV_5^1$ ,  $ERV_5^2$  y  $ERV_5^3$



Figura 4.6.6: MAE para  $ERV_{12}^1$ ,  $ERV_{12}^2$  y  $ERV_{12}^3$

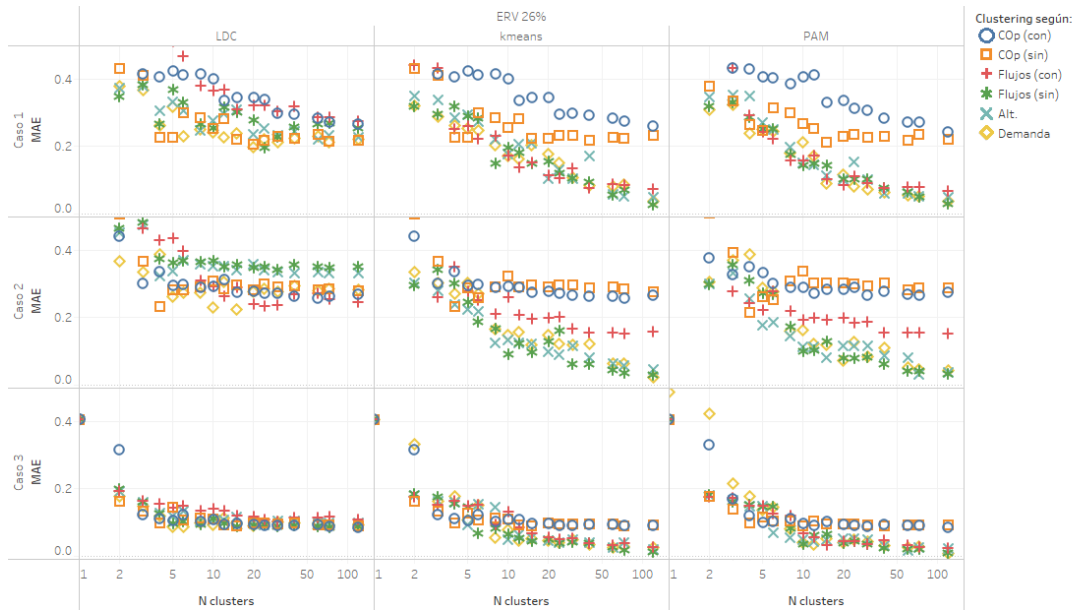


Figura 4.6.7: MAE para  $ERV_{26}^1$ ,  $ERV_{26}^2$  y  $ERV_{26}^3$



Figura 4.6.8: MAE para  $ERV_0$  y  $ERV_{50}$

Tabla 4.6.2: MAE para 120 clusters

$N^\circ$ clusters	% ERV	Caso	Método	Clustering Cop (con)	Clustering COP (sin)	Clustering Flujos (con)	Clustering Flujos (sin)	Clustering Alternativo	Clustering según demanda
120	0%	-	LDC	0.0025	0.0058	0.2071	0.0233	0.0118	0.0022
			K-means	0.0039	0.0020	0.0031	0.0054	0.0034	0.0064
			PAM	0.0019	0.0044	0.0025	0.0032	0.0038	0.0032
	Caso 1	LDC	0.1680	0.1648	0.2193	0.1678	0.1526	0.1626	
		K-means	0.1701	0.1627	0.0219	0.1516	0.1459	0.1365	
		PAM	0.1660	0.1592	0.1355	0.1493	0.1454	0.1432	
	5%	Caso 2	LDC	0.0305	0.0331	0.0378	0.0319	0.0427	0.0342
			K-means	0.0348	0.0315	0.0229	0.0094	0.0119	0.0112
			PAM	0.0297	0.0329	0.0263	0.0139	0.0129	0.0160
	Caso 3	LDC	0.0166	0.0179	0.0834	0.1368	0.0294	0.1284	
		K-means	0.1258	0.1262	0.0089	0.0099	0.0054	0.0098	
		PAM	0.0142	0.0183	0.0110	0.0051	0.0071	0.0065	
	Caso 1	LDC	0.2702	0.2234	0.2501	0.2096	0.2108	0.2158	
		K-means	0.2734	0.2067	0.0684	0.0172	0.0307	0.0130	
		PAM	0.2562	0.2183	0.0659	0.0247	0.0247	0.0262	
	12%	Caso 2	LDC	0.1583	0.1870	0.1288	0.1772	0.1706	0.1849
			K-means	0.1595	0.1818	0.0915	0.0169	0.0208	0.0245
			PAM	0.1506	0.1835	0.0826	0.0260	0.0168	0.0243
	Caso 3	LDC	0.0514	0.0549	0.0889	0.0479	0.0508	0.0517	
		K-means	0.0519	0.0545	0.0252	0.0130	0.0149	0.0128	
		PAM	0.0503	0.0510	0.0308	0.0124	0.0152	0.0087	
	Caso 1	LDC	0.2659	0.2163	0.2747	0.2546	0.2249	0.2200	
		K-means	0.2594	0.2318	0.0716	0.0248	0.0473	0.0361	
		PAM	0.2423	0.2203	0.0668	0.0284	0.0471	0.0342	
	26%	Caso 2	LDC	0.2678	0.2799	0.2457	0.3520	0.3304	0.2831
			K-means	0.2655	0.2764	0.1558	0.0293	0.0457	0.0222
			PAM	0.2749	0.2889	0.1507	0.0309	0.0357	0.0413
Caso 3	LDC	0.0850	0.0868	0.1083	0.0857	0.0958	0.0907		
	K-means	0.0907	0.0911	0.0239	0.0112	0.0171	0.0223		
	PAM	0.0842	0.0903	0.0210	0.0076	0.0210	0.0071		
50%	-	LDC	0.4283	0.1778	0.4134	0.2660	0.3231	0.1878	
		K-means	0.4216	0.1710	0.0816	0.0888	0.1167	0.0936	
		PAM	0.4086	0.1869	0.0911	0.0786	0.1109	0.0771	

4.6.1.3. Correlación

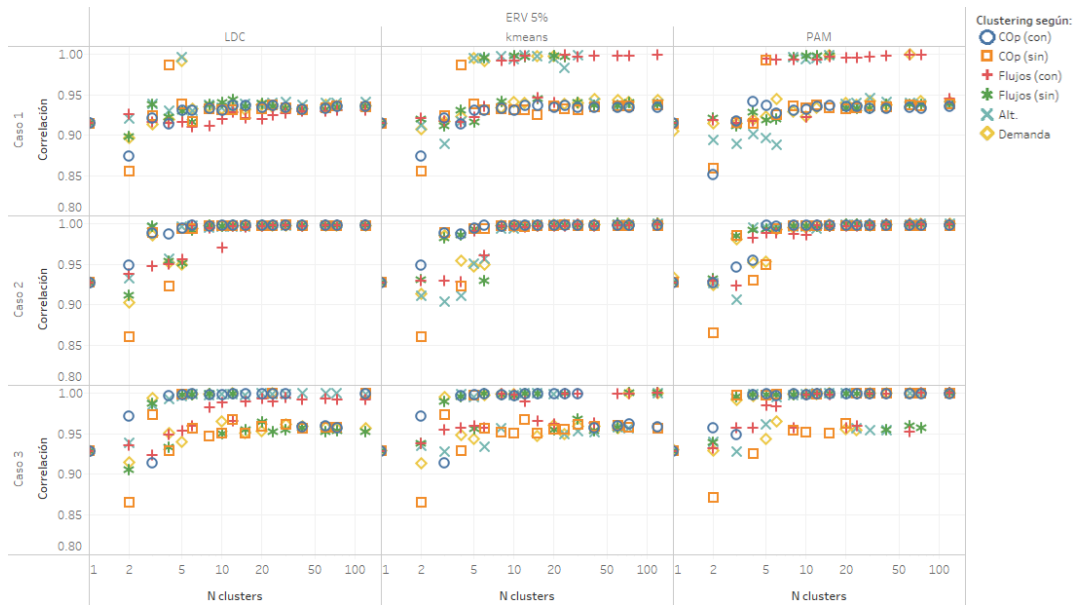


Figura 4.6.9: Correlación para  $ERV_5^1$ ,  $ERV_5^2$  y  $ERV_5^3$

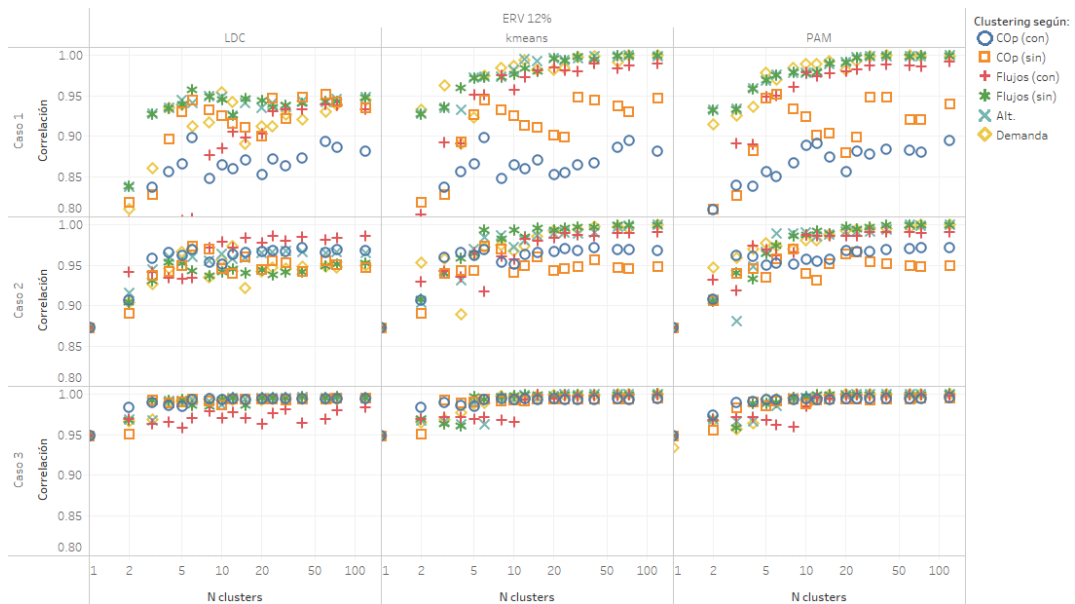


Figura 4.6.10: Correlación para  $ERV_{12}^1$ ,  $ERV_{12}^2$  y  $ERV_{12}^3$

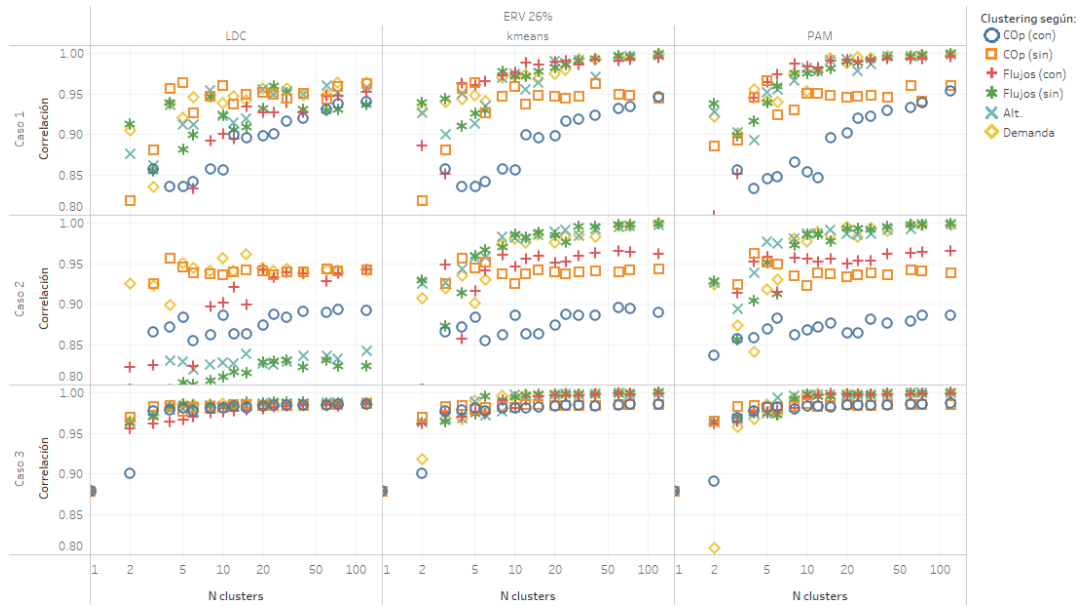


Figura 4.6.11: Correlación para  $ERV_{26}^1$ ,  $ERV_{26}^2$  y  $ERV_{26}^3$

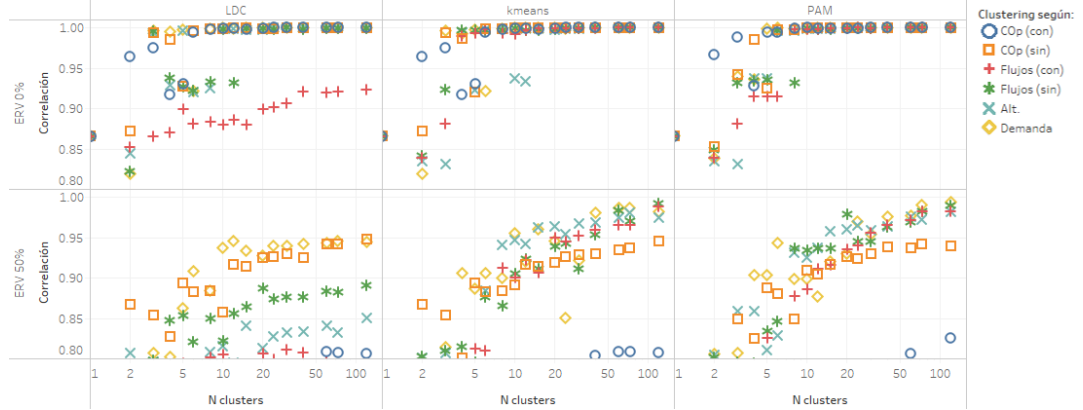


Figura 4.6.12: Correlación para  $ERV_0$  y  $ERV_{50}$

Tabla 4.6.3: Correlación para 120 clusters

N° clusters	% ERV	Caso	Método	Clustering Cop (con)	Clustering COp (sin)	Clustering Flujos (con)	Clustering Flujos (sin)	Clustering Alternativo	Clustering según demanda
120	0%	-	LDC	1.0000	1.0000	0.9228	0.9992	0.9998	1.0000
			K-means	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
			PAM	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Caso 1	LDC	0.9352	0.9349	0.9309	0.9347	0.9408	0.9362	
		K-means	0.9334	0.9365	0.9989	0.9372	0.9378	0.9428	
		PAM	0.9352	0.9376	0.9453	0.9386	0.9381	0.9412	
	5%	Caso 2	LDC	0.9979	0.9977	0.9981	0.9982	0.9972	0.9976
			K-means	0.9976	0.9978	0.9991	0.9998	0.9997	0.9997
			PAM	0.9980	0.9978	0.9989	0.9997	0.9997	0.9996
	Caso 3	LDC	0.9996	0.9996	0.9923	0.9521	0.9990	0.9567	
		K-means	0.9577	0.9565	0.9999	0.9999	1.0000	0.9999	
		PAM	0.9997	0.9996	0.9998	1.0000	0.9999	0.9999	
	Caso 1	LDC	0.8804	0.9342	0.9329	0.9473	0.9477	0.9442	
		K-means	0.8809	0.9471	0.9898	0.9995	0.9991	0.9997	
		PAM	0.8941	0.9393	0.9922	0.9994	0.9994	0.9987	
	12%	Caso 2	LDC	0.9682	0.9467	0.9857	0.9524	0.9644	0.9485
			K-means	0.9673	0.9478	0.9903	0.9996	0.9995	0.9994
			PAM	0.9718	0.9490	0.9906	0.9993	0.9996	0.9993
	Caso 3	LDC	0.9945	0.9944	0.9833	0.9960	0.9954	0.9952	
		K-means	0.9945	0.9947	0.9989	0.9998	0.9997	0.9998	
		PAM	0.9950	0.9952	0.9985	0.9998	0.9997	0.9999	
	Caso 1	LDC	0.9397	0.9618	0.9518	0.9367	0.9580	0.9632	
		K-means	0.9463	0.9437	0.9942	0.9993	0.9975	0.9984	
		PAM	0.9527	0.9603	0.9952	0.9990	0.9980	0.9987	
	26%	Caso 2	LDC	0.8913	0.9420	0.9419	0.8227	0.8422	0.9423
			K-means	0.8899	0.9430	0.9612	0.9986	0.9978	0.9994
			PAM	0.8853	0.9384	0.9654	0.9986	0.9983	0.9979
	Caso 3	LDC	0.9861	0.9868	0.9869	0.9870	0.9879	0.9856	
		K-means	0.9857	0.9858	0.9992	0.9998	0.9996	0.9993	
		PAM	0.9867	0.9851	0.9993	0.9999	0.9994	0.9999	
50%	-	LDC	0.8062	0.9479	0.7780	0.8906	0.8501	0.9439	
		K-means	0.8070	0.9449	0.9888	0.9916	0.9747	0.9813	
		PAM	0.8249	0.9391	0.9827	0.9891	0.9816	0.9933	

## Capítulo 5

# Conclusiones y Recomendaciones

Para concluir esta de tesis, se presentan en este capítulo las principales conclusiones y aportes de este trabajo, además de algunas recomendaciones para futuros trabajos que puedan ser desarrollado para dar continuidad a este proyecto.

### 5.1. Aportes

El objetivo principal de este trabajo es la representación reducida de la demanda y generación renovable variable de un sistema de potencia, considerando formulaciones que simplifiquen sus relaciones de temporalidad y espacialidad, para la aplicación de estas como entradas del problema de planificación de la transmisión eléctrica. De este modo, la primera parte del trabajo se basó en la reducción de la temporalidad a través del agrupamiento de la demanda neta del sistema a través de k-means y PAM, contrastando estos resultados con la técnica comúnmente empleada de la LDC.

Respecto a la revisión del estado del arte de este problema, si bien existen trabajos relacionados con la reducción dimensional del problema de expansión de la transmisión, estos generalmente están enfocados en la variable estocástica o espacial del sistema, representando generalmente la temporalidad a través de la LDC. Otros desarrollos utilizan técnicas complejas de tratamiento de datos de demanda a través de post-procesos de datos, sin embargo el trabajo desarrollado presenta los cimientos para poder comparar el resto de técnicas, realizando la reducción temporal de la demanda de manera directa en sus datos de entrada, procurando conservar la información espacial relevante en el problema.

De manera general, en este trabajo se han desarrollado casos de estudio para diferentes niveles de penetración renovable y localizaciones de estas, con lo cual es posible concluir de mejor manera sobre los efectos que tienen las diferentes técnicas utilizadas en diferentes niveles de variabilidad del sistema.

Además, respecto al modelo de planificación utilizado, se desarrolló un modelo no-lineal con relajación de variables de decisión, lo cual hace que los resultados sean más comparables debido a la reducción del gap de las soluciones obtenidas a través de este modelo, versus otros estudios que utilizan técnicas con un mayor error relativo, como en [52] en donde utilizan un modelo disyuntivo relajado (formulación de McCormick),

induciendo a un gap de error más alto y por ende a una pérdida de precisión de resultados, la cual al comparar errores muy pequeños se vuelve significativa.

En la segunda parte del trabajo, considerando los resultados de la primera, se desarrollan metodologías de post-procesamiento de variables de interés para el problema de expansión de la transmisión. La propuesta general es el agrupamiento de la data a partir de los datos obtenidos del pre-procesamiento de datos de entrada del TEP en un flujo óptimo de potencia. De este modo las salidas relativas a costos de operación y flujos por la líneas serán procesadas y clusterizadas para rescatar grupos de demanda representativos. De este modo, se proponen metodologías de agrupación basadas en las salidas de un sub-problema, las cuales al procesarse formarán los grupos representativos de demanda para el TEP.

Si bien en [52] se desarrolla una metodología similar, las diferencias principales son respecto al modelo y casos de estudio utilizados, además de la comparación con el agrupamiento de datos basado directamente en la demanda neta como entrada del sistema.

## 5.2. Conclusiones generales

Las principales conclusiones de este trabajo son:

### ■ Número de clusters:

- Al aumentar el número de *clusters* tiende a mejorar la calidad de la solución, independientemente de la métrica escogida. Dicha mejora no es siempre monotónica, es decir, a veces un subconjunto de condiciones operacionales de menor cardinalidad se comporta mejor que uno de mayor cardinalidad. También se observa que el número de *clusters* para los cuales se alcanza un determinado nivel de calidad de la solución es dependiente de la diversidad espacial de la demanda neta (o la penetración renovable). Es decir, a mayor diversidad, se necesita un mayor número de *clusters* para un determinado nivel de calidad de la solución.

### ■ Métricas de análisis de resultados:

- Respecto a la métricas utilizadas para el análisis de resultados, se propuso utilizar una “FO recalculada” cuyo análisis permite comparar de una forma realista las decisiones de construcción de nuevas líneas en el TEP. La elección de esta métrica se debe a que el análisis directo de la FO obtenida con los respectivos *clusters* introduce errores de precisión respecto a los costos de operación del sistema, debido a que los pesos de los bloques de demanda calculados determinan todo un periodo de operación, afectando directamente al valor final de la FO. El uso de esta métrica permitió comparar de mejor manera las soluciones obtenidas.
- A partir del análisis de la FO recalculada, se establece que un número adecuado de *clusters* es 10, ya que en ese punto las soluciones se comienzan a estabilizar.

**■ Diversidad espacial y generación ERV:**

- Considerar la diversidad espacial de la demanda neta al momento de clusterizar condiciones operacionales permite obtener agrupamiento de mejor calidad, como lo demuestran las comparaciones respecto al agrupamiento en base a LDC (sin considerar diversidad espacial). El agrupamiento en base a LDC es competitivo solo en casos de alta correlación entre las demandas por barra.
- Los resultados numéricos indican que la selección de una metodología adecuada depende directamente del nivel de variabilidad del sistema, siendo generalmente mejor para casos de estudio con baja variabilidad el agrupar a partir de la demanda y flujos de potencia sin límites de capacidad, realizando el *cluster* con las técnicas de k-means y PAM. Sin embargo, para casos de alta variabilidad el agrupamiento basado en PAM destaca mayormente, siendo su mayor restricción su requerimiento computacional asociado.
- Al igual que en el caso de la LDC, se puede apreciar que la técnica de k-means es mejor en casos con alta correlación espacial de la demanda, pero a niveles de variabilidad de demanda elevados, mejores soluciones se obtienen a través de PAM.
- Otra observación importante es que, si bien las técnicas de agrupamiento a partir de los flujos sin límites de capacidad de líneas no fueron las mejores, sus resultados fueron más robustos que los demás, manteniéndose de forma constante en un rango de soluciones aceptables para el problema, independiente de la variabilidad espacial y temporal.

**■ Conclusiones generales:**

- De manera general, PAM es mejor que k-means para casos con mayor diversidad espacial de la demanda.
- Agrupar en base a demanda neta por barra considerando su espacialidad da resultados de similar calidad que considerando un análisis post-óptimo al hacer un análisis transversal de los distintos casos de estudio. Es decir, no se observan mejoras significativas al hacer agrupamiento en base a resultados post-óptimos.

### 5.3. Recomendaciones

Futuras extensiones de este trabajo pueden incluir nuevas formulaciones del modelo de expansión de transmisión, con la incorporación de restricciones inter-temporales en el problema y posibles aplicaciones a metodologías de selección de días típicos. Respecto a la segunda parte del trabajo, se propone desarrollar metodologías que representen de mejor manera el valor representativo del *cluster* seleccionado, ya que, en el presente trabajo se utilizaba el promedio, sin embargo según las características del problema este valor puede ser determinado a través de algunas técnicas más complejas para aumentar su representatividad. De este modo, adaptando la técnica de *clustering* y estudiando

las características del sistema se puede desarrollar alguna metodología de entrada-salida que logre captar de mejor manera la entrada reducida del TEP. Otra posible extensión del trabajo es la aplicación de esta metodología en la selección de días representativos, y de este modo aplicar estas técnicas en modelos con restricciones inter-temporales.

# Anexos A

## Resumen

La rápida integración de las energías renovables y sus tecnologías han dado paso a nuevos desafíos en los Sistemas Eléctricos de Potencia (**SEP**), los cuales deben adaptarse a estas nuevas condiciones operacionales, siendo cada vez más la cantidad de variables consideradas dentro de su planificación estratégica. Es así como, por ejemplo, dada la alta variabilidad de este tipo de energías, se hace cada vez mayor la necesidad de incrementar los niveles de flexibilidad operativa del sistema, haciendo la operación del sistema cada vez más compleja. De este modo, surge la necesidad de buscar nuevas metodologías que ayuden a simplificar sus cálculos, sin perder la precisión requerida en estos.

Por otra parte, las interconexiones de los SEP (integración energética entre sistemas) son cada vez más frecuentes; por lo cual, el problema de la planificación de expansión de transmisión ha adquirido mayor dimensión y su implementación se requiere en sistemas cada vez más grandes.

En general, en los problemas de planificación de los SEP existen un gran número de variables dependientes e independientes que afectan su comportamiento, las cuales pueden ser indexadas de manera temporal, espacial, y por escenario, entre otras. Relativo al modelo del sistema, al tratarse de un estudio de mediano y largo plazo del SEP, este tipo de problemas poseen un gran número de variables de trabajo y probablemente de complejo manejo computacional, siendo su tratamiento dificultoso. Es por ello que el problema debe ser tratado de una forma simplificada, agregando variables para de este modo representar la operación de manera reducida, procurando perder la menor cantidad de información relevante en el proceso.

Una simplificación comúnmente utilizada en estos estudios es la discretización de las variables de entrada en bloques. En el caso de la demanda, esta ha sido tratada históricamente en los problemas de planificación eléctrica de forma discreta, trabajándose en bloques a través de la Curva de Duración de la Demanda o Carga -*Load Duration Curve (LDC)*<sup>1</sup>-, cuyo objetivo es aproximar la curva de demanda de un sistema en  $L$  bloques, donde cada bloque de demanda define un nivel y duración. La ventaja de esta

---

<sup>1</sup>La Curva de Duración de la Demanda ilustra la variación de la demanda de un sistema en orden descendiente, en lugar de su orden cronológico. Esta es utilizada generalmente para ilustrar la relación entre los requisitos de capacidad de generación y sus respectivos usos, dado un cierto incremento de carga.

metodología es que en vez de una representación detallada de 8760 horas al año, la demanda puede ser representada solo por  $L < 8760$ . Así, esta reducción en el tamaño del modelo permite expandir el horizonte de tiempo a más años (décadas) o aumentar el detalle de otros aspectos del modelo. Sin embargo, su desventaja radica en la pérdida de secuencialidad de los datos, y con ello información relevante relativa a temporalidad y causalidad del problema.

Otra variable de entrada del sistema con gran relevancia en este estudio es la generación de energías renovables variables (**ERV**)<sup>2</sup>, la cual se considera como ingresos de energía de disponibilidad limitada que posee variabilidad e incertidumbre, y cuyos perfiles de generación poseen alta correlación con su localización geográfica.

Es de considerar, además, que la influencia de estas variables sobre el sistema depende directamente de la escala del sistema que se estudia. En algunos casos se considera esta entrada en la *Demanda Neta*<sup>3</sup> del sistema, modelando su variabilidad directamente en la demanda, y simplificando de este modo espacialmente esta variable de entrada.

En resumen, este trabajo se basa en el análisis de datos multi-variables relativos a demanda y generación ERV y su aplicación en la resolución del problema de expansión de la transmisión. Se busca proponer y aplicar metodologías para la representación reducida de estas variables, considerando formulaciones que simplifiquen sus relaciones de temporalidad y espacialidad, minimizando en el proceso la pérdida de información relevante para la resolución del problema.

El primer estudio consiste en la reducción de escenarios de demanda en base a técnicas de *clustering* aplicadas a la demanda considerando su localización espacial, en donde, a través de diferentes escenarios de generación ERV se contrastan resultados numéricos respecto a la reducción LDC y las técnicas de k-means y PAM. Se observa que el considerar la espacialidad del sistema mejora las soluciones obtenidas respecto al caso LDC, y que a medida que se tiene una mayor variabilidad de generación ERV en el sistema se requiere de técnicas de *clustering* más complejas. En general los mejores resultados fueron obtenidos con PAM, sin embargo tiene la desventaja de que es un proceso con alto requerimiento computacional.

El segundo estudio consiste en un análisis post-óptimo de las soluciones de un OPF aplicado directamente en los datos de entrada del problema, en el cual se realiza un *clustering* respecto a las salidas que se consideren relevantes en cuanto al problema de expansión de la transmisión eléctrica. De este modo, se realiza *clustering* basado en costos de operación y flujos de potencia del sistema, con el objetivo de rescatar variables importantes relativas a la transmisión del sistema de potencia. Los resultados obtenidos son variados, ya que si bien hay algunas técnicas que se comportan mejor bajo algunos escenarios de ERV, es importante contrastar al igual que en el primer estudio, el requerimiento computacional asociado.

A modo de resumen, en la Figura A.0.1 se presenta un esquema que ilustra el trabajo propuesto con las metodologías de trabajo implementadas.

---

<sup>2</sup>En particular, se considerarán como energías renovables variables la generación solar y eólica

<sup>3</sup>La demanda neta considera a la generación renovable como negativa y se la resta al perfil de demanda original

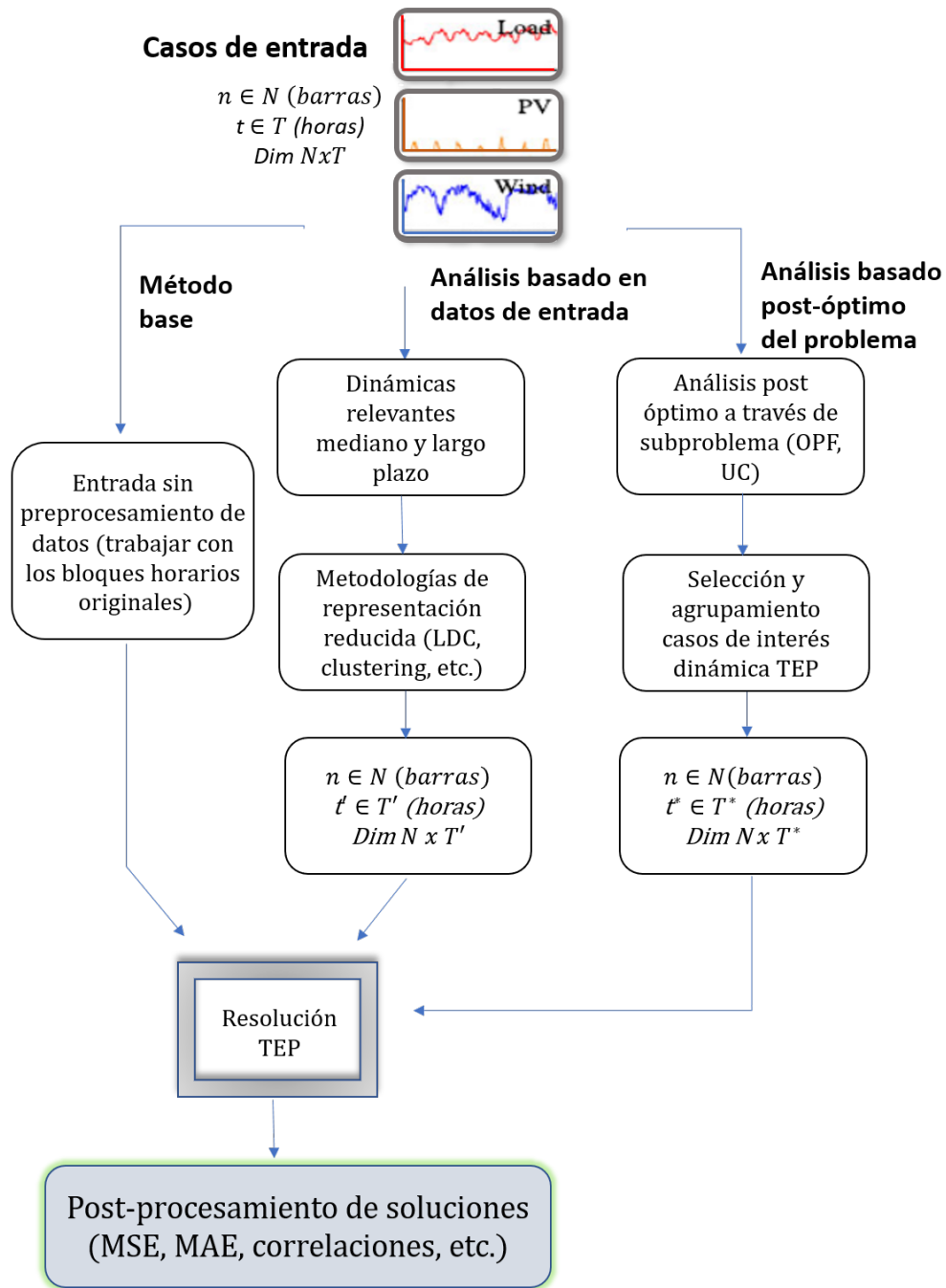


Figura A.0.1: Esquema del trabajo propuesto para estudio por caso.



# Referencias

- [1] Ignacio Aravena Solís. Representación discreta de variables independientes multidimensionales en la operación de sistemas eléctricos de potencia para estudios de mediano y largo plazo. Master's thesis, UTFSM, Valparaíso, Chile, 2013.
- [2] Pradeep Rai and Shubha Singh. A survey of clustering techniques. *International Journal of Computer Applications*, 7(12):1–5, 2010.
- [3] Benjamin Kroposki, Brian Johnson, Yingchen Zhang, Vahan Gevorgian, Paul Denholm, Bri-Mathias Hodge, and Bryan Hannegan. Achieving a 100 % renewable grid: Operating electric power systems with extremely high levels of variable renewable energy. *IEEE Power and Energy Magazine*, 15(2):61–73, 2017.
- [4] Geoffrey Rothwell and Thomas Gómez. Electricity economics. *IEEE series on power engineering*, 2003.
- [5] Wei Fan, Xiaohong Guan, and Qiaozhu Zhai. A new method for unit commitment with ramping constraints. *Electric Power Systems Research*, 62(3):215–224, 2002.
- [6] Antonio Frangioni and Claudio Gentile. Solving nonlinear single-unit commitment problems with ramping constraints. *Operations Research*, 54(4):767–775, 2006.
- [7] Xiao-Ping Zhang, Christian Rehtanz, and Bikash Pal. *Flexible AC transmission systems: modelling and control*. Springer Science & Business Media, 2012.
- [8] Fang Z Peng. Flexible ac transmission systems (facts) and resilient ac distribution systems (racds) in smart grid. *Proceedings of the IEEE*, 105(11):2099–2115, 2017.
- [9] Pierluigi Siano. Demand response and smart grids—a survey. *Renewable and sustainable energy reviews*, 30:461–478, 2014.
- [10] Ruilong Deng, Zaiyue Yang, Mo-Yuen Chow, and Jiming Chen. A survey on demand response in smart grids: Mathematical models and approaches. *IEEE Transactions on Industrial Informatics*, 11(3):570–582, 2015.

- 
- [11] Antonio Frangioni, Claudio Gentile, and Fabrizio Lacalandra. Solving unit commitment problems with general ramp constraints. *International Journal of Electrical Power & Energy Systems*, 30(5):316–326, 2008.
- [12] Antonio Frangioni, Claudio Gentile, and Fabrizio Lacalandra. Sequential lagrangian-milp approaches for unit commitment problems. *International Journal of Electrical Power & Energy Systems*, 33(3):585–593, 2011.
- [13] Salar Fattahi, Morteza Ashraphijuo, Javad Lavaei, and Alper Atamtürk. Conic relaxations of the unit commitment problem. *Energy*, 134:1079–1095, 2017.
- [14] Morteza Ashraphijuo, Salar Fattahi, Javad Lavaei, and Alper Atamtürk. A strong semidefinite programming relaxation of the unit commitment problem. In *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pages 694–701. IEEE, 2016.
- [15] András Prékopa. *Stochastic programming*, volume 324. Springer Science & Business Media, 2013.
- [16] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. Robust optimization. princeton series in applied mathematics, 2009.
- [17] Florin Capitanescu. Critical review of recent advances and further developments needed in ac optimal power flow. *Electric Power Systems Research*, 136:57–68, 2016.
- [18] Carleton Coffrin, Hassan L Hijazi, and Pascal Van Hentenryck. The qc relaxation: A theoretical and computational study on optimal power flow. *IEEE Transactions on Power Systems*, 31(4):3008–3018, 2016.
- [19] Stefan Pfenninger, Adam Hawkes, and James Keirstead. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews*, 33:74–86, 2014.
- [20] Richard P O’Neill, Eric A Krall, Kory W Hedman, and Shmuel S Oren. A model and approach to the challenge posed by optimal power systems planning. *Mathematical Programming*, 140(2):239–266, 2013.
- [21] Omid Ziaee, Omid Alizadeh-Mousavi, and F Fred Choobineh. Co-optimization of transmission expansion planning and tsc placement considering the correlation between wind and demand scenarios. *IEEE Transactions on Power Systems*, 33(1):206–215, 2018.
- [22] Yury Dvorkin, Ricardo Fernández-Blanco, Yishen Wang, Bolun Xu, Daniel S Kirschen, Hrvoje Pandžić, Jean-Paul Watson, and Cesar A Silva-Monroy. Co-planning of investments in transmission and merchant energy storage. *IEEE Transactions on Power Systems*, 33(1):245–256, 2018.
- [23] Esteban Gil, Ignacio Aravena, and Raúl Cárdenas. Generation capacity expansion planning under hydro uncertainty using stochastic mixed integer programming and scenario reduction. *IEEE Transactions on Power Systems*, 30(4):1838–1847, 2015.

- 
- [24] Antonio J Conejo, LB Morales, S Jalal Kazempour, and Afzal S Siddiqui. Investment in electricity generation and transmission. *Cham, Zug, Switzerland: Springer International Publishing*, 2016.
- [25] François Lévêque. *Competitive electricity markets and sustainability*. Edward Elgar Publishing, 2007.
- [26] Alexandre Moreira, David Pozo, Alexandre Street, and Enzo Sauma. Reliable renewable generation and transmission expansion planning: Co-optimizing system's resources for meeting renewable targets. *IEEE Transactions on Power Systems*, 32(4):3246–3257, 2017.
- [27] Matthias Fripp. Switch: a planning tool for power systems with large shares of intermittent renewable energy. *Environmental science & technology*, 46(11):6371–6378, 2012.
- [28] Carlos Suazo-Martínez, Eduardo Pereira-Bonvallet, and Rodrigo Palma-Behnke. A simulation framework for optimal energy storage sizing. *Energies*, 7(5):3033–3055, 2014.
- [29] Salvador Pineda and Juan Miguel Morales. Chronological time-period clustering for optimal capacity expansion planning with storage. *IEEE Transactions on Power Systems*, 2018.
- [30] JA Aguado, S de la Torre, and A Triviño. Battery energy storage systems in transmission network expansion planning. *Electric Power Systems Research*, 145:63–72, 2017.
- [31] Zechun Hu, Fang Zhang, and Baowei Li. Transmission expansion planning considering the deployment of energy storage systems. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–6. IEEE, 2012.
- [32] Carlos Bustos, Enzo Sauma, Sebastián de la Torre, José A Aguado, Javier Contreras, and David Pozo. Energy storage and transmission expansion planning: substitutes or complements? *IET Generation, Transmission & Distribution*, 12(8):1738–1746, 2017.
- [33] Ali Hajebrahimi, Amir Abdollahi, and Masoud Rashidinejad. Probabilistic multiobjective transmission expansion planning incorporating demand response resources and large-scale distant wind farms. *IEEE Systems Journal*, 11(2):1170–1181, 2017.
- [34] Charles I Nweke, Frank Leañez, Glenn R Drayton, and Mohan Kolhe. Benefits of chronological optimization in capacity planning for electricity markets. In *Power System Technology (POWERCON), 2012 IEEE International Conference on*, pages 1–6. IEEE, 2012.
- [35] Hugh Rudnick and Juan Zolezzi. Planificación y expansión de la transmisión en mercados eléctricos competitivos. *Pontificia Universidad Católica de Chile*, 2000.

- [36] Reza Hemmati, Rahmat-Allah Hooshmand, and Amin Khodabakhshian. State-of-the-art of transmission expansion planning: Comprehensive review. *Renewable and Sustainable Energy Reviews*, 23:312–319, 2013.
- [37] Abolfazl Arabpour, Mohammad Reza Besmi, and Pouria Maghouli. Transmission expansion planning with linearized ac load flow by special ordered set method. *Journal of Energy Engineering*, 144(2):04018008, 2018.
- [38] Andrés H Domínguez, Antonio H Escobar, and Ramón 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. *Electric Power Systems Research*, 143:611–623, 2017.
- [39] Wenyuan Li and Paul Choudhury. Probabilistic planning of transmission systems: Why, how and an actual example. In *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, pages 1–8. IEEE, 2008.
- [40] Wenyuan Li. Framework of probabilistic power system planning. *CSEE Journal of Power and Energy Systems*, 1(1):1–8, 2015.
- [41] Carlos Ruiz and Antonio J Conejo. Robust transmission expansion planning. *European Journal of Operational Research*, 242(2):390–401, 2015.
- [42] Xiaohu Zhang, Kevin Tomsovic, and Aleksandar Dimitrovski. Security constrained multi-stage transmission expansion planning considering a continuously variable series reactor. *IEEE Transactions on Power Systems*, 32(6):4442–4450, 2017.
- [43] Victor Hinojosa and Francisco Gonzalez-Longatt. Preventive security-constrained dcopf formulation using power transmission distribution factors and line outage distribution factors. *Energies*, 11(6):1497, 2018.
- [44] Mohammad Majidi-Qadikolai and Ross Baldick. A generalized decomposition framework for large-scale transmission expansion planning. *IEEE Transactions on Power Systems*, 33(2):1635–1649, 2018.
- [45] Ibrahim Alhamrouni, Azhar Khairuddin, Mohamed Salem, and Abdelrahman Alnajjar. Review on transmission expansion planning models. In *Applied Mechanics and Materials*, volume 818, pages 129–133. Trans Tech Publ, 2016.
- [46] Ignacio Aravena, Esteban Gil, and Anthony Papavasiliou. Alternative load cases selection method for transmission expansion planning. 2014. unpublished.
- [47] Andreas Wächter and Lorenz T Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57, 2006.
- [48] HSL HSL. A collection of fortran codes for large scale scientific computation, 2002, 2002.

- 
- [49] Guillermo Vinasco, Marcos J Rider, and Ruben Romero. A strategy to solve the multistage transmission expansion planning problem. *IEEE Transactions on Power Systems*, 26(4):2574–2576, 2011.
- [50] Anthony V Fiacco and Garth P McCormick. *Nonlinear programming: sequential unconstrained minimization techniques*, volume 4. Siam, 1990.
- [51] Pedro M Castro. Tightening piecewise mccormick relaxations for bilinear problems. *Computers & Chemical Engineering*, 72:300–311, 2015.
- [52] Quentin Ploussard, Luis Olmos, and Andrés Ramos. An operational state aggregation technique for transmission expansion planning based on line benefits. *IEEE Transactions on Power Systems*, 32(4):2744–2755, 2016.
- [53] Comisión-Nacional-Energía. Propuesta de expansión del sistema de transmisión 2018-1, 2018. [Web; accedido el 09-09-2018].
- [54] Murty P Bhavaraju, JD Hebson, and Willian Wood. Emerging issues in power system planning. *Proceedings of the IEEE*, 77(6):891–898, 1989.
- [55] Angel Marin and Javier Salmeron. Electric capacity expansion under uncertain demand: decomposition approaches. *IEEE transactions on power systems*, 13(2):333–339, 1998.
- [56] Sara Lumbreras, Andrés Ramos, and Fernando Banez-Chicharro. Optimal transmission network expansion planning in real-sized power systems with high renewable penetration. *Electric Power Systems Research*, 149:76–88, 2017.
- [57] Michael C Caramanis, Richard D Tabors, Kumar S Nochur, and Fred C Schweppe. The introduction of nondispatchable technologies a decision variables in long-term generation expansion models. *IEEE Transactions on Power Apparatus and Systems*, (8):2658–2667, 1982.
- [58] Sonja Wogrin. *Generation expansion planning in electricity markets with bilevel mathematical programming techniques*. PhD thesis, Universidad Pontificia Comillas, 2013.
- [59] Sonja Wogrin, Pablo Dueñas, Andrés Delgadillo, and Javier Reneses. A new approach to model load levels in electric power systems with high renewable penetration. *IEEE Transactions on Power Systems*, 29(5):2210–2218, 2014.
- [60] Diego Alejandro Tejada-Arango, Maya Domeshek, Sonja Wogrin, and Efraim Centeno. Enhanced representative days and system states modeling for energy storage investment analysis. *IEEE Transactions on Power Systems*, 2018.
- [61] L Baringo and AJ Conejo. Correlated wind-power production and electric load scenarios for investment decisions. *Applied energy*, 101:475–482, 2013.

- [62] Heejung Park and Ross Baldick. Transmission planning under uncertainties of wind and load: Sequential approximation approach. *IEEE Transactions on Power Systems*, 28(3):2395–2402, 2013.
- [63] Duehee Lee and Ross Baldick. Load and wind power scenario generation through the generalized dynamic factor model. *IEEE Transactions on Power Systems*, 32(1):400–410, 2017.
- [64] Cliff Grigg, Peter Wong, Paul Albrecht, Ron Allan, Murty Bhavaraju, Roy Billinton, Quan Chen, Clement Fong, Suheil Haddad, Sastry Kuruganty, et al. The iee reliability test system-1996. a report prepared by the reliability test system task force of the application of probability methods subcommittee. *IEEE Transactions on power systems*, 14(3):1010–1020, 1999.
- [65] Comisión-Nacional-Energía. Estudio de previsión de demanda 2015-2035 (2050), 2015. [Web; accedido el 09-09-2018].
- [66] Marcel Boiteux. *La tarification des demandes en pointe: Application de la théorie de la vente au coût marginal*. place Henri-Bergson, 1949.
- [67] Boiteux. Peak-load pricing. *The Journal of Business*, 33(2):157–179, 1960.
- [68] Efraim Centeno, Javier Reneses, and Julin Barquin. Strategic analysis of electricity markets under uncertainty: A conjectured-price-response approach. *IEEE Transactions on Power Systems*, 22(1):423–432, 2007.
- [69] Yihsu Chen and Benjamin F Hobbs. An oligopolistic power market model with tradable no/sub x/permits. *IEEE Transactions on Power systems*, 20(1):119–129, 2005.
- [70] Mohan Munasinghe. Electricity pricing; a comprehensive framework. *IEEE Transactions on Power Apparatus and Systems*, (8):3959–3968, 1981.
- [71] Frederic H Murphy and Yves Smeers. Generation capacity expansion in imperfectly competitive restructured electricity markets. *Operations research*, 53(4):646–661, 2005.
- [72] Ralph Turvey and Dennis Anderson. *Electricity economics: Essays and case studies*. 1977.
- [73] K Poncelet, H Höschle, A Virag, E Delarue, and W D’haeseleer. Selection of representative days for investment planning models. *kU Leuven, TME Working Paper WP EN2015-10*, [https://www.mech.kuleuven.be/en/tme/research/energy\\_environment/Pdf/wpen2015-10b.pdf](https://www.mech.kuleuven.be/en/tme/research/energy_environment/Pdf/wpen2015-10b.pdf), 2015.
- [74] Kris Poncelet, Hanspeter Höschle, Erik Delarue, Ana Virag, and William D’haeseleer. Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. *IEEE Transactions on Power Systems*, 32(3):1936–1948, 2017.

- 
- [75] Paul Nahmmacher, Eva Schmid, Lion Hirth, and Brigitte Knopf. Carpe diem: A novel approach to select representative days for long-term power system modeling. *Energy*, 112:430–442, 2016.
- [76] MS ElNozahy, MMA Salama, and R Seethapathy. A probabilistic load modelling approach using clustering algorithms. In *Power and Energy Society General Meeting (PES), 2013 IEEE*, pages 1–5. IEEE, 2013.
- [77] Abeer Almaimouni, Atinuke Ademola-Idowu, J Nathan Kutz, Ahlmahz Negash, and Daniel Kirschen. Selecting and evaluating representative days for generation expansion planning. In *2018 Power Systems Computation Conference (PSCC)*, pages 1–7. IEEE, 2018.
- [78] Yixian Liu, Ramteen Sioshansi, and Antonio J Conejo. Hierarchical clustering to find representative operating periods for capacity-expansion modeling. *IEEE Transactions on Power Systems*, 33(3):3029–3039, 2018.
- [79] Richard J Roiger. *Data mining: a tutorial-based primer*. Chapman and Hall/CRC, 2017.
- [80] Martin Maechler, Peter Rousseeuw, Anja Struyf, Mia Hubert, Kurt Hornik, et al. Cluster: cluster analysis basics and extensions. *R package version*, 1(2):56, 2012.
- [81] Christian Hennig. fpc: Flexible procedures for clustering. *R package version*, 2(2):0–3, 2010.
- [82] Tagaram Soni Madhulatha. Comparison between k-means and k-medoids clustering algorithms. In *Advances in Computing and Information Technology*, pages 472–481. Springer, 2011.
- [83] Antonio Gómez-Expósito, Antonio J Conejo, and Claudio Canizares. *Electric energy systems: analysis and operation*. CRC press, 2018.
- [84] Kris Poncelet, Hanspeter Höschle, Erik Delarue, Ana Virag, and William D’haeseleer. Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. *IEEE Transactions on Power Systems*, 32(3):1936–1948, 2016.
- [85] Desta Z Fitiwi, F de Cuadra, L Olmos, and M Rivier. 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*, 90:1360–1376, 2015.
- [86] M Sun, F Teng, I Konstantelos, and G Strbac. An objective-based scenario selection method for transmission network expansion planning with multivariate stochasticity in load and renewable energy sources. *Energy*, 145:871–885, 2018.

- 
- [87] Claude-Alexandra Fezeu, Keith Bell, Jiansong Ding, Patrick Panciatici, and Marie-Sophie Debry. Simplified representation of a large transmission network for use in long-term expansion planning. In *2014 Power Systems Computation Conference*, pages 1–7. IEEE, 2014.
- [88] Quentin Ploussard, Luis Olmos, and Andrés Ramos. An efficient network reduction method for transmission expansion planning using multicut problem and kron reduction. *IEEE Transactions on Power Systems*, 33(6):6120–6130, 2018.
- [89] Quentin Ploussard. *Efficient reduction techniques for a large-scale Transmission Expansion Planning problem*. PhD thesis, KTH Royal Institute of Technology, 2019.
- [90] Wenyuan Li. *Probabilistic transmission system planning*, volume 65. John Wiley & Sons, 2011.
- [91] J Swarndeep Saket and Sharnil Pandya. An overview of partitioning algorithms in clustering techniques.
- [92] P. Kaufman, L. & Rousseeuw. Clustering by Means of Medoids. In *Statistical Data Analysis Based on the L1 Norm and Related Methods*, pages 405–416, North-Holland, 1987.
- [93] Leonard Kaufman and Peter J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley-Interscience, Hoboken, NJ, USA, 2005. ISBN 0-471-87876-6.
- [94] Cort J Willmott and Kenji Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.
- [95] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.
- [96] William E Hart, Carl D Laird, Jean-Paul Watson, David L Woodruff, Gabriel A Hackebeitl, Bethany L Nicholson, and John D Sirola. *Pyomo-optimization modeling in python*, volume 67. Springer, 2017.
- [97] Guido Van Rossum and Fred L Drake. Python language reference manual. 2003.
- [98] A HSL. collection of fortran codes for large-scale scientific computation. See <http://www.hsl.rl.ac.uk>, 2007.

*Donde haya un árbol que plantar,  
plántalo tú.  
Donde haya un error que enmendar,  
enmiéndalo tú.  
Donde haya un esfuerzo que todos esquivan,  
hazlo tú.  
Sé tú el que aparta la piedra del camino*  
Gabriela Mistral

