



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

MODELO DE PROYECCIÓN DE ESCENARIOS DE CARBONO  
NEUTRALIDAD BAJO INCERTIDUMBRE

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL  
ELÉCTRICO

VICENTE SEPÚLVEDA FIGUEROA

PROFESOR GUÍA:  
CARLOS BENAVIDES FARÍAS

MIEMBROS DE LA COMISIÓN:  
RODRIGO PALMA BEHNKE  
RIGOBERTO TORRES AVILA

Este trabajo ha sido financiado por  
Centro de Energía

SANTIAGO DE CHILE

2022

RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO  
POR: VICENTE SEPÚLVEDA FIGUEROA  
FECHA: 2022  
PROF. GUÍA: SR. CARLOS BENAVIDES

## MODELO DE PROYECCIÓN DE ESCENARIOS DE CARBONO NEUTRALIDAD BAJO INCERTIDUMBRE

En el contexto de cambio climático y meta de carbono neutralidad por parte de Chile al año 2050, el Centro de Energía de la Universidad de Chile, en conjunto con otras instituciones, realizó un modelo capaz de proyectar distintos escenarios de emisiones de gases de efecto invernadero, considerando parámetros macroeconómicos, objetivos de políticas públicas, y datos multisectoriales, todo bajo un enfoque de incertidumbre. El modelo fue desarrollado en el contexto de un estudio solicitado por el Banco Interamericano de Desarrollo.

La generación de escenarios de este modelo considera un muestreo de los parámetros de modelación utilizando un muestreo por hipercubo latino. Es decir, divide el rango de incertidumbre de cada parámetro en muestras distintas (discretización), y combina cada una de estas muestras de forma aleatoria con las muestras de los otros parámetros. El modelo desarrollado fue evaluado considerando hasta 1000 escenarios futuros posibles. Por lo tanto, solo se modela una parte del total de escenarios posibles.

La propuesta de este trabajo considera diseñar e implementar una nueva metodología que permita reconocer una representación más amplia de los escenarios posibles futuros, representando en un número finito de simulaciones lo que podría pasar al simular todas las combinaciones posibles, considerando los rangos de incertidumbre.

Esta nueva metodología establece un algoritmo que detecta patrones o características en común entre distintas trayectorias de emisiones, costos de inversión y operación, generando trayectorias representativas. Y a partir de estas, se generan nuevos escenarios utilizando muestreo de hipercubo latino.

La metodología desarrollada en esta memoria fue aplicada a un caso de estudio y se compararon los resultados con una metodología previamente desarrollada. Los resultados obtenidos consisten en un aumento del rango de incertidumbre en un 16% de las emisiones GEI acumuladas en el periodo, 30% en el costo de inversión acumulado, 11% en el costo operacional acumulado y un 43% en las emisiones GEI del año 2050. Es decir, se incrementó el rango de incertidumbre en 12 millones de toneladas para el año 2050.

Finalmente, se logra expandir el rango de incertidumbre incrementando el universo de escenarios posibles. Al mismo tiempo, la metodología propuesta permite orientar la búsqueda de escenarios y parámetros significativos en modelos bajo incertidumbre.



*Para Lía*

# Agradecimientos

Primero, agradecer a mi nueva familia. A Lía, la personita que da fuerzas y motivación, la razón de que trasnochar no sea tan terrible si cuando despiertas sonríes y nos buscas. Agradezco poder estar cerca de ti, verte crecer, verte feliz. A la Yeka, quien ha estado en gran parte de mi historia. Gracias por estar siempre, por la buena compañía, por los café en pleno estudio, los abrazos y juegos improvisados. Esta memoria se escribió con gran esfuerzo de ambos, criando, estudiando y trabajando.

A mi familia de siempre, mi mamá, la ela y la javi. Con ejemplos aprendí a darlo todo, a no dormir, al doble trabajo, a que la familia se cuida y quiere, porque es lo que siempre está, en las buenas y en las malas.

A mis hermanos de otras madres, con quienes he estado en los mejores y peores momentos, en la cancha y en la vida, pero estamos y seguimos, cada uno en lo suyo, compartiendo de vez en cuando.

Empecé la carrera motivado por ayudar a personas con el uso de nuevas tecnologías. Luego, me gustó lo eléctrico, la magia de conectar cables y ver su efecto en una luz o una pantalla, y saber qué pasa. Finalmente, me quedé con la energía, eso que mueve el mundo, y que puede salvarlo. Agradecer a quienes me encontré en el camino, sobre todo con quienes pasé noches haciendo trabajos y proyectos a última hora. Mención especial a Rodrigo, a quien conocí en la fila de la inducción, y terminamos pasando por tanto.

A Carlos, mi tutor. Agradecer su tiempo y paciencia, las oportunidades que me ha dado, y que espero haberlas aprovechado bien. Gracias a él descubrí un mundo distinto, un mundo de colegas, de familias, de sueños y trabajo duro.

Finalmente, al Centro de Energía y al DIE, su gente y sus proyectos. Me parece genial lo que hacen, y me alegra pensar que puedo contribuir al menos con un granito de arena.

# Tabla de Contenido

Introducción.....	1
1.1 Motivación y antecedentes .....	1
1.2. Descripción del problema .....	5
1.3. Objetivos.....	7
1.4. Estructura del informe.....	7
Marco Teórico y Estado del Arte .....	8
2.1. Modelo de proyección de escenarios .....	8
2.2. Muestreo de hipercubo latino (LHS).....	9
2.2.1. Proyección de escenarios utilizando LHS .....	10
2.3. Algoritmos de reducción de escenarios.....	12
2.3.1. Aprendizaje no supervisado .....	12
2.3.2. Distancia euclidiana .....	13
2.3.3. Normalización de datos .....	14
2.3.4. Clustering k-means .....	15
2.3.5. Selección de k y distribución de datos .....	17
2.3.6. Análisis de componentes principales (PCA).....	18
2.4. Estado del arte .....	22
2.4.1. Modelos de proyección y de escenarios.....	22
2.4.2. Muestreo de hipercubo latino .....	23
Metodología y Aportes del Trabajo de Memoria .....	24
3.1. Formalización del problema.....	24
3.1.1. Combinación de parámetros .....	24
3.1.2. Escenarios críticos .....	24
3.1.3. Formulación del problema .....	25
3.2. Resultados que se esperan obtener .....	26
3.3. Metodología propuesta .....	27
3.3.1. Etapa 1 - Simulación de modelo de referencia .....	29
3.3.2. Etapa 2 - Carga y análisis de datos .....	29

3.3.3. Etapa 3 - Agrupación o <i>clustering k-means</i> .....	38
3.3.4. Etapa 4 - Combinación de parámetros.....	43
3.3.5. Etapa 5 – Proyección de nuevos escenarios.....	43
Resultados.....	46
4.1. Resultados por modelos.....	47
4.1.1. Histograma y dispersión.....	47
4.1.2. Dispersión y rangos de incertidumbre .....	54
4.1.3. Trayectorias de los escenarios.....	57
4.2. Resultados por sectores.....	60
4.2.1. Sector comercio .....	60
4.2.2. Sector industria y minería .....	65
4.2.3. Sector público.....	72
4.2.4. Sector residencial.....	76
4.2.5. Sector transporte .....	82
4.2.6. Sector generación eléctrica.....	90
Análisis de resultados.....	96
5.1. Modelo de referencia y metodología propuesta .....	96
5.2. Análisis por sectores .....	99
5.2.1. Sector comercio .....	99
5.2.2. Sector industria y minería .....	100
5.2.3. Sector público.....	102
5.2.4. Sector residencial.....	102
5.2.5. Sector transporte .....	103
5.2.6. Sector generación eléctrica.....	105
.....	105
5.3. Carbono neutralidad y tormenta perfecta .....	108
5.4. Aprendizaje no supervisado.....	109
5.5. Uso de la metodología y oportunidades.....	109
Conclusiones.....	110
Bibliografía .....	111
Anexo .....	113
Parámetros con incertidumbre .....	113

# Índice de Tablas

Tabla 1: Ejemplo de matriz LHS o muestreo de hipercubo latino, para 10 escenarios (o muestras) y 4 parámetros. ....	9
Tabla 2: Ejemplo de normalización de matriz de datos. ....	14
Tabla 3: Matriz de LHS, correspondiente a la incertidumbre aplicada en cada escenario en los parámetros del modelo de referencia. ....	30
Tabla 4: Parámetros de menor desviación estándar en grupo de escenarios de valores máximos y mínimos de emisiones acumuladas, sector industria y minería. ....	44
Tabla 5: Distribución de escenarios proyectados en nueva matriz LHS. ....	45
Tabla 6: Comparación de valores máximos y mínimos de las métricas de desempeño, para modelo de referencia u original y modelo propuesto. ....	96
Tabla 7: Comparación de rango de las métricas de desempeño y aumento porcentual, para modelo de referencia u original y modelo propuesto. ....	96
Tabla 8: Parámetros de sector comercio bajo incertidumbre. ....	113
Tabla 9: Parámetros de sector industria y minería bajo incertidumbre. ....	113
Tabla 10: Parámetros de sector público bajo incertidumbre. ....	115
Tabla 11: Parámetros de sector residencial bajo incertidumbre. ....	115
Tabla 12: Parámetros de sector transporte bajo incertidumbre. ....	116
Tabla 13: Parámetros de sector generación eléctrica bajo incertidumbre. ....	117

# Índice de Figuras

Figura 1: Trayectoria de emisiones del escenario de Carbono Neutralidad hacia el 2050: mitigación y captura. Fuente: Ministerio de Energía. ....	2
Figura 2: Resultados de emisiones según supuestos de referencia, agrupado por sectores. Fuente: Banco Interamericano de Desarrollo [4]. ....	3
Figura 3: Resultados de emisiones ante distintos escenarios, según medidas de mitigación originales (NDC), y originales + adicionales (NDC+). Fuente: Banco Interamericano de Desarrollo. ....	4
Figura 4: Diagrama de combinación completa de escenarios por sector a partir de la incertidumbre de 5 parámetros asociados al sector generación eléctrica, transporte e industria y minería (ejemplo). Fuente: Elaboración propia. ....	6
Figura 5: Diagrama de la plataforma de modelación para el análisis de la Estrategia de Descarbonización de Largo Plazo. ....	8
Figura 6: Transformación de muestras LHS a distribución normal, de media igual a 0 y desviación estándar igual a 1. ....	10
Figura 7: Ejemplo de vector que distribuye normal utilizando muestras de LHS. ....	11
Figura 8: Reducción de escenarios partiendo de 1.000 escenarios de muestra de la distribución normal estándar bidimensional (izquierda) reducidos a 100 escenarios (centro) y reducidos a 50 escenarios (derecha). Se muestran las correspondientes descomposiciones de $\mathbb{R}^2$ en regiones de Voronoi obtenidas por el algoritmo de selección hacia delante con respecto a la norma euclidiana. ....	12
Figura 9: Comparación de datos al aplicar normalización. ....	15
Figura 10: Ejemplo de resultado del algoritmo de <i>clustering k-means</i> , para $k=3$ (número de grupos), donde los puntos negros son los centroides o puntos medios de cada clúster. ....	16
Figura 11: Reducción de dimensionalidad asociada a método PCA. ....	19
Figura 12: Interpretación geométrica de componentes principales. ....	19
Figura 13: Ejemplo de PCA, componentes principales de una galleta. ....	20
Figura 14: Varianza acumulada en función del número de componentes principales. ....	21
Figura 15: Comparación de máximo y mínimo total de emisiones al año 2050, y suma de los máximos y mínimos de los sectores en distintos escenarios. ....	25
Figura 16: Diagrama de modelo propuesto utilizando metodología de combinación de parámetros, agrupación y clasificación por <i>k-means</i> , y proyección de nuevos escenarios mediante muestreo de hipercubo latino. ....	28

Figura 17: Trayectorias según incertidumbre para precio de combustible diésel, electricidad e hidrógeno, en sector industria y minería, para casos de referencia, $lhs = 0.01$ y $lhs = 0.99$ . .....	31
Figura 18: Métricas de desempeño acumuladas (periodo 2020-2050), modelo original, gráficos de dispersión e histogramas (en la diagonal). .....	33
Figura 19: Análisis por método del codo para métricas de desempeño del modelo de referencia, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul). .....	36
Figura 20: Histograma de escenarios por grupo o clúster, para $k=117$ (número de clústeres) en 1001 escenarios, para modelo original o de referencia. ....	36
Figura 21: Varianza explicada acumulada de los escenarios de valores máximos del sector residencial. ....	39
Figura 22: (a) Peso o importancia de cada parámetro en las componentes principales, y (b) pesos absolutos (magnitud), para escenarios de valores máximos del sector residencial. ....	41
Figura 23: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos del sector residencial, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul). ....	42
Figura 24: Métricas de desempeño totales o suma de sectores, obtenidas de modelo original, periodo 2020-2050. ....	48
Figura 25: Métricas de desempeño totales o suma de sectores, obtenidas de modelo propuesto, periodo 2020-2050. ....	50
Figura 26: Métricas de desempeño totales o suma de sectores, obtenidas de 2da iteración, periodo 2020-2050. ....	51
Figura 27: Métricas de desempeño totales o suma de sectores, obtenidas de 3ra iteración, periodo 2020-2050. ....	52
Figura 28: Métricas de desempeño totales o suma de sectores, obtenidas de 4ta iteración, periodo 2020-2050. ....	53
Figura 29: Gráfico de caja de emisiones totales o suma de sectores, periodo 2020-2050. ....	54
Figura 30: Gráfico de caja de emisiones totales o suma de sectores, año 2050. ....	55
Figura 31: Gráfico de caja de CAPEX total o suma de sectores, periodo 2020-2050. ....	56
Figura 32: Gráfico de caja de OPEX total o suma de sectores, periodo 2020-2050. ....	56
Figura 33: Trayectorias de emisiones totales o suma de sectores, para modelos original, propuesto e iteraciones. Periodo 2020-2050. ....	57
Figura 34: Trayectorias de CAPEX total o suma de sectores, para modelos original, propuesto e iteraciones. Periodo 2020-2050. ....	58

Figura 35: Trayectorias de OPEX total o suma de sectores, para modelos original, propuesto e iteraciones. Periodo 2020-2050. ....	59
Figura 36: PCA y varianza explicada acumulada en sector comercio, para escenarios de valores máximos en cuanto a métricas de desempeño. ....	60
Figura 37: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos (arriba) y mínimos (abajo) del sector comercio. ...	61
Figura 38: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector comercio, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul). ....	62
Figura 39: Gráficos de caja asociado a las emisiones del sector comercio. ....	62
Figura 40: Gráficos de caja asociado a costos del sector comercio. ....	63
Figura 41: Trayectorias de emisiones sector comercio, modelo original y propuesto. Periodo 2020-2050. ....	63
Figura 42: Trayectorias de CAPEX sector comercio, modelo original y propuesto. Periodo 2020-2050. ....	64
Figura 43: Trayectorias de OPEX sector comercio, modelo original y propuesto. Periodo 2020-2050. ....	64
Figura 44: PCA y varianza explicada acumulada en sector industria y minería, para escenarios de valores máximos en cuanto a métricas de desempeño. ....	65
Figura 45: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos del sector industria y minería. ....	66
Figura 46: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores mínimos del sector industria y minería. ....	67
Figura 47: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector industria y minería, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul). ....	68
Figura 48: Gráficos de caja asociado a emisiones del sector industria y minería. ...	69
Figura 49: Gráficos de caja asociado a costos del sector industria y minería. ....	69
Figura 50: Trayectorias de emisiones sector industria y minería, modelo original y propuesto. Periodo 2020-2050. ....	70
Figura 51: Trayectorias de CAPEX sector industria y minería modelo original y propuesto. Periodo 2020-2050. ....	71
Figura 52: Trayectorias de OPEX sector industria y minería, modelo original y propuesto. Periodo 2020-2050. ....	71

Figura 53: PCA y varianza explicada acumulada en sector público, para escenarios de valores máximos en cuanto a métricas de desempeño. ....	72
Figura 54: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos (arriba) y mínimos (abajo) del sector público. ....	73
Figura 55: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector público, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul).....	74
Figura 56: Gráficos de caja asociado a las emisiones del sector público.....	74
Figura 57: Trayectorias de emisiones sector público, modelo original y propuesto. Periodo 2020-2050.....	75
Figura 58: PCA y varianza explicada acumulada en sector residencial, para escenarios de valores máximos en cuanto a métricas de desempeño. ....	76
Figura 59: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos del sector residencial. ....	77
Figura 60: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores mínimos del sector residencial.....	78
Figura 61: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector residencial, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul). ....	79
Figura 62: Gráficos de caja asociado a emisiones del sector residencial. ....	79
Figura 63: Gráficos de caja asociado a costos del sector residencial. ....	80
Figura 64: Trayectorias de emisiones sector residencial, modelo original y propuesto. Periodo 2020-2050.....	80
Figura 65: Trayectorias de CAPEX residencial, modelo original y propuesto. Periodo 2020-2050.....	81
Figura 66: Trayectorias de OPEX sector residencial, modelo original y propuesto. Periodo 2020-2050.....	81
Figura 67: PCA y varianza explicada acumulada en sector transporte, para escenarios de valores máximos (arriba) y mínimos (abajo) en cuanto a métricas de desempeño. ....	83
Figura 68: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos del sector transporte. ....	84
Figura 69: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores mínimos del sector transporte. ....	85
Figura 70: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector	

transporte, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul). .....	86
Figura 71: Gráficos de caja asociado a emisiones del sector transporte.....	87
Figura 72: Gráficos de caja asociado a costos del sector transporte.....	87
Figura 73: Trayectorias de emisiones sector transporte, modelo original y propuesto. Periodo 2020-2050.....	88
Figura 74: Trayectorias de CAPEX transporte, modelo original y propuesto. Periodo 2020-2050.....	88
Figura 75: Trayectorias de OPEX sector transporte, modelo original y propuesto. Periodo 2020-2050.....	89
Figura 76: PCA y varianza explicada acumulada en sector generación eléctrica, para escenarios de valores máximos (arriba) y mínimos (abajo) en cuanto a métricas de desempeño. ....	90
Figura 77: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos del sector generación eléctrica. ....	91
Figura 78: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores mínimos del sector generación eléctrica.....	92
Figura 79: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector generación eléctrica, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul). ....	93
Figura 80: Gráficos de caja asociado a emisiones del sector generación eléctrica. .	93
Figura 81: Gráficos de caja asociado a costos del sector generación eléctrica. ....	94
Figura 82: Trayectorias de emisiones sector generación eléctrica, modelo original y propuesto. Periodo 2020-2050.....	94
Figura 83: Trayectorias de CAPEX generación eléctrica, modelo original y propuesto. Periodo 2020-2050.....	95
Figura 84: Trayectorias de OPEX sector generación eléctrica, modelo original y propuesto. Periodo 2020-2050.....	95
Figura 85: Comparación entre valores máximos (izquierda) y mínimos (derecha) para el total de emisiones del año 2050 en un mismo escenario, y la suma de sectores en distintos escenarios. ....	97
Figura 86: Diferencia entre máximos y mínimos totales y por suma de sectores, en los distintos modelos e iteraciones.....	98
Figura 87: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector comercio, modelo propuesto.....	99

Figura 88: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector industria y minería, modelo propuesto. ....	101
Figura 89: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector público, modelo propuesto. ....	102
Figura 90: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector residencial, modelo propuesto. ....	103
Figura 91: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector transporte, modelo propuesto.....	104
Figura 92: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector generación eléctrica, modelo propuesto. ....	105
Figura 93: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector generación eléctrica en modelo de referencia.....	106
Figura 94: Diferencia entre modelo original y modelo propuesto, de los promedios de muestras de LHS para parámetros del sector generación eléctrica, en 10 escenarios de valores máximos y 10 de valores mínimos, en emisiones en el año 2050. ....	106

# Capítulo 1

## Introducción

### 1.1 Motivación y antecedentes

El mundo se encuentra en una situación crítica debido al cambio climático, producto de años de sobreexplotación de los recursos, malas prácticas industriales y tardío avance de tecnologías sustentables. Por tanto, distintos países, entre ellos Chile, participaron del Acuerdo de París el año 2015, donde se comprometen a hacer todo lo que esté a su alcance por medio de contribuciones determinadas a nivel nacional (NDC por sus siglas en inglés). A su vez, incluye la obligación de que todas las partes informen periódicamente sobre sus emisiones y sobre sus esfuerzos de aplicación. Ese año Chile presentó su primer NDC [1], en el cual se comprometía a reducir su intensidad de emisiones por unidad de PIB en un 30% al año 2030, respecto al nivel alcanzado en 2007.

Luego, en el año 2020, Chile presentó su actualización de NDC, aumentando su nivel de compromiso [2]. Chile se comprometió a que sus emisiones sean menores a 95 millones tCO<sub>2</sub> el año 2030 y a alcanzar la carbono neutralidad hacia el año 2050. Ese mismo año el Ministerio de Energía publica la proyección de consumo energético nacional, presentando las medidas de mitigación que permitan lograr emisiones netas iguales a cero al año 2050 [3], en paralelo a la actualización de las contribuciones determinadas a nivel nacional. Este informe incluye medidas asociadas a electromovilidad, hidrógeno, retiro de centrales a carbón e industria sostenible, entre otras; y sus medidas corresponden al insumo fundamental del modelo de proyección de escenarios a utilizar en esta memoria.

Las medidas de mitigación son contrastadas con un escenario de referencia (o línea base), que corresponde a los niveles de emisión de carbono en el caso que se continúe con las políticas y estrategias actuales. En Figura 1, se presenta la comparación entre la trayectoria de referencia y una trayectoria de emisiones considerando las distintas medidas de mitigación, incluyendo la captura de emisiones propia de los bosques.

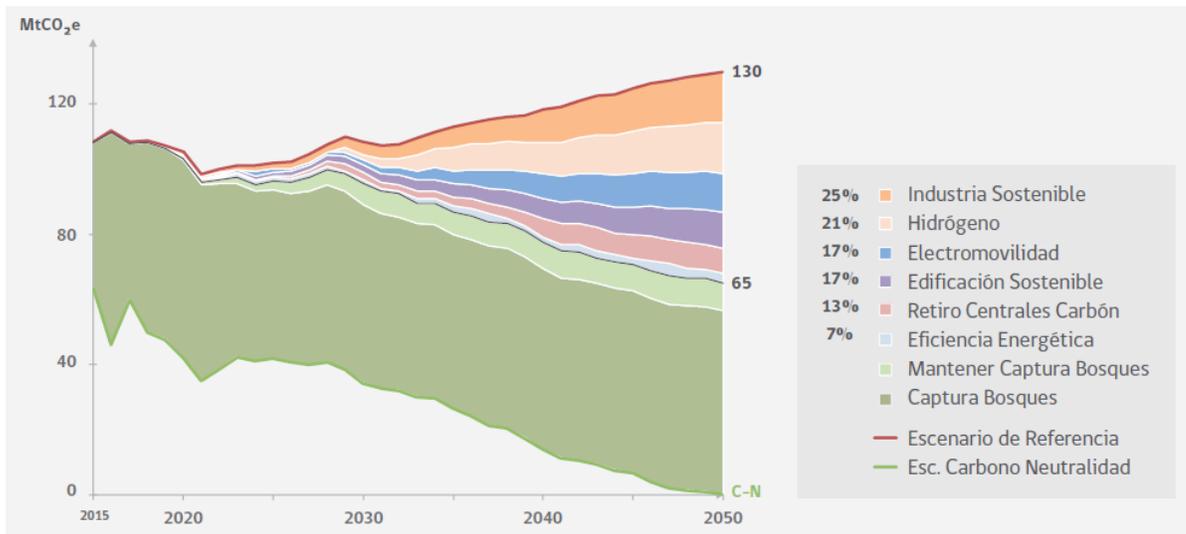


Figura 1: Trayectoria de emisiones del escenario de Carbono Neutralidad hacia el 2050: mitigación y captura. Fuente: Ministerio de Energía.

Estas trayectorias están sujetas a incertidumbre debido a la naturaleza de los distintos sectores implicados. Por ejemplo, en electromovilidad, un sector que se proyecta participe del 17% en la reducción de las emisiones de carbono al 2050, considera como meta lograr un 100% de transporte público urbano constituido por buses eléctricos, en la Región Metropolitana, al año 2040. Así también, considera un 100% de taxis eléctricos a nivel nacional, para el año 2050. Estas medidas están sujetas a distintos factores, como la instalación de electrolinerías públicas y privadas, o modelos de negocio que permitan reducir el impacto del alto costo de inversión en vehículos eléctricos, a la vez que fomentan su ahorro operacional.

Esta incertidumbre da origen a la necesidad de desarrollar una herramienta de proyección de escenarios basado en un modelo robusto, que permita tomar decisiones fundamentadas y efectivas.

El Centro de Energía de la Universidad de Chile, en colaboración con el Banco Interamericano de Desarrollo, *RAND Corporation*, CLAPES UC y el Ministerio del Medio Ambiente, realizó una herramienta computacional o modelo de simulación, donde se proyectan escenarios de sectores productivos, energéticos y medioambientales, afectados a las medidas de mitigación o captura de carbono. El proyecto busca, además de simular los distintos escenarios, establecer medidas adicionales y analizar la importancia de cada sector y medida en el resultado final [4].

En Figura 2, se presenta la trayectoria de emisiones netas totales para el escenario de referencia, según estrategia de medidas de mitigación de carbono del NDC, desglosado según emisión o captura de los distintos sectores. Esta proyección fue realizada utilizando el modelo desarrollado en estudio previo.

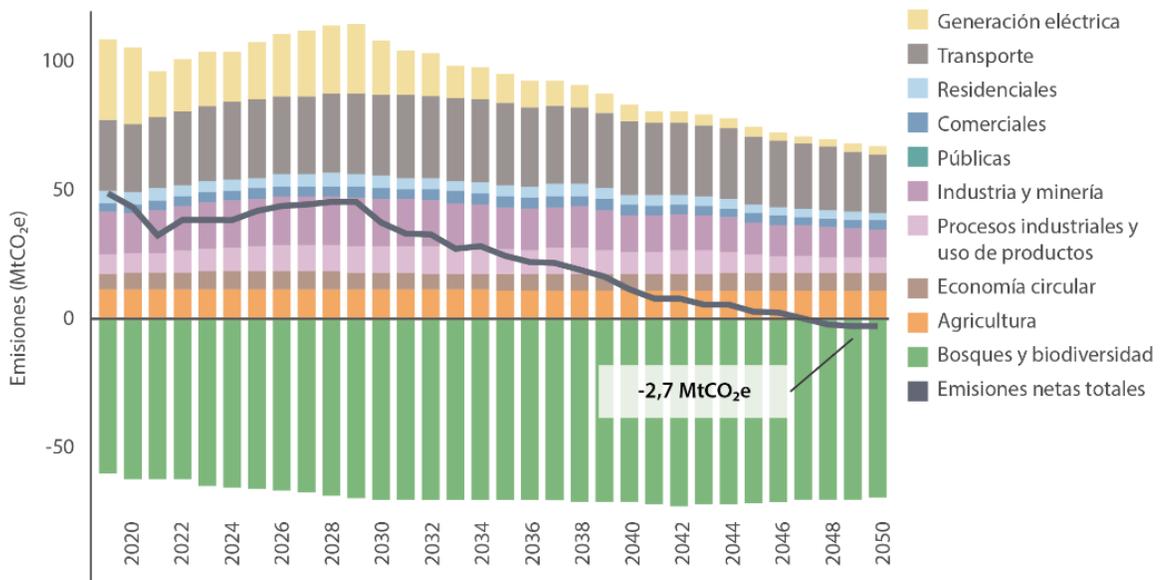


Figura 2: Resultados de emisiones según supuestos de referencia, agrupado por sectores. Fuente: Banco Interamericano de Desarrollo [4].

También, es posible observar, en Figura 2, la disminución de emisiones, principalmente en los casos de generación eléctrica (debido al ingreso de centrales de energía renovable y retiro de centrales térmicas), de transporte (dado el fomento a la electromovilidad). Así también, el aumento de captura, mediante bosques y biodiversidad (reforestación).

En Figura 3, se presentan los distintos escenarios mediante la aplicación de incertidumbre en los parámetros de modelación y el cumplimiento de las metas asociadas a cada medida de mitigación propuesta en el NDC. Algunos ejemplos de parámetros de modelación sectoriales son PIB, la población, el rendimiento vehicular, etc. Es posible observar que, para el caso NDC<sup>1</sup>, la mayoría de las emisiones netas se encuentran sobre cero al año 2050, siendo entonces más susceptible a no cumplirse la carbono neutralidad ante un cambio en las medidas de mitigación. Mientras que para el caso NDC+<sup>2</sup>, se presenta una mayoría de escenarios con emisiones bajo cero al año 2050, siendo entonces de esperar con mayor probabilidad el cumplimiento de carbono neutralidad.

<sup>1</sup> Refiere a las medidas declaradas o comprometidas en la Contribución Determinada a Nivel Nacional (NDC por sus siglas en inglés).

<sup>2</sup> NDC+ corresponde a una estrategia que incorpora medidas adicionales de mitigación de carbono respecto a las medidas de la estrategia NDC.

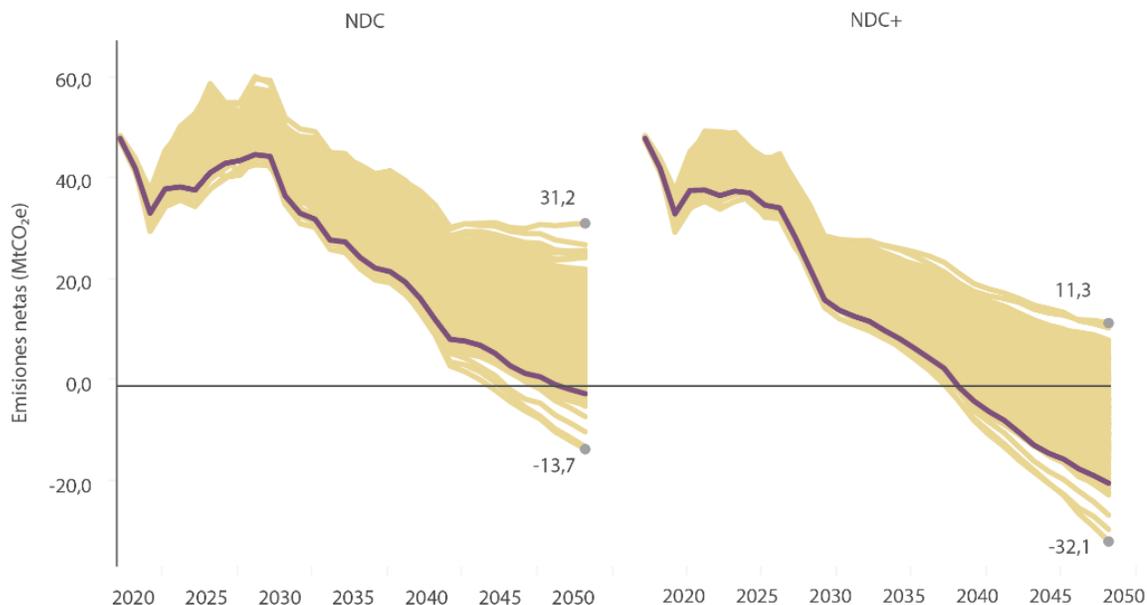


Figura 3: Resultados de emisiones ante distintos escenarios, según medidas de mitigación originales (NDC), y originales + adicionales (NDC+). Fuente: Banco Interamericano de Desarrollo.

Esta herramienta es fruto de distintas disciplinas, tales como ingenierías eléctrica, mecánica, informática, agronómica e industrial, principalmente, dado su alcance multisectorial. Tiene un fuerte componente energético, relacionado a la disciplina de ingeniería civil eléctrica, compuesto por conceptos de electromovilidad, eficiencia energética, industria sostenible, generación eléctrica, e hidrógeno verde. También, abarca conocimientos económicos, principalmente en la evaluación de CAPEX y OPEX<sup>3</sup> de cada una de las trayectorias.

<sup>3</sup> CAPEX y OPEX (por sus siglas en inglés *Capital Expenditure* y *Operational Expenditure*), corresponden a los costos de inversión y operación, respectivamente.

## 1.2. Descripción del problema

El modelo desarrollado en estudio previo [4] permite evaluar distintos escenarios mediante la aplicación de incertidumbre a cada uno de los parámetros y medidas de mitigación sectoriales, definiendo un rango posible y tomando muestras aleatorias dentro de este. Los escenarios de referencia son del tipo determinísticos, y corresponden a una selección particular de las variables inciertas subyacentes a la modelación de los sectores.

Luego, al aplicar incertidumbre, la cantidad de escenarios posibles crece exponencialmente a medida que aumenta la cantidad de parámetros y cantidad de muestras.

Para ejemplificar el crecimiento del número de escenarios posibles, se presenta, en Figura 4, las posibles combinaciones a partir de la incertidumbre de los parámetros de modelación de 3 sectores, con distinta cantidad de parámetros. Supongamos que cada parámetro puede tomar solo 2 valores o muestras, que los sectores industria y minería, y transporte, están representados por 2 parámetros cada uno, y el sector generación eléctrica está representado por un solo parámetro. Se puede observar que existen 32 potenciales escenarios dados por los valores que pueden tomar los parámetros antes descritos ( $2^5 = 32$ ).

Luego, el modelo se vuelve insostenible si se quisiera analizar cada posible combinación: por ejemplo, si se tuvieran 100 muestras por parámetro, 20 parámetros por sector, y 10 sectores, el total de escenarios posibles sería de  $100^{10 \cdot 20}$ , y el modelo aplicaría las ecuaciones a cada uno de esos casos (ecuaciones de demanda, emisiones, participación, costos de inversión y operación, entre otros). Los recursos computacionales serían restrictivos, así como el tiempo de simulación sería de horas o días.

Este problema se conoce como la "maldición de la dimensionalidad", que es inherente a los problemas combinatoriales. Esto es, explosión geométrica o exponencial del número de combinaciones al incrementar las variables (o dimensiones) del problema. Así, su exploración exhaustiva se hace impráctica o definitivamente imposible.

Como solución, el modelo desarrollado en estudio previo considera un número finito de escenarios (mil escenarios), seleccionando en cada uno de estos escenarios, de forma aleatoria, una muestra de cada parámetro, sin repetición.

En el ejemplo presente en Figura 4, se representa esta selección de muestras por parámetro para cada escenario, de forma simplificada (2 muestras, con repetición). Se observa que no se tienen todas las combinaciones posibles, y que, si las muestras "a" y "b" representaran escenarios optimistas y pesimistas, no se habrían modelado escenarios completamente optimistas ni completamente pesimistas.

Sector	Parámetros	Muestras
Industria y minería	1	1a
		1b
	2	2a
		2b
Transporte	3	3a
		3b
	4	4a
		4b
Generación eléctrica	5	5a
		5b

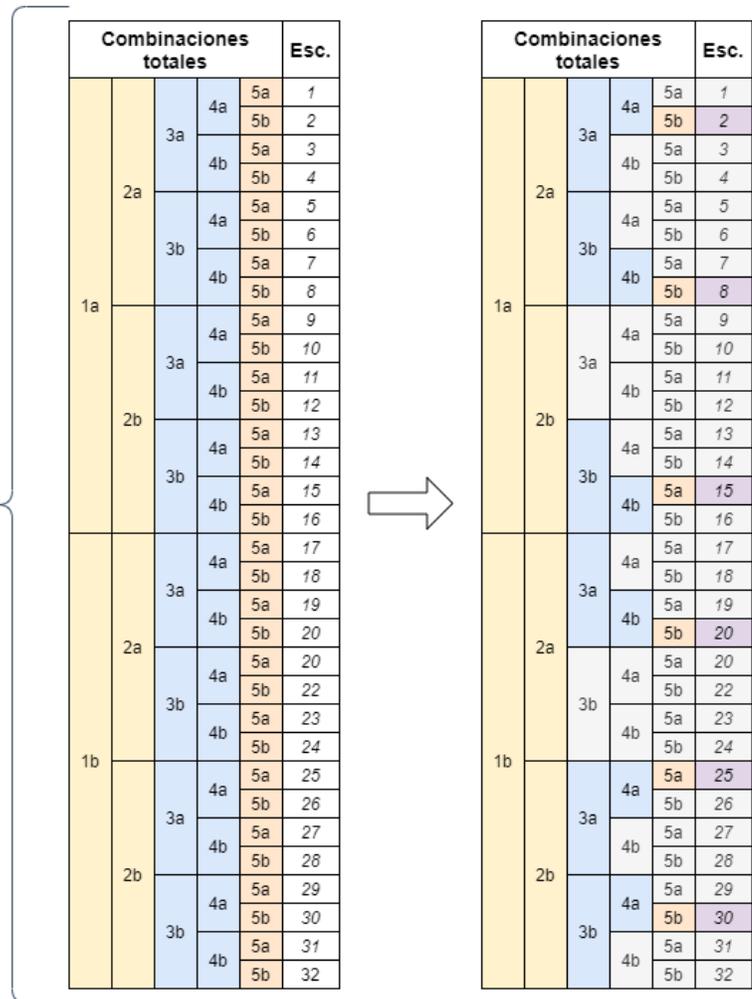


Figura 4: Diagrama de combinación completa de escenarios por sector a partir de la incertidumbre de 5 parámetros asociados al sector generación eléctrica, transporte e industria y minería (ejemplo). Fuente: Elaboración propia.

Finalmente, lo anterior determina el problema a resolver, que es la baja representatividad de escenarios posibles, tanto en la combinación de parámetros (limitado por la infinita cantidad de escenarios a simular), como por la baja probabilidad de modelar escenarios puramente críticos, donde todos los sectores se encuentren en un escenario máximo o mínimo de su rango de incertidumbre, al mismo tiempo.

### 1.3. Objetivos

La presente memoria tiene como objetivo general:

- Desarrollar una metodología que permita obtener los rangos máximos y mínimos entre los cuales pueden variar las emisiones de carbono equivalente, los costos de inversión y de operación en Chile, considerando sectores productivos y energéticos, dentro de rangos determinados de incertidumbre en sus parámetros, en un modelo de proyección de escenarios de carbono neutralidad.

Como objetivos específicos se encuentran:

- Aplicar metodología propuesta en caso de estudio. Comparar modelo original y modelo propuesto bajo una misma entrada y número de salidas, conservando valores de referencia, rangos de incertidumbre y cantidad de escenarios a simular.
- Identificar parámetros de mayor incidencia en escenarios de resultados máximos o mínimos en cuanto a las métricas de desempeño, las cuales son: emisiones de carbono equivalente, CAPEX y OPEX en el horizonte de evaluación (2020 a 2050), y emisiones de carbono equivalente en el año 2050.
- Conservar información propia del modelo sin modificaciones, tales como trayectorias y valores de referencia, modelos sectoriales y medidas de mitigación.

### 1.4. Estructura del informe

El documento de memoria está estructurado en seis capítulos, iniciando con la presente introducción. El Capítulo 2 detalla el marco teórico en el cual se basa la metodología, junto al estado del arte respecto a problemas similares. El Capítulo 3 por su parte, detalla la metodología propuesta, expresando la aplicación y justificación del modelo propuesto, expresando además los resultados esperados. El Capítulo 4 presenta los resultados obtenidos a nivel de simulación y operación del modelo propuesto, contrastando con resultados del modelo original o de referencia. Luego, en Capítulo 5 se presenta un análisis crítico de estos resultados, para finalmente, en Capítulo 6, finalizar con las principales conclusiones y trabajo futuro.

# Capítulo 2

## Marco Teórico y Estado del Arte

### 2.1. Modelo de proyección de escenarios

El modelo objeto de estudio, cuyo diagrama se presenta en Figura 5, está compuesto por los siguientes sectores: generación eléctrica, comercio, industria y minería, público, residencial y transporte. Para el sector generación eléctrica se utiliza el Modelo Energético PMR, el cual corresponde a un modelo de optimización implementado en GAMS [5]. Para el resto de los sectores se desarrollaron modelos implementados en lenguaje de programación Python. Si bien el Modelo Energético PMR también tiene representados los sectores transporte, industria y minería, comercial, público y residencial, por razones de tiempo de simulación se prefirió no utilizar un enfoque de optimización para estos sectores.

El modelo completo además incluye sectores como bosques y biodiversidad, agricultura, residuos e IPPU (o de procesos industriales y uso de productos), pero estos no serán objeto de estudio de la memoria, debido a que se modelan según procesos distintos.

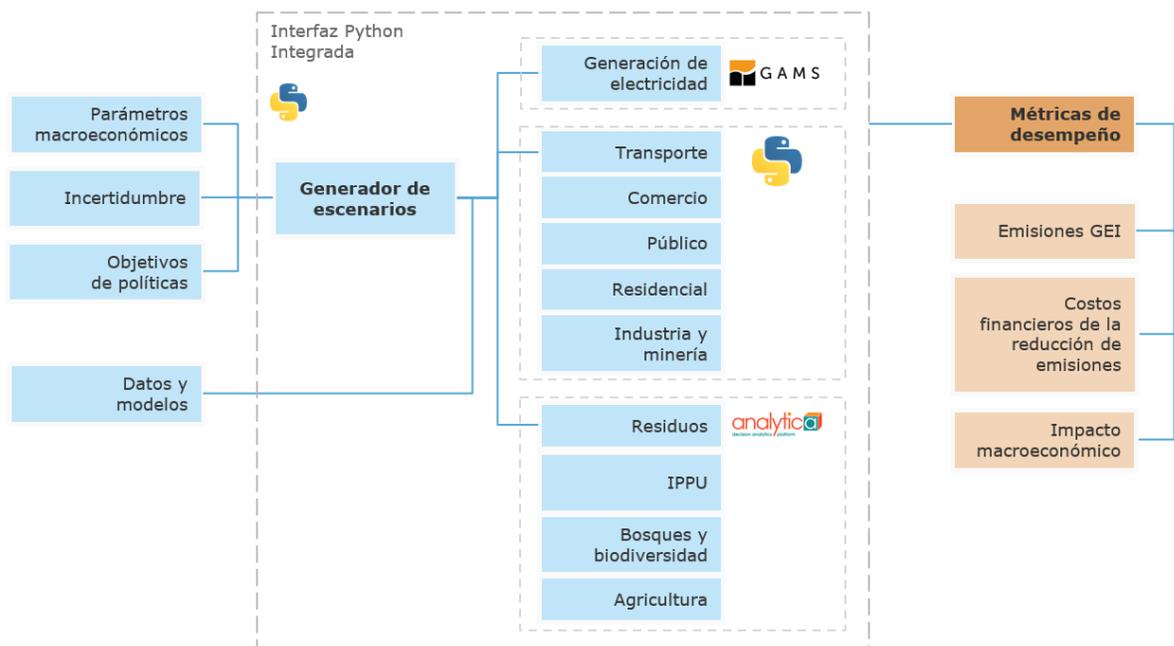


Figura 5: Diagrama de la plataforma de modelación para el análisis de la Estrategia de Descarbonización de Largo Plazo.

El modelo pertenece a lo que se conoce como tomador de decisiones sólidas o *robust decision-making*. Es decir, es un modelo que permite identificar estrategias

fundamentadas, detectar vulnerabilidades y reconocer comportamientos mutuos o correlaciones entre las distintas variables. Si bien el modelo se integró a este esquema, no lo integra intrínsecamente, ya que su principal función es la de proyectar escenarios.

Para lograr lo anterior, se introduce incertidumbre mediante el método de muestreo de hipercubo latino (LHS), modificando parámetros estratégicamente a fin de identificar las vulnerabilidades o riesgos del cumplimiento de carbono neutralidad al año 2050.

## 2.2. Muestreo de hipercubo latino (LHS)

El método de muestreo de hipercubo latino (LHS por su definición en inglés: *latin hypercube sampling*), descrito en 1979 por Michael McKay del Laboratorio Nacional de Los Álamos, es un método estadístico para generar una distribución de muestreo cuasi aleatorio, permitiendo diseñar experimentos computarizados o para simulaciones de Montecarlo. El LHS se construye de la siguiente manera: se divide un espacio en **n** secciones, donde **n** representa el número de muestras, y luego, se toma solo un valor en cada sección o intervalo. Este valor se encuentra entre 0 y 1.

En Tabla 1 se presenta un ejemplo de matriz LHS al ejecutar la función *lhs* de la librería de python *pyDOE*, que recibe el número de parámetros (4) y muestras (10), para luego discretizar los rangos entre 0 y 1 de cada parámetro según la cantidad de muestras, y tomar el valor promedio de cada intervalo discreto (*'center'*).

Tabla 1: Ejemplo de matriz LHS o muestreo de hipercubo latino, para 10 escenarios (o muestras) y 4 parámetros.

<b>LHS</b>	<b>P1</b>	<b>P2</b>	<b>P3</b>	<b>P4</b>
<b>E1</b>	0.05	0.05	0.15	0.15
<b>E2</b>	0.55	0.85	0.95	0.75
<b>E3</b>	0.25	0.25	0.45	0.25
<b>E4</b>	0.45	0.35	0.75	0.45
<b>E5</b>	0.75	0.55	0.25	0.55
<b>E6</b>	0.95	0.45	0.35	0.05
<b>E7</b>	0.35	0.95	0.05	0.65
<b>E8</b>	0.15	0.65	0.55	0.35
<b>E9</b>	0.85	0.75	0.85	0.85
<b>E10</b>	0.65	0.15	0.65	0.95

El método LHS selecciona un valor del rango discretizado aleatoriamente para cada muestra, e itera hasta seleccionar todos los valores.

Algunas variantes del método LHS permiten tomar valores dentro de cada sección que aumenten la distancia entre muestras, o cumplan alguna función respecto de la variable asignada (útil en modelos de optimización o análisis de dependencia).

De esta forma, es posible determinar los distintos valores que puede tomar un parámetro dentro de un rango, bajo distribución uniforme, es decir, cada escenario o muestra tiene la misma probabilidad de ocurrir.

Cabe destacar que el promedio de las muestras es de 0.5, siendo esta la esperanza de una variable con distribución uniforme de rango entre 0 y 1.

### 2.2.1. Proyección de escenarios utilizando LHS

Por otra parte, es posible construir un vector de valores a partir de un valor medio y su desviación estándar, aplicando distribución normal a un vector de muestras de hipercubo latino.

La transformación usada se presenta en Figura 6, la cual consiste en pasar de cada punto azul o muestra de LHS, a la curva de función de distribución acumulada (CDF por sus siglas en inglés), y luego, a su respectivos valores en el eje x, correspondientes a valores que distribuyen normalmente.

En el ejemplo, se utilizan 6 muestras, centradas en cada uno de los intervalos de la discretización entre 0 y 1 (muestreo de hipercubo latino).

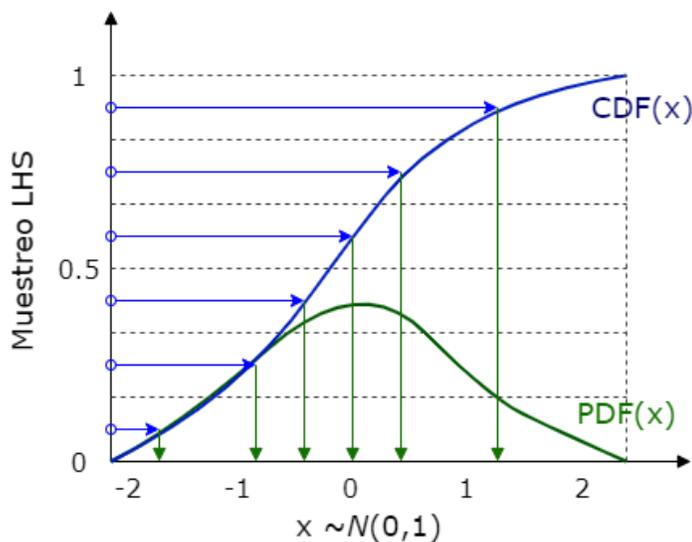


Figura 6: Transformación de muestras LHS a distribución normal, de media igual a 0 y desviación estándar igual a 1.<sup>4</sup>

---

<sup>4</sup> CDF: *cumulative distribution function*. PDF: *probability density function*.

Como ejemplo de esta transformada se tienen las variaciones presentes en Figura 7. La figura representa 4 parámetros, donde  $x$  y  $s$  son el valor medio y la desviación estándar, respectivamente.

Es posible observar que el vector generado no necesariamente toma valores entre 0 y 1, tal como se presenta en los parámetros  $c$  y  $d$ . Para efectos de construir una nueva matriz de LHS utilizando esta técnica, es necesario filtrar o modificar las muestras, cumpliendo con el rango entre 0 y 1, y que cada muestra tome un valor dentro de la discretización del rango.

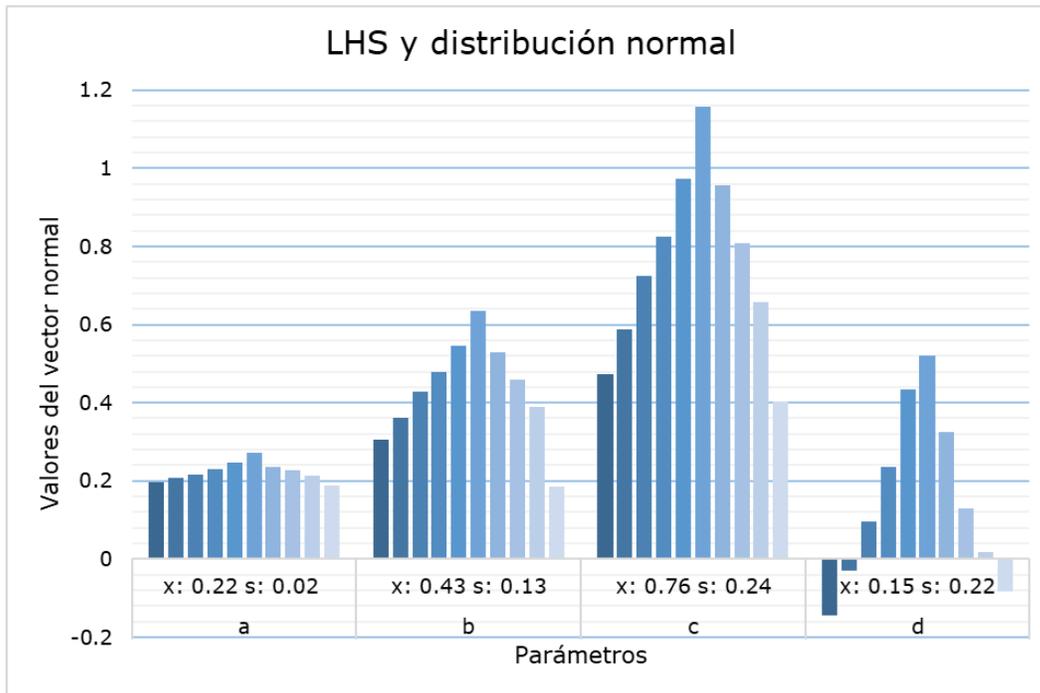


Figura 7: Ejemplo de vector que distribuye normal utilizando muestras de LHS.

Luego, es posible expandir un escenario representativo mediante la técnica mencionada, considerando el valor de cada parámetro del escenario como el valor medio del vector normal, y considerando la desviación estándar del parámetro en el grupo al cual representa. De esta forma, la proyección de nuevos escenarios presentará mayor diferencia de valor en los parámetros de mayor desviación estándar, y mantendrá un valor similar en los parámetros de menor desviación estándar, explorando nuevas combinaciones al mantener "fijos" los valores que se repiten (menor desviación estándar).

Por ejemplo, considerar el parámetro PIB de un escenario representativo del sector transporte, dentro de los escenarios de máximas emisiones GEI del mismo sector. Supongamos que el escenario representativo tiene un valor de PIB de 0.22, en cuanto a su muestra de LHS. Este valor definirá el valor promedio del nuevo vector normal. Además, se tiene una desviación estándar de 0.02 entre todos los escenarios del grupo (escenarios de máximas emisiones del sector transporte). Luego, se tendrían los valores presentes en el parámetro  $a$  de la Figura 7, para el parámetro PIB, en los nuevos escenarios proyectados.

## 2.3. Algoritmos de reducción de escenarios

Una técnica útil para detectar patrones de similitud entre escenarios es la reducción de escenarios. Se trata de agrupar en un conjunto los escenarios de características similares, y luego, definir un subconjunto de escenarios representativos cuya información sea suficiente para representar el conjunto original.

Un ejemplo de esto se presenta en [6], con una formulación matemática para la generación de árboles de escenarios en programación estocástica, y posterior reducción de escenarios por análisis de sensibilidad.

La reducción de escenarios por agrupación o clustering, presente en este caso, consiste en seleccionar muestras o resultados representativos según similitudes dentro de un mismo factor o atributo. Por ejemplo, en Figura 8, se presenta una descomposición en regiones de Voronoi, obtenidas mediante algoritmo de selección *forward* en [6].

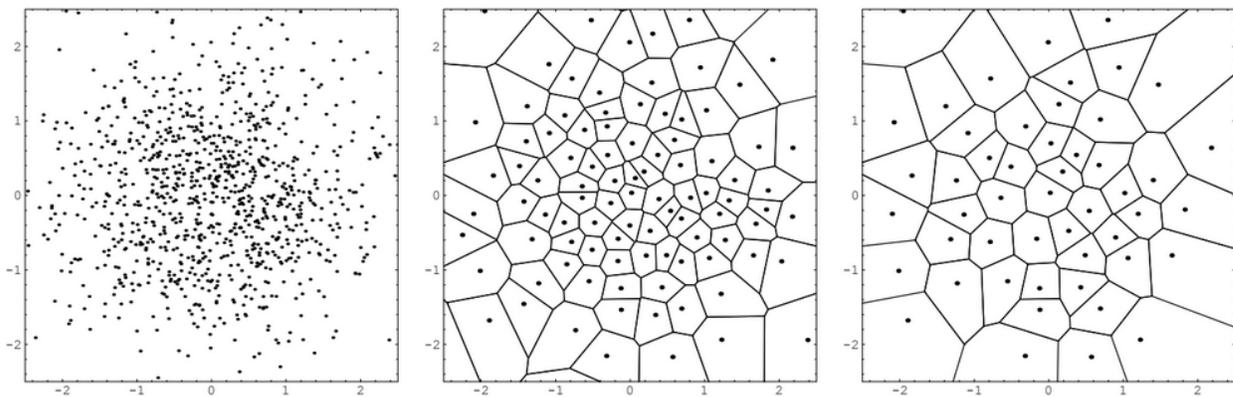


Figura 8: Reducción de escenarios partiendo de 1.000 escenarios de muestra de la distribución normal estándar bidimensional (izquierda) reducidos a 100 escenarios (centro) y reducidos a 50 escenarios (derecha). Se muestran las correspondientes descomposiciones de  $\mathbb{R}^2$  en regiones de Voronoi obtenidas por el algoritmo de selección hacia delante con respecto a la norma euclidiana.

### 2.3.1. Aprendizaje no supervisado

Los métodos de aprendizaje no supervisado se caracterizan por modelos que se ajustan a los elementos u observaciones, ya que estos presentan una característica aleatoria y/o no se reconoce un comportamiento típico o predecible. Los modelos se utilizan, por ejemplo, para agrupar datos por afinidad, definiendo una distancia que le sirva para comparar los datos entre sí.

En caso contrario, en un aprendizaje supervisado existe conocimiento de si un elemento representa una situación específica, como por ejemplo si un paciente tiene una enfermedad, o si un conjunto de puntos representa una figura.

Como ejemplo de aprendizaje no supervisado se tienen los algoritmos de *clustering* o de agrupación, los cuales se utilizan para descubrimiento de tópicos, detección de anomalías o caracterización de conjuntos de datos multidimensionales.

Además, se tiene como ejemplo de aprendizaje no supervisado la reducción de dimensionalidad, utilizando técnicas matemáticas y estadísticas convirtiendo un conjunto de datos original en un nuevo conjunto de datos de menor dimensión con cierta pérdida de información.

El aprendizaje no supervisado aplica al modelo presente en este estudio, ya que se presenta un fuerte componente aleatorio mediante la aplicación de incertidumbre por LHS, y se tienen resultados multidimensionales, como lo son las métricas de desempeño: emisiones GEI, CAPEX, OPEX y variables macroeconómicas.

A continuación, se presentarán los pasos necesarios para identificar adecuadamente escenarios representativos, mediante algoritmo de *clustering k-means* y reducción de dimensionalidad utilizando PCA (*Principal Components Analysis*).

### 2.3.2. Distancia euclidiana

Por lo general, la similitud entre elementos corresponde a la distancia de valores numéricos, por lo que suele trabajarse con métricas vectoriales, como la distancia euclidiana, definida por la siguiente expresión:

$$\|x\| = \sqrt{(x_1 - x_2)^2}$$

Donde  $x_1$  y  $x_2$  representan una variable  $x$ , de los elementos 1 y 2. Una variable puede ser, por ejemplo, la altura de una persona. Entonces, si  $x_1 = 1.54$  es la altura de la persona 1, y  $x_2 = 2.04$  es la altura de una persona 2, la distancia euclidiana para la altura entre ambas personas es  $\|x\| = \sqrt{(1.54 - 2.04)^2} = \sqrt{(-0.50)^2} = 0.50$ .

Si se tienen  $n$  variables para cada elemento (por ejemplo, para caracterizar personas según altura, edad y peso), la distancia euclidiana entre dos elementos está dada por:

$$\|X\| = \sqrt{(x_{11} - x_{21})^2 + (x_{12} - x_{22})^2 + \dots + (x_{1n} - x_{2n})^2}$$

De esta forma, es posible comparar la similitud de 2 escenarios en términos de sus métricas de desempeño (emisiones GEI, CAPEX y OPEX), o en términos de sus muestras de LHS, por ejemplo.

No obstante, es imprescindible normalizar los datos, ya que la distancia euclidiana, por su definición de resta absoluta entre 2 atributos o variables, puede tener mayor valor en variables de alta magnitud (ej.: CAPEX anual de miles de MMUSD), que en variables de menor magnitud (ej.: emisiones GEI de decenas de Mton-CO2).

### 2.3.3. Normalización de datos

Si se quisiera comparar la distancia entre dos escenarios, en términos de emisiones y CAPEX, por ejemplo, se tendría que la distancia en cuanto a CAPEX sería mayor a la distancia de emisiones, ya que las emisiones anuales son menores a 100, mientras que el CAPEX anual sobrepasa los 1000, en sus respectivas unidades de medida.

Luego, para aplicar la distancia euclidiana es sumamente importante normalizar los valores de cada atributo.

En el trabajo presente se utilizó la función *StandardScaler* de la librería *sklearn* para lenguaje python.<sup>5</sup> Esta función, permite llevar una serie o vector de datos (x) a un nuevo conjunto (z) de promedio  $\mu$  igual a cero y desviación estándar  $\sigma$  igual a uno, siguiendo la ecuación:

$$z = \frac{x - \mu}{\sigma}$$

De esta forma, es posible comparar atributos de distinta magnitud, y devolver los datos a su conjunto original, reconociendo el promedio y desviación estándar aplicado.

Por ejemplo, se presenta en Tabla 2 la transformación de una matriz de datos de columnas a, b y c, a una matriz normalizada de columnas z(a), z(b) y z(c), mediante la aplicación de la ecuación previamente definida.

Tabla 2: Ejemplo de normalización de matriz de datos.

ID	Matriz original			Matriz normalizada		
	a	b	c	z(a)	z(b)	z(c)
<b>0</b>	3.25	6.16	3.70	1.38	-0.20	1.10
<b>1</b>	-1.22	3.50	0.38	-0.59	-0.96	0.55
<b>2</b>	-1.06	8.86	-0.20	-0.51	0.58	0.45
<b>3</b>	-2.15	3.94	-8.10	-0.99	-0.84	-0.86
<b>4</b>	1.73	11.73	-10.43	0.71	1.41	-1.24
<b>Promedio</b>	<b>0.11</b>	<b>6.84</b>	<b>-2.93</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
<b>Desv. Estándar</b>	<b>2.28</b>	<b>3.46</b>	<b>6.03</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>

<sup>5</sup> Más información en página oficial de scikit learn: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>

Los datos se grafican en Figura 9, donde se puede observar que los datos de las columnas a, b y c difieren considerablemente de magnitud e incluso de signo, reflejado en su promedio y desviación estándar. Luego, al aplicar la normalización, los datos de las columnas z(a), z(b) y z(c) presentan una magnitud similar, permitiendo una comparación de distancia entre distintos ID o filas, sin tendencia a las variables de mayor magnitud.

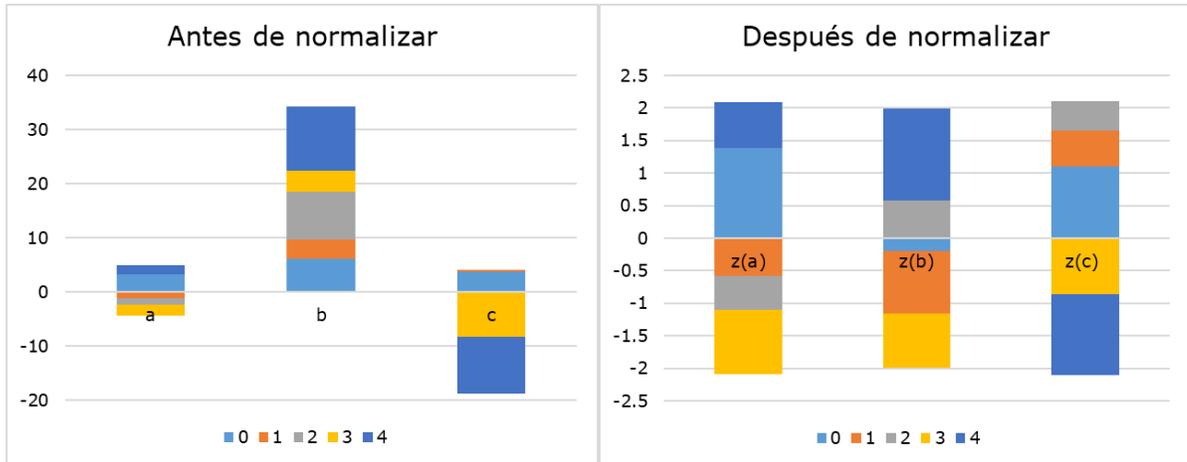


Figura 9: Comparación de datos al aplicar normalización.

### 2.3.4. Clustering k-means

Un algoritmo de *clustering* es un método de aprendizaje no supervisado que tiene como objetivo agrupar elementos de un conjunto de datos, según criterios de similitud, de forma que los elementos que hay dentro de cada grupo o clúster sean más parecidos entre ellos que cada uno de ellos respecto a otros grupos.

El método de *clustering k-means* busca minimizar la suma de la varianza de cada clúster, cuya función se presenta en Ecuación 1. Corresponde a un método no jerárquico que agrupa el conjunto de datos en k clústeres distintos y distinguibles, es decir, que cada dato pertenece a un solo grupo o clúster (ver Figura 10). El valor de k se fija por diseño del programador, utilizando métricas de validación.

Ecuación 1: función objetivo *clustering k-means*.

$$\min \sum_{k=1}^k W(C_k)$$

Donde,

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

De la Ecuación 1 se tiene que  $\mu_k$  representa el promedio o centroide del clúster k, y  $x_i$  representa cada dato o vector de datos (en caso de filas múltiples métricas o

dimensiones) dentro del clúster. Luego, se tiene que  $W(C_k)$  corresponde a la suma de la distancia al cuadrado de cada vector de datos respecto a su centroide.

Se puede observar en Figura 10 la presencia de datos que están más cerca de un centroide que no pertenece a su clúster. Esto quiere decir que la asignación de datos puede ser mejorada y por tanto identificar una menor función objetivo.

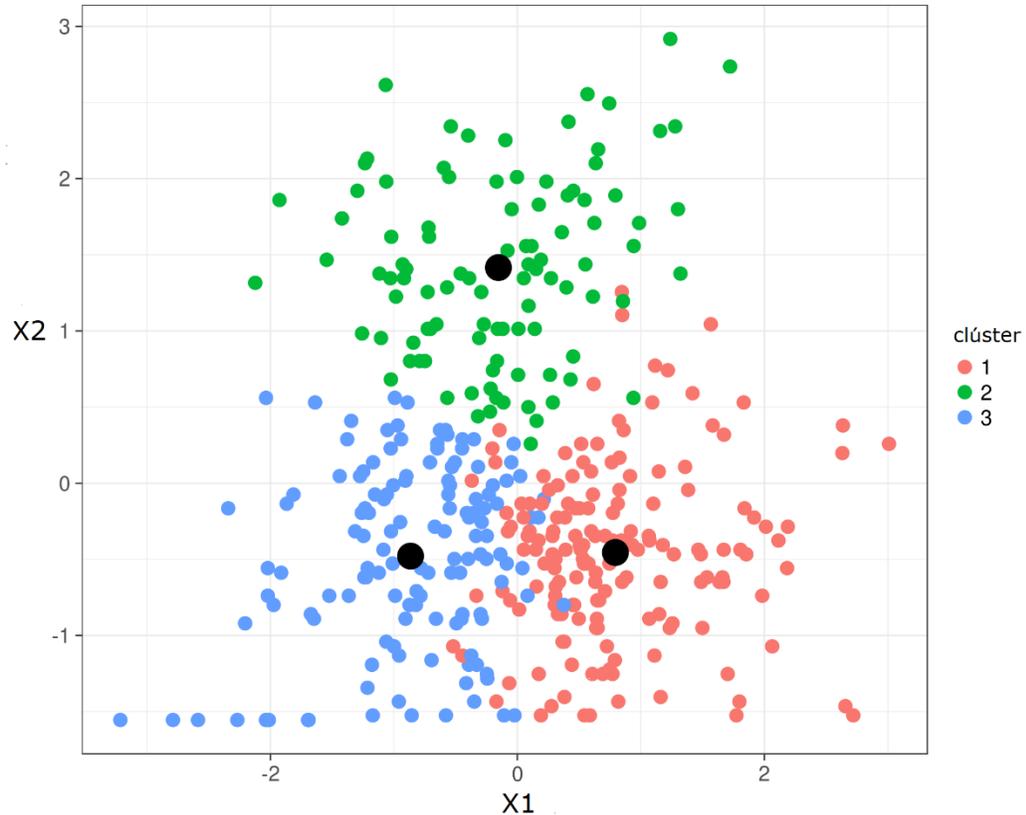


Figura 10: Ejemplo de resultado del algoritmo de *clustering k-means*, para  $k=3$  (número de grupos), donde los puntos negros son los centroides o puntos medios de cada clúster.

La asignación de datos a cada grupo se realiza de forma iterativa un número finito de veces, seleccionando la asignación de datos que presente una menor función objetivo. La cantidad de iteraciones, cantidad de grupos y selección inicial de centroides son parte esencial de la eficiencia del algoritmo, existiendo innumerables técnicas y estudios relacionados.

Utilizando esta técnica de reducción de escenarios, es posible identificar escenarios representativos independiente de la distribución que presenten sus variables. Es decir, un escenario representativo puede ser único (poco común) o puede estar representando muchos escenarios repetidos o de valores similares (escenarios comunes).

### 2.3.5. Selección de k y distribución de datos

A continuación, se presentan dos métricas de validación del algoritmo de *clustering k-means*, y consideraciones relevantes para el presente estudio. Estas métricas de validación permiten justificar la selección de número de grupos o clústeres (valor de  $k$ ), así como evaluar la distribución de los datos que se busca agrupar.

- Distorsión:

La distorsión está relacionada con la función objetivo del algoritmo de clustering. Esta función objetivo corresponde a la suma de las distancias al cuadrado de cada escenario con el centroide del grupo o clúster asignado. La distorsión corresponde al promedio de estas sumas de distancias.

Es decir, la distorsión indica que tan cerca están los escenarios del promedio o centroide de su respectivo clúster.

Por lo tanto, a mayor cantidad de grupos, menor será el valor de la función objetivo y de la distorsión, llegando a cero en caso de que se tengan tantos grupos como escenarios.

Dado que se busca minimizar la función objetivo mediante la asignación óptima de escenarios a cada grupo o clúster (usualmente, esta asignación se realiza de forma iterativa e inicialmente de forma aleatoria), se busca un bajo valor de distorsión considerando una baja cantidad de grupos, logrando una efectiva agrupación o clasificación.

En términos de reducción de escenarios, conviene asignar una cantidad de grupos cuya función objetivo se encuentre dentro del codo de la curva de distorsión, ya que indica que, de agregar un nuevo grupo, la ganancia será menor que la pérdida de quitar un grupo.

- Método Silhouette:

El método Silhouette representa la comparación entre la distancia de cada escenario con los clústeres más cercanos, y la distancia del escenario con los escenarios de su mismo clúster.

El coeficiente o indicador de Silhouette denotado por  $s$ , para una muestra está dado por Ecuación 2.

Ecuación 2: Indicador de Silhouette.

$$s = \frac{b - a}{\max(a, b)}$$

Donde **a** corresponde al promedio de la distancia entre la muestra y el resto de los elementos del mismo clúster. Y **b** corresponde a la distancia entre la muestra y el resto de las muestras del clúster más cercano.

Luego, el indicador de Silhouette de un conjunto de datos corresponde al promedio del indicador de todas las muestras.

Se pueden dar los siguientes casos:

- Si  $s$  tiende a -1, los elementos se encuentran en clúster equivocado, ya que están más cerca de un clúster cercano que del propio.
- Si  $s$  tiende a 0, los clústeres pueden presentar solapamiento.
- Mientras que, si  $s$  tiende a 1, los elementos se encuentran en clústeres adecuados en cuanto a la similitud de sus pares y diferencia con otros clústeres.

Se busca el valor más cercano a uno, ya que esto significa que la distancia promedio de cada elemento respecto a los otros grupos, es mayor que la distancia del elemento con los de su mismo grupo.

Finalmente, ambas métricas pueden ser complementarias al escoger el número de clústeres, así como con restricciones del diseño y objetivos de la reducción de escenarios.

- Restricciones de diseño

Siguiendo la línea de aprendizaje no supervisado, puede ocurrir que, al realizar tratamiento a un conjunto de datos, se deba recurrir a simplificaciones o restricciones acordes a los objetivos del estudio.

Por ejemplo, en contexto de categorizar clientes de una tienda, puede ocurrir que solo se tuviese presupuesto para 4 campañas publicitarias. Luego, sería conveniente clasificar a los clientes a un número cercano de 4 grupos, y si hay más grupos, priorizar los que posean características atractivas en términos de marketing.

En el caso presente, si se trabajan con 1000 escenarios, y se quisiera proyectar al menos 5 escenarios por cada escenario representativo, se deberían escoger hasta 200 grupos y así obtener 200 escenarios representativos o centroides.

### 2.3.6. Análisis de componentes principales (PCA)

Un punto débil del algoritmo de *clustering k-means*, es la capacidad de agrupar datos de múltiples y numerosas dimensiones o variables. Esto ya que a medida que aumenta la cantidad de variables, debe recurrir a una mayor cantidad de grupos con tal de representar cada una de las combinaciones de las variables.

En el caso del presente estudio, se tienen sectores como industria y minería o el sector transporte, que presentan 50 parámetros con rango de incertidumbre distinto de cero. Luego, el algoritmo de *clustering* buscaría identificar similitudes entre escenarios para al menos 50 resultados distintos, lo cual es difícil independiente de la cantidad de escenarios contemplados. Para estos casos, es conveniente utilizar PCA.

### 2.3.6.1. Reducción de dimensionalidad

El análisis de componentes principales corresponde a un método de reducción de dimensionalidad, que permite explicar la misma información de  $n$  muestras u observaciones con  $p$  parámetros, en  $z$  componentes, como se presenta en Figura 11.

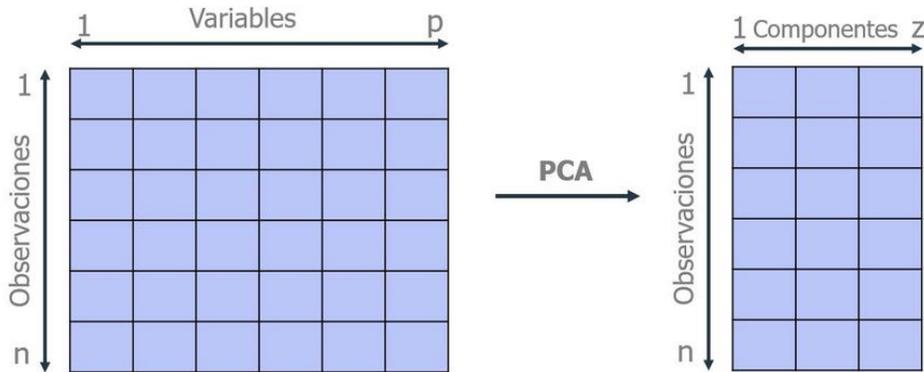


Figura 11: Reducción de dimensionalidad asociada a método PCA.

Cada componente principal corresponde un vector que permite explicar la varianza de un conjunto de datos. Por ejemplo, en Figura 12, las flechas verde y roja representan las direcciones de dos componentes principales ortogonales (condición del método PCA). La flecha roja corresponde a la primera componente principal, y es el vector que mayor varianza de los datos explica. Luego, la flecha verde es la segunda componente principal, y es el segundo vector que explica mayor varianza.

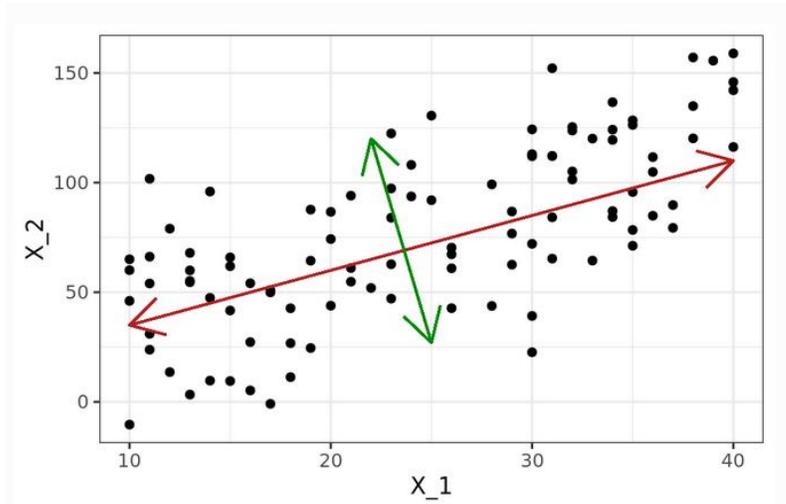


Figura 12: Interpretación geométrica de componentes principales.

Formalmente, cada componente principal es un vector propio que corresponde a una dirección de la matriz de datos asociada. Cuando se multiplica una matriz por alguno de sus vectores propios, se obtiene un múltiplo del vector propio, es decir, el resultado es ese mismo vector multiplicado por un número o escalar. Al valor por el

que se multiplica el vector propio resultante se le conoce como valor propio. A todo vector propio le corresponde un valor propio y viceversa.

Luego, las componentes principales se ordenan de mayor a menor valor propio, orden que coincide con las componentes principales que explican una mayor variabilidad. Es decir, la primera componente principal es el vector propio de valor propio más alto.

Por ejemplo, una forma gráfica de visualizar la información entregada por las componentes principales es el ejemplo de la galleta, presente en Figura 13. En esta región o espacio utilizado por la galleta, se tienen dos ejes de alta variabilidad y uno de baja variabilidad, por lo que se podría reducir la dimensión del problema a una proyección en el plano, despreciando el último eje ( $x_3$ ).

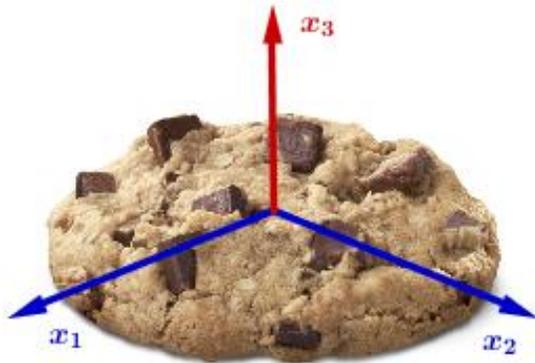


Figura 13: Ejemplo de PCA, componentes principales de una galleta.

#### 2.3.6.2. Varianza explicada acumulada

Por lo general, dada una matriz de datos de dimensiones  $n \times p$ , el número de componentes principales que se pueden calcular es como máximo  $n-1$  o  $p$  (el menor de los dos valores es el limitante) [7].

Un criterio para escoger el número de componentes principales utilizado es evaluando la proporción de varianza explicada acumulada. Un ejemplo de esta curva se presenta en Figura 14. El criterio es seleccionar la cantidad de componentes principales desde donde la varianza acumulada disminuye su crecimiento. No obstante, queda a criterio del diseñador y en función de los requerimientos del estudio.

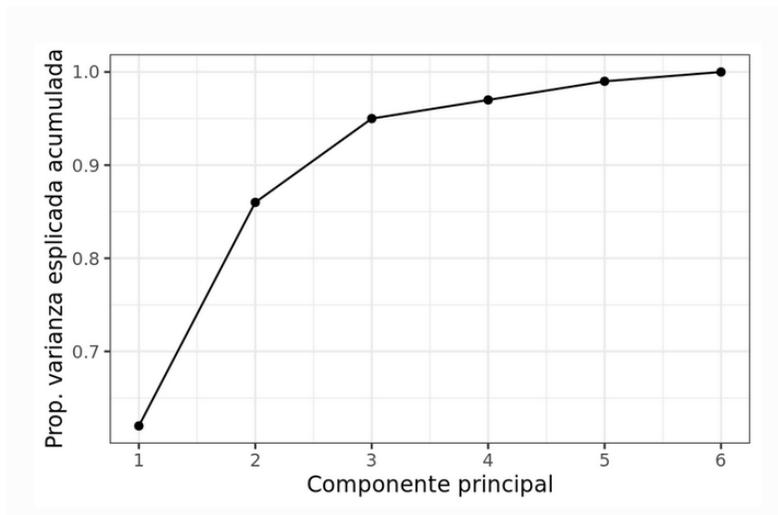


Figura 14: Varianza acumulada en función del número de componentes principales.

En el ejemplo de la Figura 14, se deberían escoger 2 componentes principales, ya que con esto se cubre cerca de un 85% de la información contenida en la matriz de datos.

### 2.3.6.3. Peso de cada variable

Otra utilidad del PCA es la de identificar el peso de cada parámetro o columna en la representatividad de la matriz de datos. Cada componente principal  $Z_i$  se obtiene por combinación lineal de los parámetros originales. La primera componente principal del grupo de parámetros  $[X_1, X_2, \dots, X_p]$  es la combinación lineal normalizada de dichos parámetros:

$$Z_1 = \varphi_{11}X_1 + \varphi_{21}X_2 + \dots + \varphi_{p1}X_p$$

La combinación lineal es normalizada si:

$$\sum_{j=1}^p \varphi_{j1}^2 = 1$$

Las demás componentes principales se obtienen de forma similar, con la restricción de ser ortogonales a las componentes principales calculadas previamente.

Los términos  $\varphi_{11}, \dots, \varphi_{p1}$  se denominan *loadings* y definen las componentes. Por ejemplo,  $\varphi_{11}$  es el *loading* de la variable  $X_1$  de la primera componente principal. Los *loadings* pueden interpretarse como el peso o importancia que tiene cada parámetro en cada componente y, por lo tanto, ayudan a conocer qué tipo de información recoge cada una de las componentes.

Finalmente, es posible reconocer los parámetros de mayor peso en la recopilación de información y variabilidad de una matriz de datos, mediante el estudio de sus *loadings* absolutos en las componentes principales consideradas.

## 2.4. Estado del arte

En esta sección, se presenta la revisión del estado del arte asociado a modelos de proyección, modelos de múltiples escenarios, y del estado del arte del método de muestreo de hipercubo latino. Se tiene un resumen de cada estudio revisado, más la aplicación o relación con el estudio presente.

De la revisión bibliográfica realizada, no se encontraron estudios que aborden específicamente la problemática de combinatoria en modelos de proyección de escenarios bajo incertidumbre.

### 2.4.1. Modelos de proyección y de escenarios

A continuación, se presentan soluciones a problemas asociados a modelamiento de escenarios futuros, mediante técnicas de reducción de escenarios o proyecciones estocásticas.

Se investigaron como opción de reemplazo al modelo de referencia, principalmente en cuanto al uso del método de muestreo de hipercubo latino (LHS) para la aplicación de incertidumbre a las trayectorias de referencia.

En [8], se agrega incertidumbre por medio de un modelo probabilístico, a un modelo determinístico asociado a las trayectorias de emisiones de GEI<sup>6</sup> en Chile. Se evalúan las variables endógenas y exógenas cuyo comportamiento es posible asociar a una variable aleatoria, y se asigna una función de probabilidad según la literatura asociada a expertos y estudios de proyección. De esta forma, sería posible obtener trayectorias en función de su probabilidad, identificando rangos mínimos y máximos según criterios estadísticos.

En [9] se propone una formulación basada en escenarios para la gestión óptima de un sistema, que consiste en programación no lineal estocástica en dos etapas. Maneja la gran escala del modelo de optimización bajo incertidumbre mediante la reducción de escenarios, según análisis de sensibilidad, sin perder una solución óptima final.

En [10], se utilizan cadenas de Márkov como entrada a una simulación de Monte Carlo, para determinar escenarios de despacho considerando una alta penetración de fuentes de generación solar y eólica. Este estudio considera la formulación de proyecciones de escenarios, donde a partir de algunas muestras del pasado se pueden obtener estados a futuro, iterativamente, utilizando probabilidades de transición.

En [11], se tiene un modelo de planificación de la generación y almacenamiento (PGA), donde se busca determinar la participación de cada tecnología en el mix óptimo, considerando las estocasticidades propias de los modelos de planificación de

---

<sup>6</sup> Gases de efecto invernadero.

un sistema eléctrico de potencia. Para esto, se aplica  $n$  veces el modelo de PGA determinístico en base a  $n$  escenarios distintos, detectando las principales características mediante información mutua. Luego, se aplica una reducción de escenarios mediante técnica de *clustering*. A esto, se aplica un análisis de post optimalidad basado en índices de adaptabilidad del PGA, y eventualmente se aplica un ajuste al modelo iterando el proceso.

## 2.4.2. Muestreo de hipercubo latino

En el estado del arte también se encuentran otras aplicaciones del muestreo de hipercubo latino, orientadas a modelos de optimización y de problemas probabilísticos. El modelo de referencia utiliza un LHS básico, es decir, sin modificaciones, por lo que se investiga la posibilidad de aplicar otros tipos de muestreos similares.

En [12], se utiliza el muestreo jerárquico de hipercubo latino<sup>7</sup> y muestreo de hipercubo latino con dependencia<sup>8</sup> para evaluar distintas formas de distribuir energía en un sistema de potencia, generando un conjunto de escenarios y contingencias de la red reducidos pero representativos, que cubren toda la gama de condiciones de funcionamiento y los extremos. Esta metodología permite reducir la dimensión de las entradas y gastos computacionales.

También respecto a sistemas de potencia, en [13] se utiliza el muestreo de hipercubo latino para modelar múltiples escenarios probabilísticos bajo incertidumbre, con enfoque en las características estocásticas e interactivas de la generación mixta, especialmente en redes con fuentes de energía renovable. Con esto, se propone una estrategia de respuesta activa a la demanda, con criterios de seguridad.

En [14], se plantea una estrategia para mejorar la eficiencia de un muestreo óptimo de hipercubo latino (OLHS), mejorando el rendimiento de convergencia en términos de tiempo. El diseño óptimo de hipercubo latino (OLHD) permite un considerar solo un número limitado de puntos de evaluación para representar el espacio total de posibilidades, principalmente en experimentos donde una tomar una muestra resulta muy caro.

Por último, destacar el resultado presente en [15], donde se estudia la eficiencia de métodos de agrupación (o *clustering*) modificado en problemas probabilísticos de flujos de potencia en redes de alto componente renovable, y su respectiva incertidumbre. Se tiene que el cálculo del flujo de potencia con los métodos de agrupación es más preciso que utilizando el método LHS y es más rápido que el método de simulación de Monte Carlo.

---

<sup>7</sup> Corresponde a conjuntos de LHS recursivamente divisibles en subconjuntos de LHS.

<sup>8</sup> Realiza permutaciones de una muestra de hipercubo latino ya generada, según análisis estadístico.

# Capítulo 3

## Metodología y Aportes del Trabajo de Memoria

### 3.1. Formalización del problema

#### 3.1.1. Combinación de parámetros

El modelo de referencia utilizado en esta memoria [4] realiza una discretización de los rangos de incertidumbre aplicados a algunos parámetros específicos. Utilizando la metodología de muestreo de hipercubo latino (LHS), esta discretización asigna valores aleatoriamente a cada parámetro, preocupándose de que ningún valor se repita.

El modelo de referencia tiene un total de 862 parámetros, donde 161 presentan rangos de incertidumbre mayores a cero para el horizonte de evaluación 2019-2050, es decir, los valores de los parámetros cambian según los escenarios explorados. Un escenario se entiende como la simulación del modelo a partir de un conjunto de datos diferente al de los demás escenarios, en al menos un dato o parámetro. Por ejemplo, se pueden simular distintos escenarios de recurso hídrico, desde extrema sequía a un año sumamente lluvioso.

Si cada uno de estos 161 parámetros pudiera tomar solo 2 valores, por ejemplo, 0 o 1 (pesimista u optimista), se necesitaría simular  $2^{161}$  escenarios para obtener todos los resultados posibles, lo cual es cerca de  $3 \cdot 10^{48}$  simulaciones.

Si en vez de 2 valores, cada parámetro pudiera tomar cien o mil valores, las simulaciones necesarias se escapan de una lógica humana.

#### 3.1.2. Escenarios críticos

Una situación actual del modelo, ejemplificada en la Figura 15, es la notable diferencia entre el máximo de las emisiones totales<sup>9</sup> y la suma de los valores máximos de emisiones por sector, en los escenarios explorados, en este caso, para el año 2050. Esta situación indica que el modelo de referencia no representa un escenario donde todos los sectores presenten máximas emisiones al mismo tiempo. Lo mismo ocurre en cuanto al mínimo de emisiones.

---

<sup>9</sup> Emisiones totales: suma de las emisiones de los sectores en un mismo escenario.

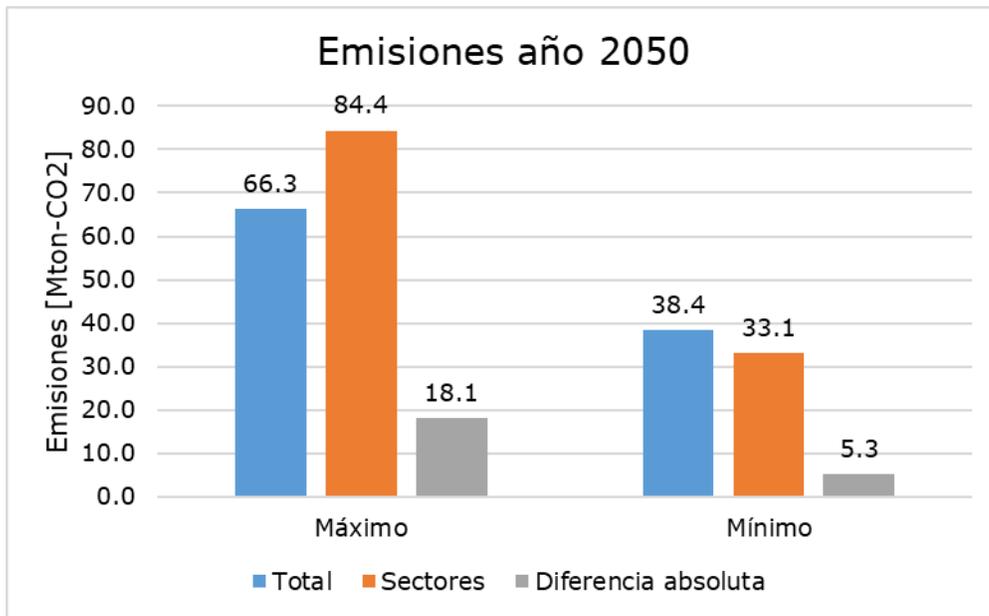


Figura 15: Comparación de máximo y mínimo total de emisiones al año 2050, y suma de los máximos y mínimos de los sectores en distintos escenarios.

Lo anterior se debe a que la combinatoria de muestres es aleatoria de distribución uniforme (por tanto, equiprobable). Luego, es improbable que por azar coincidan solo máximos o solo mínimos, a menos que el número de escenarios tienda a infinito.

Por otra parte, corresponde mencionar que cada sector posee su propio modelo de simulación. Por lo tanto, el resultado de cada sector es mayoritariamente independiente, a excepción del impacto de variables macroeconómicas como población o producto interno bruto (PIB). Luego, cada sector posee trayectorias, niveles de emisiones y costos financieros particulares.

El modelo de generación eléctrica, si bien recibe como dato de entrada la demanda eléctrica proyectada por los sectores industria y minería, comercio, público, residencial y transporte, presenta un modelo independiente en cuanto a sus parámetros y desarrollo de optimización. Es decir, la demanda eléctrica solo presenta una trayectoria definida, por escenario, para el sector generación eléctrica, ya que este se simula de forma posterior a los resultados del resto de los sectores.

### 3.1.3. Formulación del problema

El problema por resolver en este trabajo de memoria es la baja representatividad de escenarios posibles, tanto en la combinación de parámetros, como por la inexistencia de escenarios críticos, donde todos los sectores presenten máximas emisiones al mismo tiempo (igualmente en cuanto a los mínimos de emisiones y en otras métricas de desempeño).

Por lo tanto, se busca un aumento del rango de incertidumbre expresado por las métricas de desempeño tales como emisiones de carbono equivalente, CAPEX y OPEX.

## 3.2. Resultados que se esperan obtener

Del trabajo se espera obtener una metodología replicable y adaptable a distintas entradas, que permita obtener un modelo de proyecciones de amplio rango de incertidumbre. Se espera poder reconocer los distintos supuestos e importancia de las medidas de mitigación en los resultados de los distintos escenarios.

La nueva metodología deberá cumplir con las siguientes características:

- Obtener un rango más amplio de las trayectorias de las distintas métricas de desempeño, tales como CAPEX, OPEX y emisiones GEI. Así también, obtener un rango más amplio del valor de emisiones GEI al año 2050, cuya motivación surge de los compromisos adoptados y meta de carbono neutralidad.
- Identificar parámetros de mayor incidencia en escenarios de resultados máximos o mínimos en cuanto a las métricas de desempeño, mediante método de Análisis de Componentes Principales, a partir de los resultados del modelo de referencia.
- Reducción de escenarios similares, permitiendo la proyección de nuevos escenarios, con enfoque en identificar escenarios de valores máximos o mínimos en alguna de las métricas de desempeño.
- La reducción de escenarios se realiza mediante algoritmo de *clustering k-means*, por lo que es imprescindible comparar la distribución de métricas de desempeño en los distintos escenarios, buscando validar la clasificación y consiguiente selección de escenarios representativos o centroides.

También, se espera validar la metodología propuesta mediante iteraciones de esta, comparando el aumento, o disminución, del rango de incertidumbre por sector y total de sectores.

### 3.3. Metodología propuesta

El objetivo principal es desarrollar una metodología que permita obtener los rangos máximos y mínimos entre los cuales pueden variar las emisiones de carbono equivalente, considerando sectores productivos y energéticos, dentro de rangos determinados de incertidumbre en sus parámetros.

Además, se busca identificar parámetros de mayor incidencia en escenarios de resultados máximos o mínimos en cuanto a las métricas de desempeño, las cuales son: emisiones de carbono equivalente, CAPEX y OPEX en el horizonte de evaluación (2020 a 2050), y emisiones de carbono equivalente en el año 2050.

El enfoque principal de la metodología propuesta consiste en determinar los factores comunes o patrones entre los escenarios máximos y mínimos de las 4 métricas de desempeño consideradas: emisiones acumuladas, CAPEX y OPEX en el periodo 2020-2050, y las emisiones del año 2050. Es decir, se busca identificar las condiciones necesarias y casos en que se obtienen escenarios máximos y mínimos de las métricas mencionadas.

Para verificar el comportamiento de la metodología propuesta, esta se aplica a un mismo caso de estudio utilizado en el modelo de referencia. Este caso de estudio se compone de las medidas de mitigación de la actualización de la NDC 2020 [2], y trayectorias de referencia bajo supuestos actuales de proyecciones que consideran un PIB asociado a la pandemia por COVID-19.

La metodología propuesta parte de la continuación de un modelo de referencia, ya que a partir de los resultados o métricas de desempeño (emisiones GEI, CAPEX y OPEX) y la incertidumbre de sus respectivos escenarios, genera una nueva matriz de incertidumbre (o matriz LHS), y vuelve a simular. En Figura 16 se presenta el diagrama de la metodología propuesta.

Esta continuación permite realizar iteraciones de la metodología propuesta. Si bien los resultados principales se componen de la comparación entre el modelo de referencia y el modelo con metodología propuesta, se realizan 3 iteraciones adicionales, teniéndose los siguientes resultados:

- Original: resultados del modelo de referencia.
- Propuesta: resultados del modelo con metodología propuesta, posterior a la simulación del modelo de referencia.
- 2da iteración: resultados de aplicar la metodología propuesta por segunda vez (es decir, a partir de los resultados de "Propuesta").
- 3ra iteración: resultados de aplicar la metodología propuesta por tercera vez (es decir, a partir de los resultados de "2da iteración").
- 4ta iteración: resultados de aplicar la metodología propuesta por tercera vez (es decir, a partir de los resultados de "3ra iteración").

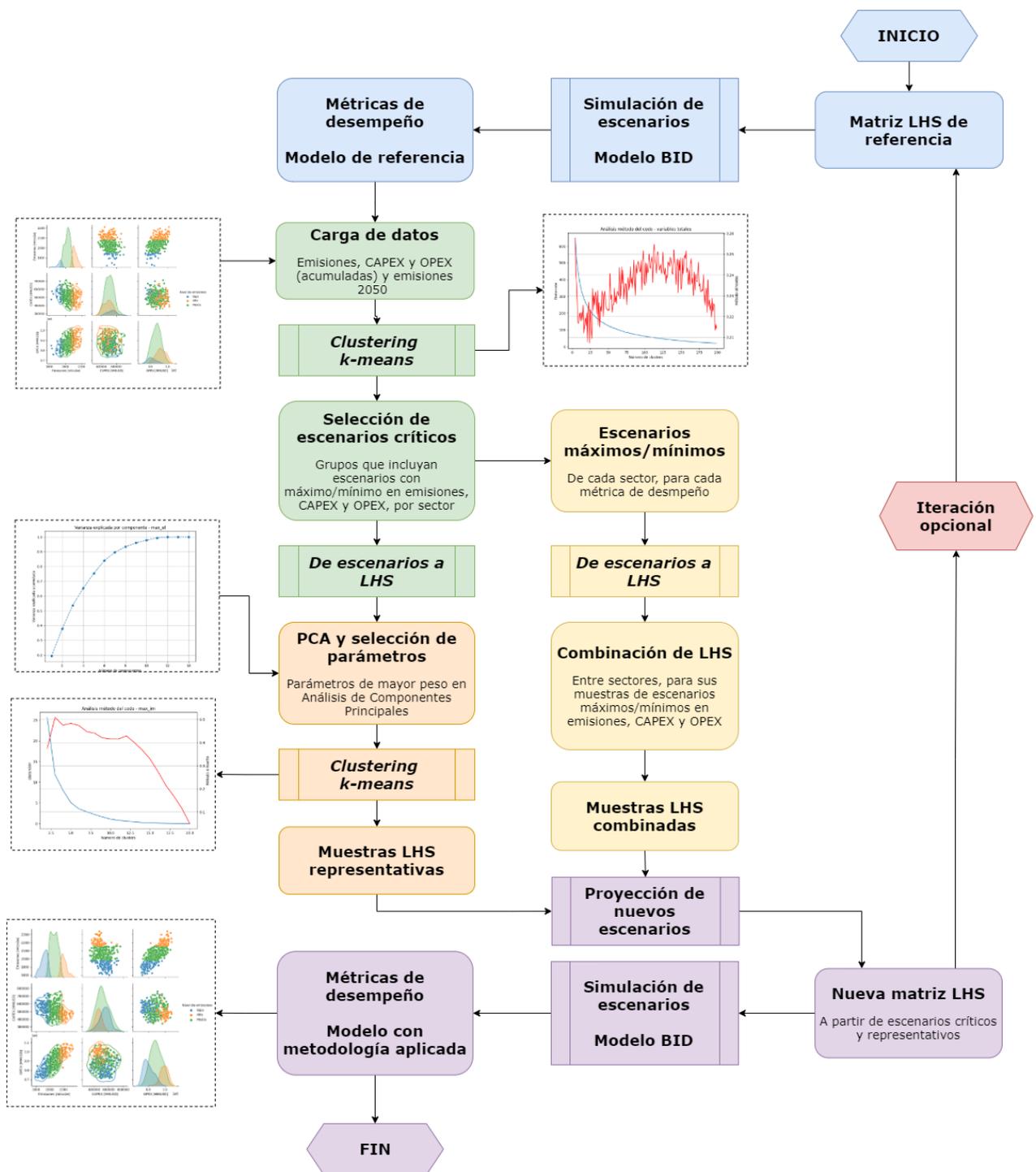


Figura 16: Diagrama de modelo propuesto utilizando metodología de combinación de parámetros, agrupación y clasificación por *k-means*, y proyección de nuevos escenarios mediante muestreo de hipercubo latino.

Del proceso se distinguen 5 etapas principales, las cuales se detallarán presentando su motivación, pasos a seguir y datos de entrada y salida:

1. Simulación de modelo de referencia (azul).
2. Carga y análisis de datos (verde).
3. Agrupación o *clustering k-means* (naranja).
4. Combinación de parámetros (amarillo).
5. Proyección de nuevos escenarios (morado).

### 3.3.1. Etapa 1 - Simulación de modelo de referencia

Previo a la aplicación de la metodología propuesta, es necesario contar con al menos una simulación previa. Esto ya que la metodología propuesta utiliza las salidas y entradas del modelo anterior, para aprender sobre los patrones comunes y valores que predisponen un escenario de valores máximos o mínimos en alguna de las métricas de desempeño.

Para el modelo original o de referencia, es indispensable que su incertidumbre sea generada de forma aleatoria, para poder visualizar fehacientemente los resultados de la metodología propuesta sin sesgos de una aplicación previa. Basta con verificar que el código genere una matriz LHS y no solo la lea.

Para las iteraciones de la metodología propuesta, se parte de una matriz LHS modificada, por lo que esta tenderá a valores extremos (cerca de 0 o a 1), asociados a los escenarios de valores máximos y mínimos de las métricas de desempeño.

### 3.3.2. Etapa 2 - Carga y análisis de datos

El objetivo de esta etapa es obtener los datos del modelo original o de referencia, para luego realizar clasificación y reducción de sus parámetros y escenarios, con la finalidad de identificar las muestras de LHS de los escenarios de valores máximos y mínimos en alguna de las métricas de desempeño.

#### 3.3.2.1. Carga de datos

Los datos obtenidos son:

- Trayectorias de referencia:

Corresponde a los valores de cada parámetro dentro del horizonte de evaluación. Este puede ser constante para todo este horizonte. Un parámetro típicamente constante es, por ejemplo, el rendimiento o la intensidad energética. Los parámetros que varían en el tiempo son los costos de inversión, los precios de los energéticos, la participación de alguna tecnología, etc.

No son datos obtenidos dentro de la metodología propuesta, pero se mencionan para explicar las muestras de LHS y rangos de incertidumbre, información que sí es utilizada en la metodología propuesta y que están sumamente relacionadas.

- Muestras de LHS o incertidumbre del modelo:

Corresponde a la matriz de LHS, donde las filas corresponden a cada uno de los escenarios a evaluar (1000 filas), y las columnas corresponden a cada uno de los parámetros del modelo (862 columnas más columna de *future\_id* o número de escenario explorado).

Sus valores van entre 0 y 1, según la distribución cuasi aleatoria propia del muestreo de hipercubo latino. De esta matriz, se obtiene la variación de las trayectorias de cada parámetro, respecto de su referencia.

La siguiente tabla muestra un ejemplo de lo mencionado anteriormente. Se tienen tres parámetros del sector industria y minería: precio de los combustibles diésel, electricidad e hidrógeno. Se evalúan 1000 escenarios o futuros plausibles.

Para cada futuro estos parámetros toman un valor entre 0 y 1. Luego, el valor que toma el parámetro para cada escenario es igual a  $\text{factor}(t) \times \text{valor de referencia}(t)$ , donde  $t$  corresponde al año y  $\text{factor}(t)$  corresponde a un valor que crece linealmente hasta el año 2050, tal que proyecte la trayectoria a un valor dentro del rango de incertidumbre, según la muestra de LHS asociada (si la muestra es 1, la trayectoria se proyecta al rango máximo de incertidumbre definida al año 2050).

Tabla 3: Matriz de LHS, correspondiente a la incertidumbre aplicada en cada escenario en los parámetros del modelo de referencia.

<b>Id futuro</b>	<b>Industria y minería combustible diésel</b>	<b>Industria y minería combustible electricidad</b>	<b>Industria y minería combustible hidrógeno</b>
<b>1</b>	0.30682456	0.2759417	0.41503782
<b>2</b>	0.5467762	0.29056929	0.42842161
...	...	...	...
<b>999</b>	0.592967174	0.245534008	0.773897747
<b>1000</b>	0.20239768	0.30317494	0.12268431

El modelo BID presenta 862 parámetros, y en esta experiencia se consideran 1000 escenarios proyectados.

- Rangos de incertidumbre:

Corresponde a los factores que determinan las trayectorias a evaluar utilizando las muestras de LHS. Para cada parámetro se define un valor "mínimo 2050" y un valor "máximo 2050". Si ambos son iguales, el parámetro se modela sin incertidumbre (trayectoria constante en todos los escenarios).

Si por ejemplo se tiene: mínimo 2050 = 0.9 y máximo 2050 = 1.1, entonces el valor de la trayectoria en el año 2050 se modela dentro de un rango de  $\pm 10\%$  respecto al valor de la trayectoria de referencia en el mismo año.

Para los valores previos al del año 2050 (es decir, entre los años 2020 y 2049), la desviación respecto de la trayectoria de referencia crece linealmente hasta alcanzar el valor del año 2050.

En Figura 17 se presenta el rango de incertidumbre para los 3 parámetros del ejemplo anterior. Es decir, se tienen 3 trayectorias de referencia, y 6 escenarios proyectados según rangos de incertidumbre (mínimo 2050 y máximo 2050 igual a 0.9 y 1.1 en este caso). La muestra de LHS con valores 0.01 está asociada al valor mínimo del parámetro, mientras que el valor LHS igual 0.99 está asociado al valor máximo.

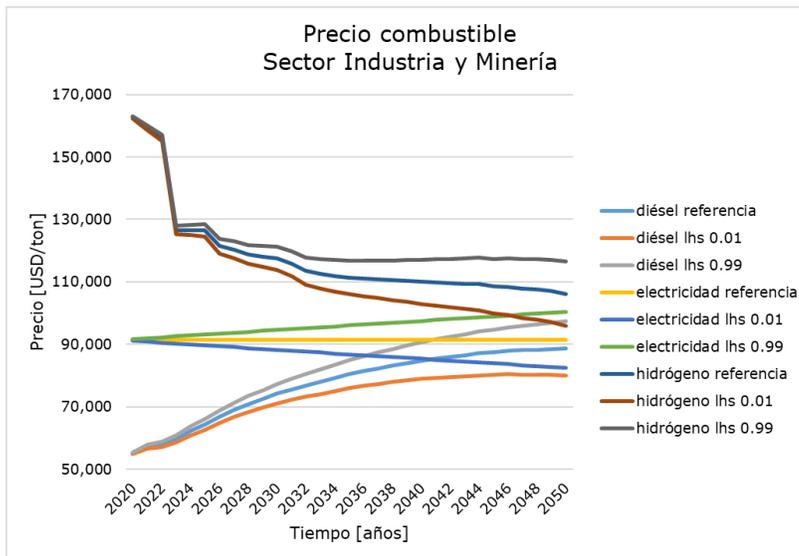


Figura 17: Trayectorias según incertidumbre para precio de combustible diésel, electricidad e hidrógeno, en sector industria y minería, para casos de referencia,  $lhs = 0.01$  y  $lhs = 0.99$ .

Cabe destacar que las trayectorias mencionadas corresponden a distintos escenarios, recordando que es poco probable que 3 parámetros coincidan en sus muestras con un valor 0.01 o 0.99 en un mismo escenario (similar a que un evento que ocurre 1 entre 400 veces, ocurra 3 veces consecutivas).

- Métricas de desempeño:

Se utilizan las salidas del modelo: emisiones GEI, CAPEX y OPEX. Por tanto, la metodología propuesta no considera el impacto macroeconómico como métrica de desempeño (por ejemplo, el impacto en el PIB de cada escenario evaluado).

- Emisiones GEI: se tienen las emisiones anuales de los 6 sectores, y las emisiones anuales de subsectores. Por ejemplo, algunos subsectores del sector transporte son la aviación, camiones y taxis.

La metodología propuesta solo considera las emisiones por sectores (es decir, la suma o total de los subsectores).

Para concentrar el análisis de escenarios en el horizonte de evaluación, se trabaja las emisiones acumuladas en el periodo de evaluación. Es decir, se considera el "área bajo la curva" o la suma de emisiones entre el año 2020 y el año 2050.

Se trabaja de esta forma considerando que el efecto invernadero "tiene memoria", considerando entonces su relación con la temperatura promedio global, cuya meta es la de no superar en más de 2 °C, con respecto a los niveles preindustriales, y proseguir los esfuerzos para limitar este aumento de la temperatura a 1,5 °C [2].

De hecho, la propuesta del NDC de Chile define como meta un presupuesto de emisiones de GEI de hasta 1.100 MtCO<sub>2</sub>eq para el periodo 2020-2030 [4].

- Emisiones GEI 2050: se utilizan los valores de emisiones al año 2050, considerando el objetivo principal de tener un Chile carbono neutral a ese año.
- Costos: corresponden a los costos de inversión en tecnologías (CAPEX), y el costo de operación y mantenimiento anual (OPEX), asociados a las medidas sectoriales.

Análogamente a las emisiones, se considera el total de costos durante el periodo 2020-2050.

Por simplificación, el modelo no considera estas métricas en el sector Público, debido a la baja participación de las emisiones de este sector en el total de las emisiones, y por tanto no incluyen medidas de reducción de emisiones.

En Figura 18 se presenta en la diagonal la distribución mediante histogramas de las 3 métricas obtenidas del periodo 2020-2050 del modelo original, para el total (suma) de los sectores.

En los gráficos de dispersión, se presentan las relaciones CAPEX – emisiones, emisiones – OPEX y CAPEX – OPEX. En el triángulo inferior, además, se presentan curvas de agrupación representando la concentración de los datos por nivel de emisiones.

Los colores representan el nivel de emisiones del escenario, clasificado según el siguiente criterio:

- Nivel bajo:  $emisiones \in \left[ \text{mínimo}, \frac{\text{máximo} - \text{mínimo}}{3} \right[$
- Nivel medio:  $emisiones \in \left[ \frac{\text{máximo} - \text{mínimo}}{3}, 2 \cdot \frac{\text{máximo} - \text{mínimo}}{3} \right[$
- Nivel alto:  $emisiones \in \left[ 2 \cdot \frac{\text{máximo} - \text{mínimo}}{3}, \text{máximo} \right[$

Por tanto, cada nivel no tiene necesariamente la misma cantidad de escenarios.

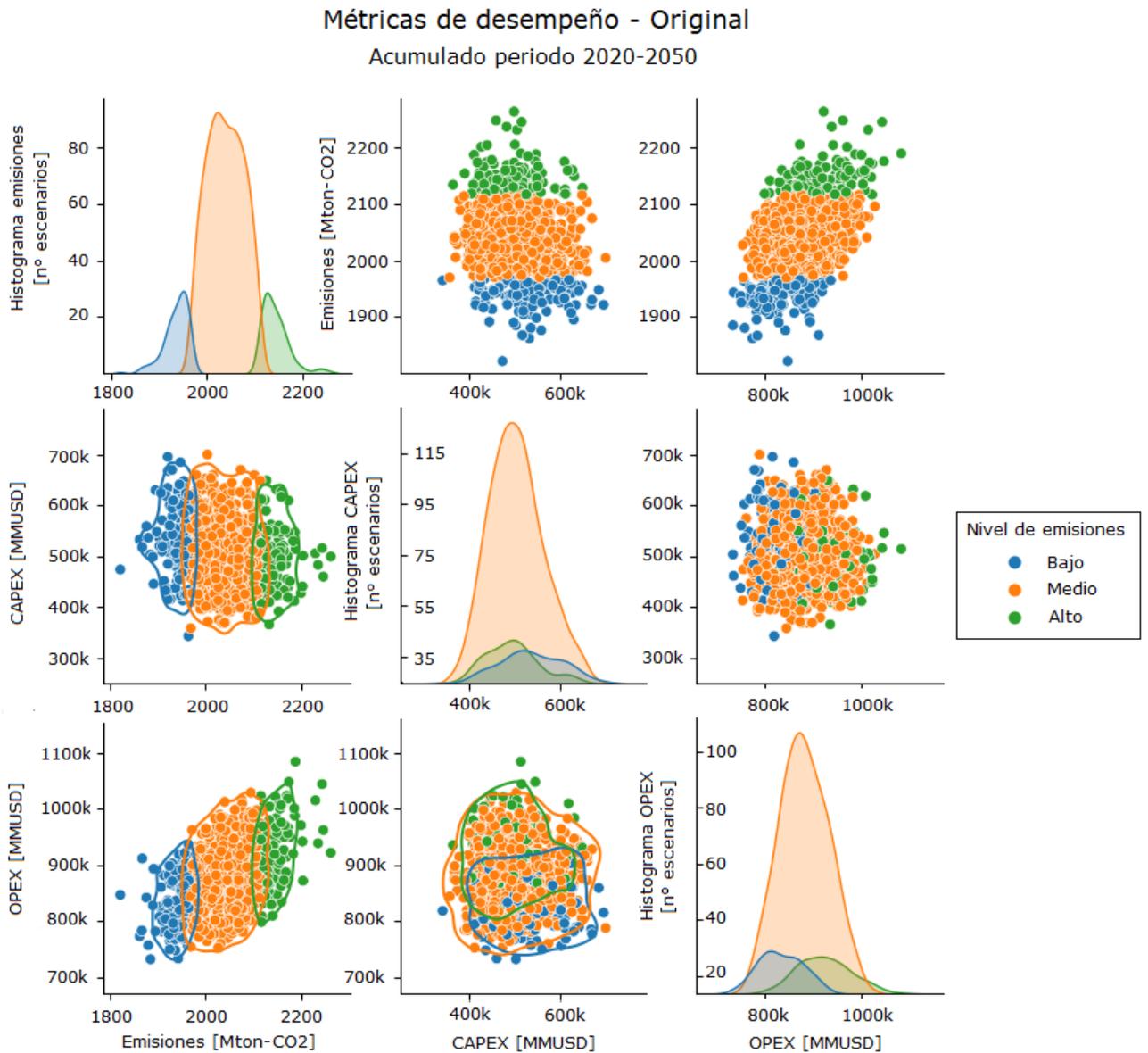


Figura 18: Métricas de desempeño acumuladas (periodo 2020-2050), modelo original, gráficos de dispersión e histogramas (en la diagonal).

Se puede observar una correlación entre emisiones-CAPEX y emisiones OPEX:

- Los escenarios de nivel bajo en emisiones se concentran en escenarios de bajo OPEX, y lo contrario ocurre en escenarios de alto nivel de emisiones. Mientras que los escenarios de nivel medio de emisiones se distribuyen en distintos escenarios de OPEX.
- Los escenarios de alto nivel de emisiones se concentran en escenarios de bajo CAPEX, mientras que los niveles de emisiones medio y bajo se distribuyen en los distintos escenarios de CAPEX.

Además, se tiene una menor cantidad de escenarios en los niveles bajo y alto en emisiones. Este tipo de análisis se realizará en cada modelo o iteración de la metodología propuesta.

Cabe destacar que el solapamiento de los niveles de emisiones en el histograma de emisiones se debe a que se trata de una curva del tipo envolvente de lo que serían gráficos de barras. Es por esto que la curva no pasa directamente a cero entre un nivel y otro.

#### 3.3.2.2. Clustering k-means

Dado lo anterior, se plantea utilizar *clustering k-means* para identificar similitud entre escenarios en cuanto a sus métricas de desempeño. Así también, es posible detectar escenarios cuyas métricas sean poco comunes (por ejemplo, valores máximos o mínimos extremos).

Para seleccionar la cantidad de grupos o clústeres, se utilizan las métricas de validación conocidas como distorsión y método *Silhouette*, recordando de la sección "Marco Teórico y Estado del Arte" lo siguiente:

- Distorsión:

La distorsión está relacionada con la función objetivo del algoritmo de clustering. Esta función objetivo corresponde a la suma de las distancias al cuadrado de cada escenario con el centroide del grupo o clúster asignado. La distorsión corresponde al promedio de estas sumas de distancias.

Es decir, la distorsión indica que tan cerca están los escenarios del promedio o centroide de su respectivo clúster.

Por lo tanto, a mayor cantidad de grupos, menor será el valor de la función objetivo y de la distorsión, llegando a cero en caso de que se tengan tantos grupos como escenarios.

Dado que se busca minimizar la función objetivo mediante la asignación óptima de escenarios a cada grupo o clúster (usualmente, esta asignación se realiza de forma iterativa e inicialmente de forma aleatoria), se busca un bajo valor de distorsión considerando una baja cantidad de grupos, logrando una efectiva agrupación o clasificación.

En términos de reducción de escenarios, conviene asignar una cantidad de grupos cuya función objetivo se encuentre dentro del codo de la curva de distorsión, ya que indica que, de agregar un nuevo grupo, la ganancia será menor que la pérdida de quitar un grupo.

- Método Silhouette:

El método Silhouette representa la comparación entre la distancia de cada escenario con los clústeres más cercanos, y la distancia del escenario con los escenarios de su mismo clúster.

Se busca el valor más cercano a uno, ya que esto significa que la distancia promedio de cada elemento respecto a los otros grupos, es mayor que la distancia del elemento con los de su mismo grupo.

Luego, se utilizan ambas métricas de validación para escoger el número de clústeres. La propuesta consiste en seleccionar el número de clústeres de mayor indicador de método Silhouette, dentro del rango del codo asociado al indicador de distorsión .

Según Figura 19, donde se presentan las métricas de validación de *clustering k-means* para las métricas de desempeño del modelo de referencia, se puede concluir que:

- Por distorsión (curva azul), la cantidad de grupos debería encontrarse entre  $k = 80$  y  $k = 120$ , aproximadamente.
- Por indicador de *Silhouette*, los puntos máximos se concentran entre  $k = 75$  y  $k = 175$ .
- Combinando ambos criterios, el algoritmo de la metodología propuesta escoge  $k = 117$ , ya que es el máximo indicador de *Silhouette* que coincide con la pérdida de ganancia de la métrica de distorsión.

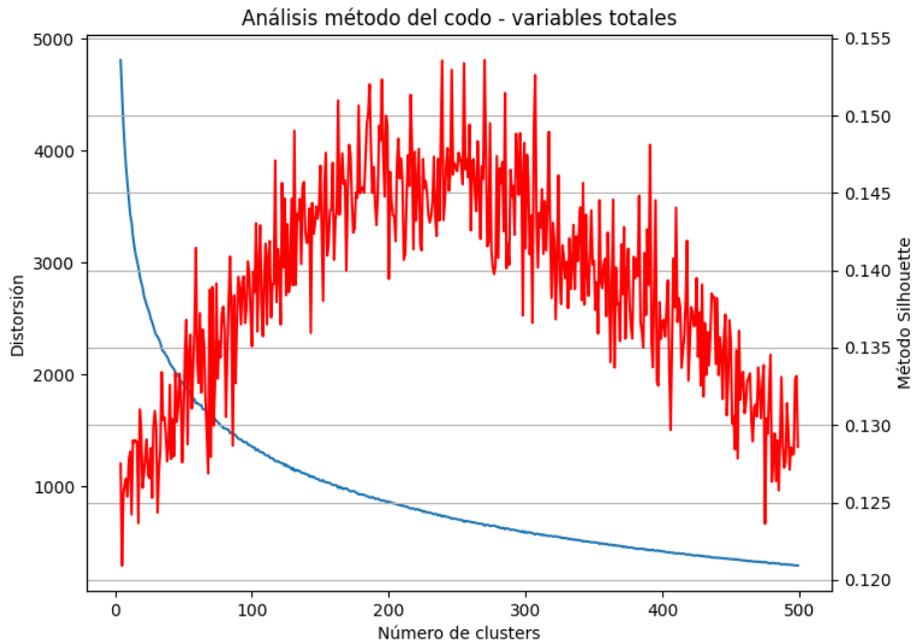


Figura 19: Análisis por método del codo para métricas de desempeño del modelo de referencia, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul).

Recordar que, al ser un método de aprendizaje no supervisado, no hay una solución única y la cantidad de grupos depende del criterio del diseñador, justificado por métricas como las mencionadas.

Con  $k = 117$  (número de clústeres) para 1001 escenarios, se tienen grupos o clústeres de 9 escenarios en promedio, variando entre 1 a 19 escenarios, tal como se presenta en Figura 20.

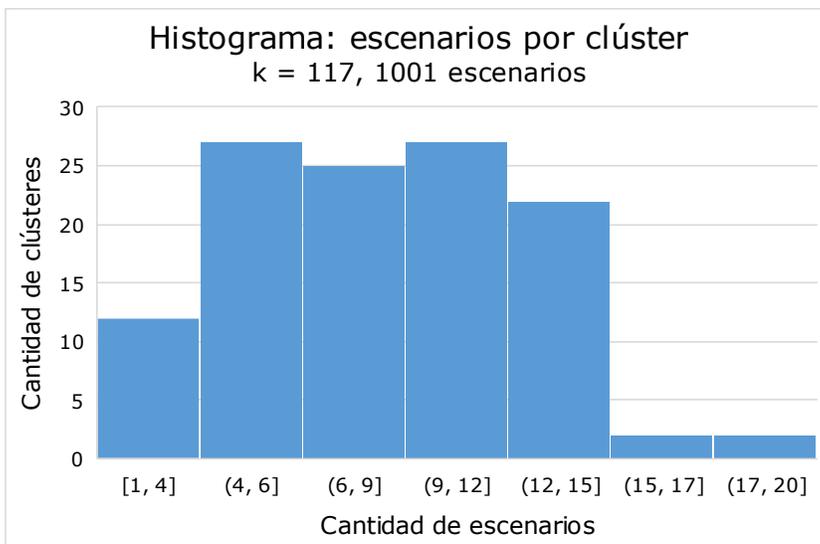


Figura 20: Histograma de escenarios por grupo o clúster, para  $k=117$  (número de clústeres) en 1001 escenarios, para modelo original o de referencia.

### 3.3.2.3. Selección de escenarios

Una vez agrupados los escenarios según similitud de las métricas de desempeño, se procede a seleccionar los escenarios de valores máximos y mínimos por sector para las mismas 4 métricas: emisiones de carbono equivalente, CAPEX y OPEX acumulados en el periodo 2020-2050, y emisiones de carbono equivalente en el año 2050.

El proceso es el siguiente:

- 1) Para cada sector se busca el id del escenario que alcanza el valor máximo para una de las métricas.
- 2) Se busca el grupo o clúster al cual pertenece el id del escenario.
- 3) Se seleccionan todos los id de los escenarios pertenecientes al grupo o clúster.
- 4) Se repite desde 1) para otra métrica de desempeño.
- 5) Finalmente, se obtienen los id de los escenarios donde se presenten escenarios con valores máximos en las métricas de desempeño. Puede haber id de escenarios que presenten valor máximo en más de una métrica de desempeño.

Lo mismo se realiza para los escenarios de valores mínimos en las métricas de desempeño.

De esta forma, se obtienen 2 conjuntos de id de escenarios para cada sector y para el total o suma de sectores: un conjunto para los valores máximos, y un conjunto para los valores mínimos.

Dado que los conjuntos se forman a partir de grupos o clúster, que pueden coincidir entre métricas de desempeño, no necesariamente presentan la misma cantidad de id de escenarios.

### 3.3.2.4. De escenarios a LHS

A partir de los id de escenarios seleccionados anteriormente, se obtienen las muestras de LHS asociadas. Es decir, si el id de un escenario es 100, se obtiene la muestra 100 de la matriz de muestras LHS. Esto, con la finalidad de identificar patrones que propician un escenario de valor máximo o mínimo en los distintos sectores.

Recordar que la matriz de muestras LHS corresponde a la matriz de valores entre 0 y 1, que permite discretizar el rango de incertidumbre de cada uno de los parámetros de forma aleatoria y distribuida uniformemente.

Por ejemplo, si las muestras LHS seleccionadas a partir de los escenarios de valores máximos, para un parámetro, tienden a 1, se podría inferir que si este parámetro toma valores del rango superior de incertidumbre, el escenario tomará un valor máximo en alguna de las métricas de desempeño, aunque no es concluyente por si solo.

De esta etapa, para cada sector y caso (escenarios de valores máximos o mínimos), se pasa de un conjunto de  $n$  id de escenarios, a una matriz de  $n$  filas y 863 columnas (862 parámetros más columna asociada al id de la fila).

Antes de continuar con las etapas de agrupación o combinación, se realizan dos acciones claves de la metodología propuesta.

#### 3.3.2.5. Selección de parámetros con incertidumbre.

De los 862 parámetros del modelo de referencia, a solo 161 se le definió un rango de incertidumbre.

Además, los parámetros generalmente están asociados para modelar un sector en específico, siendo unos pocos parámetros transversales a todos los sectores, como lo son la población y el PIB.

Dado lo anterior, para cada matriz mencionada anteriormente (una por cada sector y caso), se seleccionan los parámetros propios del sector con rango de incertidumbre definida distinta de cero.

Por ejemplo, el sector comercio tiene definido solo 3 parámetros con incertidumbre: elasticidad de comercio, PIB y población. Luego, las matrices del sector comercio tendrán 3 columnas asociadas a los parámetros.

Los parámetros con rango de incertidumbre definida mayor a cero de cada sector se presentan en sección Anexos.

### 3.3.3. Etapa 3 - Agrupación o *clustering k-means*

A partir de las matrices de LHS para cada sector y caso (escenarios de valores máximos o mínimos en alguna de las métricas de desempeño), se busca reducir las filas o escenarios con la finalidad de obtener un subconjunto representativo. Para esto, se utiliza al algoritmo de *clustering k-means*.

Para obtener los escenarios más representativos, se toma una muestra de parámetros mediante el estudio de componentes principales, ya que la cantidad de parámetros de algunos sectores dificulta la aplicación del algoritmo *k-means*: se tienen más parámetros que escenarios, por lo que el algoritmo se inclina a tener la misma cantidad de grupos que de escenarios, sin aplicar ninguna reducción de escenarios.

#### 3.3.3.1. PCA o Análisis de Componentes Principales

Según lo visto en sección 2.3.62.3.6. Análisis de componentes principales, es posible representar la información contenida en una matriz de datos mediante sus componentes principales.

Cada componente principal representa un vector propio de la matriz representada, es decir, es un vector capaz de describir cada escenario mediante la multiplicación del vector con un escalar o valor propio.

La cantidad de componentes principales a utilizar se puede definir mediante el criterio de varianza explicada acumulada. En Figura 21, se presenta un ejemplo del análisis del modelo de referencia, donde se presenta la varianza explicada acumulada de los escenarios de valores máximos en métricas de desempeño, para el sector residencial.

Se puede observar que, de un máximo de 15 componentes principales, con 6 componentes principales es posible explicar un 80% de la varianza acumulada. O lo que es igual, existen 9 direcciones o componentes principales que en total explican un 20% de la varianza acumulada.

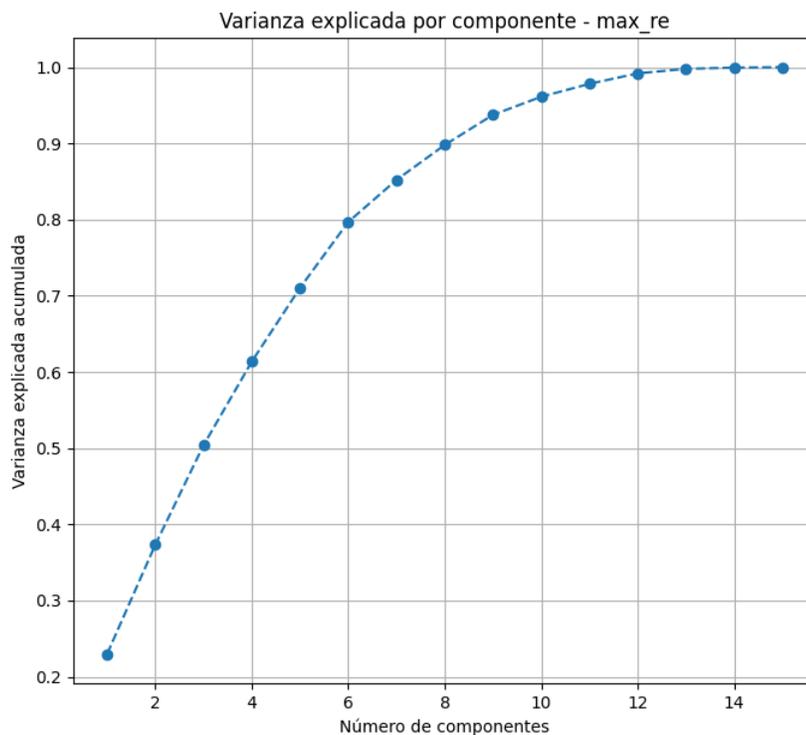


Figura 21: Varianza explicada acumulada de los escenarios de valores máximos del sector residencial.

En la propuesta metodológica, se selecciona la cantidad de componentes principales desde donde la curva disminuye su crecimiento o pendiente. Es decir, se consideran las componentes principales que mejor expliquen la varianza acumulada.

Luego de seleccionar la cantidad de componentes principales consideradas, se procede a reducir la matriz de muestras LHS de cada sector y caso. Para esto, se

utiliza el análisis de peso o importancia de cada parámetro en las distintas componentes principales.

Recordando de la sección 2.3.6.3 Peso de cada variable, la primera componente principal del grupo de parámetros  $[X_1, X_2, \dots, X_p]$  es la combinación lineal normalizada de dichos parámetros:

$$Z_1 = \varphi_{11}X_1 + \varphi_{21}X_2 + \dots + \varphi_{p1}X_p$$

La combinación lineal es normalizada si:

$$\sum_{j=1}^p \varphi_{j1}^2 = 1$$

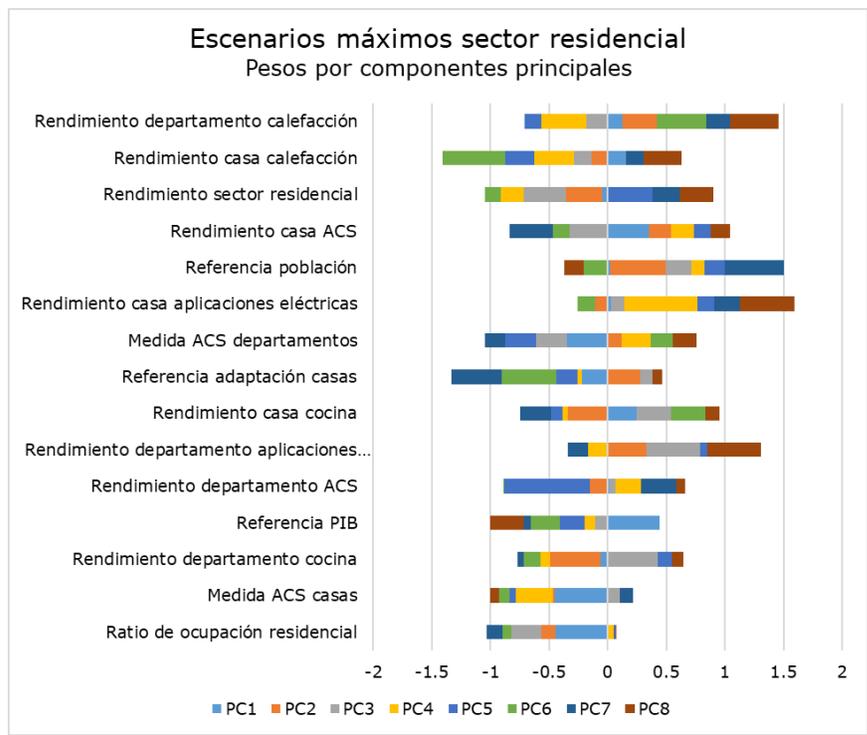
El cálculo de las siguientes componentes principales cumple las mismas condiciones, sumado a que deben ser ortogonales a las componentes calculadas previamente.

Los términos  $\varphi_{11}, \dots, \varphi_{p1}$  se denominan *loadings* y definen las componentes. Por ejemplo,  $\varphi_{11}$  es el *loading* de la variable  $X_1$  de la primera componente principal. Los *loadings* pueden interpretarse como el peso o importancia que tiene cada parámetro en cada componente y, por lo tanto, ayudan a conocer qué tipo de información recoge cada una de las componentes.

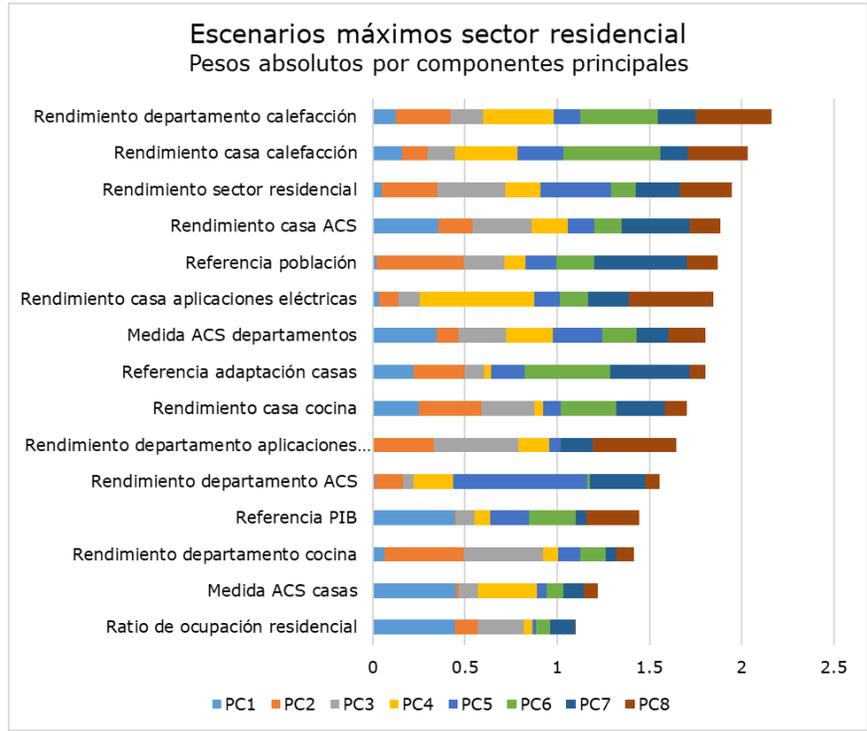
En Figura 22(a) se presenta como ejemplo los pesos o importancia de cada parámetro para cada componente principal, del sector residencial. Para una mejor comparación, se consideran los pesos absolutos (sin signos), como los presentes en Figura 22(b).

Se tiene que los parámetros de menor peso en el total de componentes principales son los que presentan una tendencia constante dentro de la matriz de datos. Mientras que los parámetros de mayor peso presentan mayor variabilidad.

Luego, dado que el conjunto de escenarios corresponde al de valores máximos en las métricas de desempeños, para el sector residencial, se tiene que los parámetros de tendencia constante son los parámetros que definen escenarios de valores máximos del sector (por ejemplo, que mantengan en todo el conjunto valores cercanos a 0), mientras que los parámetros de mayor variabilidad, pueden ser explorados con la finalidad de identificar qué valores deben presentar para obtener escenarios de valores máximos (esto ya que no se concentran en un valor en específico, sino que distribuyen en todo el rango entre 0 y 1, recordando que pertenecen al muestreo de LHS).



(a)



(b)

Figura 22: (a) Peso o importancia de cada parámetro en las componentes principales, y (b) pesos absolutos (magnitud), para escenarios de valores máximos del sector residencial.

Luego, al aplicar el algoritmo de *clustering k-means*, es conveniente clasificar los parámetros de mayor variabilidad, ya que los parámetros de tendencia constante no presentarán mayor diferencia entre un centroide y otro.

La metodología propuesta considera una cantidad de parámetros tal que permita obtener curvas pronunciadas de métricas de validación del algoritmo *clustering k-means*. Es decir, que se pueda distinguir una pérdida de ganancia en términos de distorsión, y picos definidos en la curva de método Silhouette.

Siguiendo el mismo ejemplo del sector residencial, se consideran los parámetros cuyos pesos absolutos acumulados y ponderados por la varianza explicada de sus componentes principales (es decir, dando mayor prioridad a los primeros componentes principales) están sobre el percentil 60, siendo estos:

- Rendimiento de la calefacción en departamentos.
- Rendimiento del agua caliente sanitaria (ACS) en casas.
- Rendimiento de cocina en casa.
- Rendimiento de calefacción en casa.
- Rendimiento general del sector residencial.
- Y cumplimiento de la medida asociada al ACS en departamentos.

Luego, aplicando el algoritmo de *clustering k-means* considerando solo estos 6 parámetros de la matriz, se obtienen las curvas de métricas de validación presentes en Figura 23. Al considerar ambas curvas, se tiene que, de 18 escenarios, es posible considerar 9 escenarios representativos, ya que en  $k=9$  se tiene el máximo indicador de método Silhouette, y se encuentra dentro de la disminución de ganancia en la curva de distorsión.

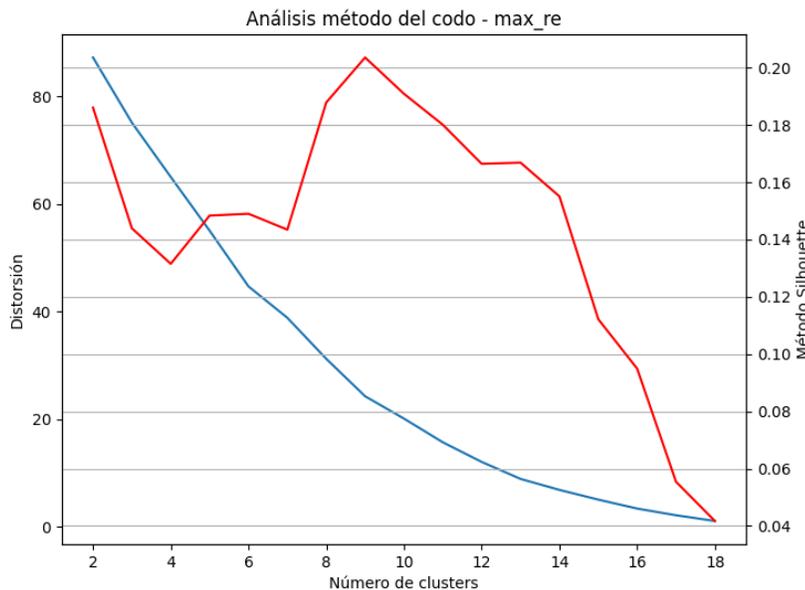


Figura 23: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos del sector residencial, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul).

Finalmente, de esta etapa se obtienen los escenarios representativos para cada sector y caso de valores máximos o mínimos, los cuales serán posteriormente proyectados.

### 3.3.4. Etapa 4 - Combinación de parámetros

Este subproceso, nacido de la idea de obtener el máximo de emisiones totales de forma simultánea al máximo de emisiones de cada sector (e igualmente con los mínimos), considera el acoplamiento de los valores de parámetros correspondientes a cada sector, cuyo escenario corresponda a un valor máximo (o mínimo) de cada una de las métricas de desempeño consideradas.

Por ejemplo, los valores de la muestra LHS de los parámetros del sector comercio del escenario de máximas emisiones para este sector, se combinan con los valores de la muestra de LGS de los parámetros del sector industria y minería del escenario de máximas emisiones para este sector, y así sucesivamente. Lo mismo ocurre para el caso con mínimas emisiones, y el resto de las métricas de desempeño.

Con esto se obtienen 8 muestras, 4 para los valores máximos (una muestra por métrica de desempeño) y 4 para los valores mínimos, y servirán de insumo para la expansión de escenarios presente en el siguiente proceso.

### 3.3.5. Etapa 5 – Proyección de nuevos escenarios

Para esta etapa, es necesario el cálculo de desviación estándar para cada parámetro, en sus respectivas matrices de valores máximos o mínimos, obtenidos de la etapa de carga de datos.

#### 3.3.5.1. Cálculo de desviación estándar.

La desviación estándar es un cálculo estadístico que permite cuantificar la similitud entre distintos valores, dado por la siguiente fórmula:

$$\sigma = \sqrt{\frac{\sum(\bar{x} - x_i)^2}{N}}$$

Donde  $\bar{x}$  corresponde al promedio de los datos, y  $N$  a la cantidad de datos. Es decir, en esencia representa la distancia promedio de los datos respecto a su valor medio.

Dado que la metodología consiste en la exploración de escenarios a partir de escenarios representativos, los parámetros de estos nuevos escenarios variarán tanto en sus muestras de LHS como desviación estándar hayan tenido, siguiendo la siguiente lógica:

- Los parámetros de baja desviación estándar en el conjunto de escenarios críticos ya lograron un valor crítico y por tanto se busca que lo mantengan.

- Los parámetros de alta desviación estándar en el conjunto de escenarios críticos deben explorar nuevos valores a fin de descubrir un valor que predisponga un escenario crítico.

Para justificar lo anterior, se presenta en

Tabla 4 los parámetros del sector industria y minería de menor desviación estándar, en el caso de 50 escenarios de valores máximos, y 50 escenarios de valores mínimos, ambos en cuanto a emisiones acumuladas en el periodo 2020-2050<sup>10</sup>.

Tabla 4: Parámetros de menor desviación estándar en grupo de escenarios de valores máximos y mínimos de emisiones acumuladas, sector industria y minería.

<b>Caso</b>	<b>Parámetro</b>	<b>Promedio</b>	<b>Desviación Estándar</b>
<b>Mínimo de emisiones</b>	Medida 2 – electrificación mina abierta.	0.84	0.13
	Medida 3 – reemplazo fuente de calor otras industrias.	0.86	0.14
	Rendimiento energía útil cobre.	0.21	0.20
<b>Máximos de emisiones</b>	Rendimiento energía útil cobre.	0.86	0.13
	Medida 2 – electrificación mina abierta.	0.22	0.16
	Motor hidrógeno cobre.	0.55	0.22

Se puede observar que el parámetro de medida de mitigación 2 – electrificación mina abierta y parámetro rendimiento de energía útil del cobre, presentan promedios opuestos de sus muestras de LHS.

Al tener baja desviación estándar, presentan valores similares a sus promedios en todos los escenarios del conjunto. Por tanto, se puede concluir que son parámetros representativos de su condición de valores máximos o mínimos de emisiones en el sector.

### 3.3.5.2. Proyección de escenarios representativos

Finalmente, a partir de cada escenario representativo compuesto por vectores de 862 parámetros cada uno, obtenidos de Etapa 4 - Combinación de parámetros y Etapa 5 – Proyección de nuevos escenarios, se proyectan nuevos escenarios, con tal de lograr la misma cantidad de escenarios del modelo de referencia, es decir, 1000 escenarios.

La proyección de cada escenario se realiza mediante la técnica presente en sección 2.2.1. Proyección de escenarios utilizando LHS. Es decir, el valor de cada

---

<sup>10</sup> Este caso de 50 escenarios no es parte de la propuesta, se utiliza como demostración.

parámetro en cada escenario representativo representa el promedio de la distribución, y se aplica la desviación estándar del parámetro dentro de su matriz por sector y caso de valores máximos o mínimos.

En

Tabla 5, se presenta un ejemplo del diseño en cuanto a la cantidad de escenarios proyectados por sector y por caso crítico. Los valores buscan mantener una distribución equitativa, pero considerando la cantidad de escenarios representativos que posee cada sector. Es por esto que sectores como comercio y público presentan menos proyecciones, mientras que industria y minería, y transporte, presentan más.

Tabla 5: Distribución de escenarios proyectados en nueva matriz LHS.

<b>Sector o caso crítico</b>	<b>Proyecciones por escenario representativo o crítico</b>	<b>Distribución de la matriz LHS</b>
<b>Total</b>	6	196
<b>Comercio</b>	4	70
<b>Industria y minería</b>	8	154
<b>Público</b>	4	45
<b>Residencial</b>	6	112
<b>Transporte</b>	8	182
<b>Generación eléctrica</b>	6	133
<b>Emisiones 2020-2050</b>	19 y 22 (*)	30
<b>Emisiones 2050</b>	19	26
<b>CAPEX</b>	19	16
<b>OPEX</b>	19	16
<b>Total</b>		<b>1000</b>

(\*) Se tiene una variable de holgura, a fin de lograr los 1000 escenarios.

Cabe destacar que, si un parámetro sobrepasa los límites del muestreo LHS, es decir, supera el rango de  $[0, 1]$ , este se fija en 0 o en 1, según corresponda. De esta forma, la nueva matriz de incertidumbre no presenta la misma distribución que la matriz original, y por tanto se presentan más de un valor cercano a cero o a uno en un mismo parámetro.

Lo anterior es esencial para lograr el aumento del rango de incertidumbre en las métricas de desempeño, ya que la combinación de muestras estará orientada a los extremos del rango de incertidumbre de cada parámetro.

# Capítulo 4

## Resultados

A continuación, se presentan los resultados de la aplicación de la metodología propuesta. Para comparar la metodología propuesta con la metodología utilizada en estudio previo, la evaluación se realiza considerando el escenario denominado "supuestos actuales" (supuestos NDC actualizados del NDC), y las medidas de mitigación del NDC (sin incluir las medidas complementarias que fueron incorporadas en el escenario NDC+).

Se utiliza el caso de estudio mencionado ya que de este se obtuvieron las métricas de desempeño como entradas a la metodología propuesta. No obstante, con la misma matriz LHS generada es posible modelar los otros casos de estudio.

Además, se presenta la iteración de la metodología propuesta, por tanto, se tienen los siguientes resultados:

- Modelo original: corresponde a la ejecución del modelo sin la metodología propuesta, y, por tanto, genera una matriz LHS de forma tradicional (cuasi aleatoria).
- Modelo propuesto: corresponde a la ejecución del modelo con la metodología propuesta, es decir, obtiene los datos del modelo original, aplica la metodología propuesta y vuelve a ejecutar el modelo con una nueva matriz LHS generada.
- 2da iteración: aplicación de la metodología a partir de los resultados de modelo propuesto.
- 3era iteración: aplicación de la metodología a partir de los resultados de 2da iteración.
- 4ta iteración: aplicación de la metodología a partir de los resultados de 3ra iteración.

Primero, se presentan los resultados propios de la aplicación de la metodología propuesta, con enfoque en la comparación entre el modelo original, propuesto e iteraciones.

Y finalmente, se presentan los resultados por sector con enfoque en sus componentes principales y primera aplicación de la metodología propuesta (es decir, sin considerar las iteraciones).

## 4.1. Resultados por modelos

En esta sección, se presentan los resultados con la finalidad de comparar entre el modelo de referencia (llamado original en los resultados), y el modelo propuesto o modelo de referencia con metodología propuesta aplicada.

Además, se presentan tres iteraciones de la metodología propuesta, donde las entradas corresponden a los datos del modelo previamente simulado (ej.: 2da iteración toma de referencia los resultados del modelo propuesto).

Los resultados de las métricas de desempeño acumuladas (CAPEX, OPEX y emisiones en el periodo 2020-2050) se presentan con gráficos de dispersión e histogramas, gráficos de caja y gráficos con las trayectorias de cada escenario.

### 4.1.1. Histograma y dispersión

A continuación, se grafican las métricas de desempeño acumuladas del periodo 2020-2050 mediante gráficos mixtos de dispersión e histogramas (ubicados en la diagonal), buscando las relaciones entre emisiones, CAPEX y OPEX en cada uno de los 1001 escenarios evaluados (1000 proyectados más escenario de referencia).

En los gráficos de dispersión, se presentan las relaciones CAPEX – emisiones, emisiones – OPEX y CAPEX – OPEX. En el triángulo inferior, además, se presentan curvas de agrupación representando la concentración de los datos por nivel de emisiones.

Los sub gráficos presentes en la diagonal corresponden a histogramas de la variable respecto a los niveles de emisiones.

Los colores representan el nivel de emisiones del escenario, clasificado según el siguiente criterio:

- Nivel bajo:  $emisiones \in \left[ \text{mínimo}, \frac{\text{máximo} - \text{mínimo}}{3} \right[$
- Nivel medio:  $emisiones \in \left[ \frac{\text{máximo} - \text{mínimo}}{3}, 2 \cdot \frac{\text{máximo} - \text{mínimo}}{3} \right[$
- Nivel alto:  $emisiones \in \left[ 2 \cdot \frac{\text{máximo} - \text{mínimo}}{3}, \text{máximo} \right[$

Es decir, se divide el rango de emisiones en 3 partes iguales en magnitud, y se clasifican los escenarios según el intervalo al que pertenece. De la definición anterior, cada nivel no incluye, necesariamente, la misma cantidad de escenarios.

Los gráficos que se presentan a continuación se realizaron utilizando la librería *seaborn*, del lenguaje python. Específicamente, corresponde a función *pairplot*.<sup>11</sup> Y fueron modificados en formato para agregar información detallada en cuanto a los ejes y valores presentes.

---

<sup>11</sup> Más información en <https://seaborn.pydata.org/generated/seaborn.pairplot.html>

#### 4.1.1.1. Modelo original

En Figura 24 se presentan las métricas de desempeño del modelo original, correspondiente a las variables totales (o suma de los sectores) respecto a emisiones, CAPEX y OPEX, para el periodo comprendido entre el año 2020 a 2050.

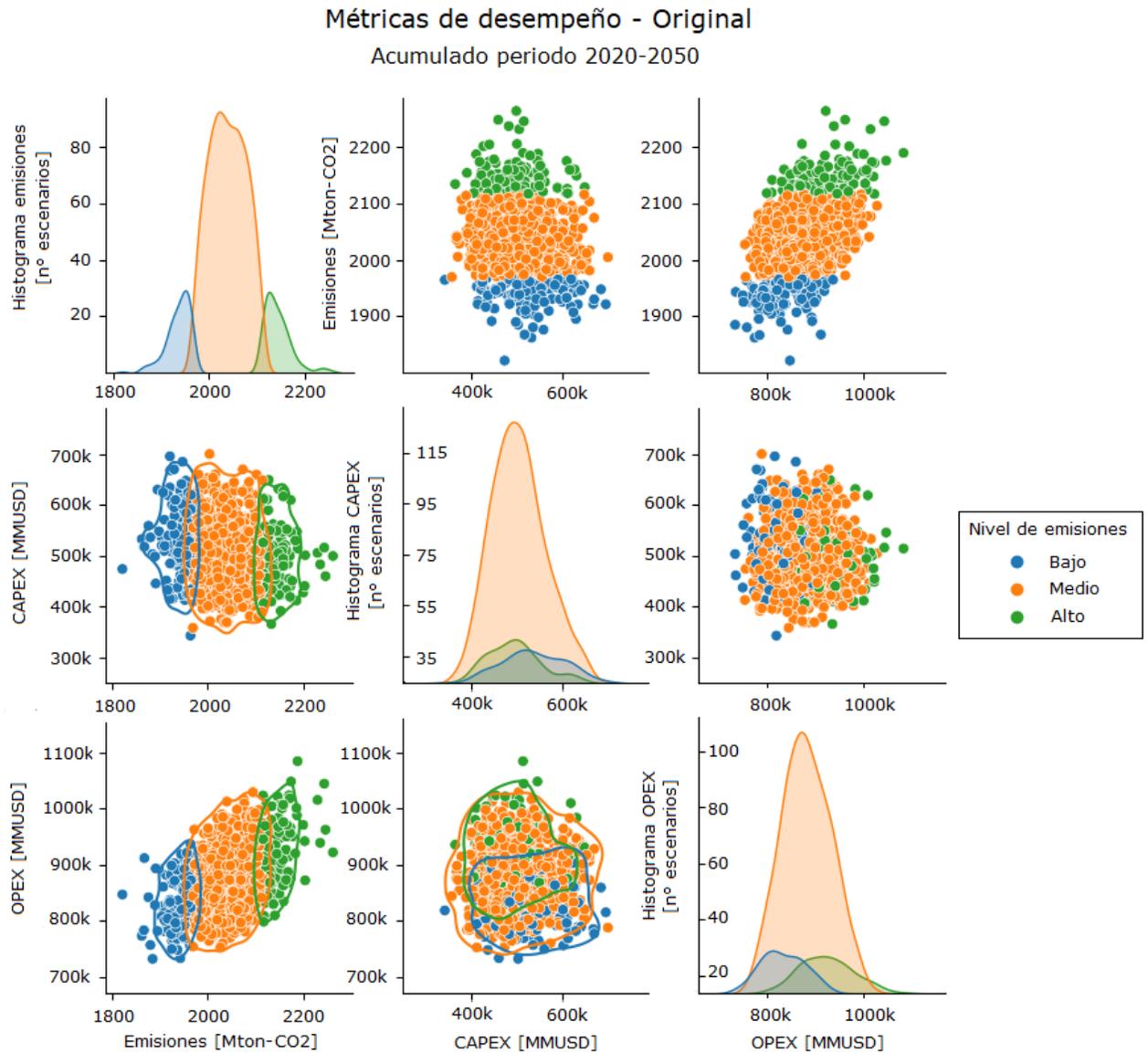


Figura 24: Métricas de desempeño totales o suma de sectores, obtenidas de modelo original, periodo 2020-2050.

Respecto a los histogramas de escenarios, para las emisiones acumuladas en el periodo, se puede observar que la mayoría de los escenarios se encuentran en un nivel medio de emisiones (color naranja). Los niveles bajo y alto presentan una cantidad similar de escenarios.

Respecto a histograma del indicador CAPEX, o sub gráfico de en medio, se tiene que las emisiones altas se concentran en los escenarios de menor CAPEX, mientras que niveles bajos y medio son transversales a todos los escenarios de CAPEX.

Por otra parte, se tiene que a mayor OPEX, mayor nivel de emisiones presentan los escenarios, esto ya que el nivel alto (verde) se encuentra principalmente en escenarios con mayor indicador OPEX. Caso contrario ocurre a un menor valor del indicador OPEX, donde se tienen niveles bajos de emisiones. En cuanto a las emisiones medias, estas se encuentran en todos los valores del indicador OPEX.

Esta relación OPEX-emisiones se puede observar mediante una pendiente de los datos en el gráfico de dispersión ubicado en la parte inferior-izquierda. Incluso, se tiene que el escenario de mayor valor de OPEX, está dentro de los mayores valores de emisiones acumuladas. Caso análogo se tiene con el escenario de menor valor de OPEX, cuyo escenario se encuentra dentro de los escenarios de menores valores de emisiones acumuladas.

#### 4.1.1.1. Modelo propuesto

En Figura 25 se presentan las métricas de desempeño del modelo propuesto.

Se puede observar que aumenta la cantidad de escenarios en los extremos de emisiones acumuladas (niveles de emisiones bajos y altos), y se intensifican las relaciones mencionadas anteriormente entre emisiones-CAPEX y emisiones-OPEX. Estas son:

- Los niveles altos de emisiones se encuentran en escenarios de bajo CAPEX. Niveles medios y bajos de emisiones se presentan en distintos niveles de CAPEX.
- Los niveles altos de emisiones se encuentran en escenarios de alto OPEX. Por el contrario, los niveles bajos de emisiones se encuentran en escenarios de bajo OPEX.

Una nueva observación, es que en el modelo propuesto se distribuyen escenarios donde en modelo original existían vacíos o pocos escenarios. Es decir, se tiene una mayor dispersión en el universo de escenarios.

### Métricas de desempeño - Propuesta

Acumulado periodo 2020-2050

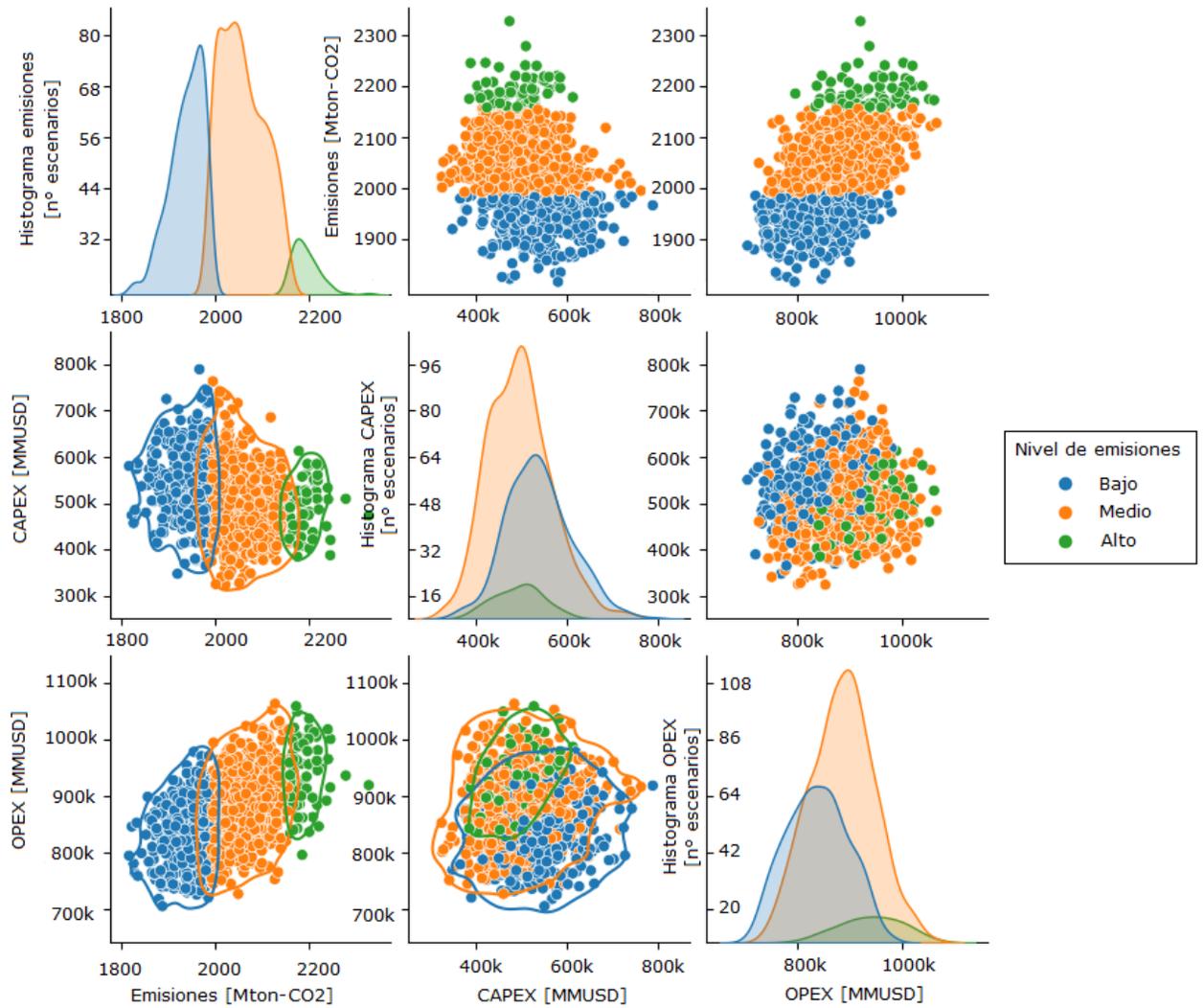


Figura 25: Métricas de desempeño totales o suma de sectores, obtenidas de modelo propuesto, periodo 2020-2050.

#### 4.1.1.2. Segunda iteración

En Figura 26, se presentan las métricas de desempeño del modelo propuesto iterado 2 veces.

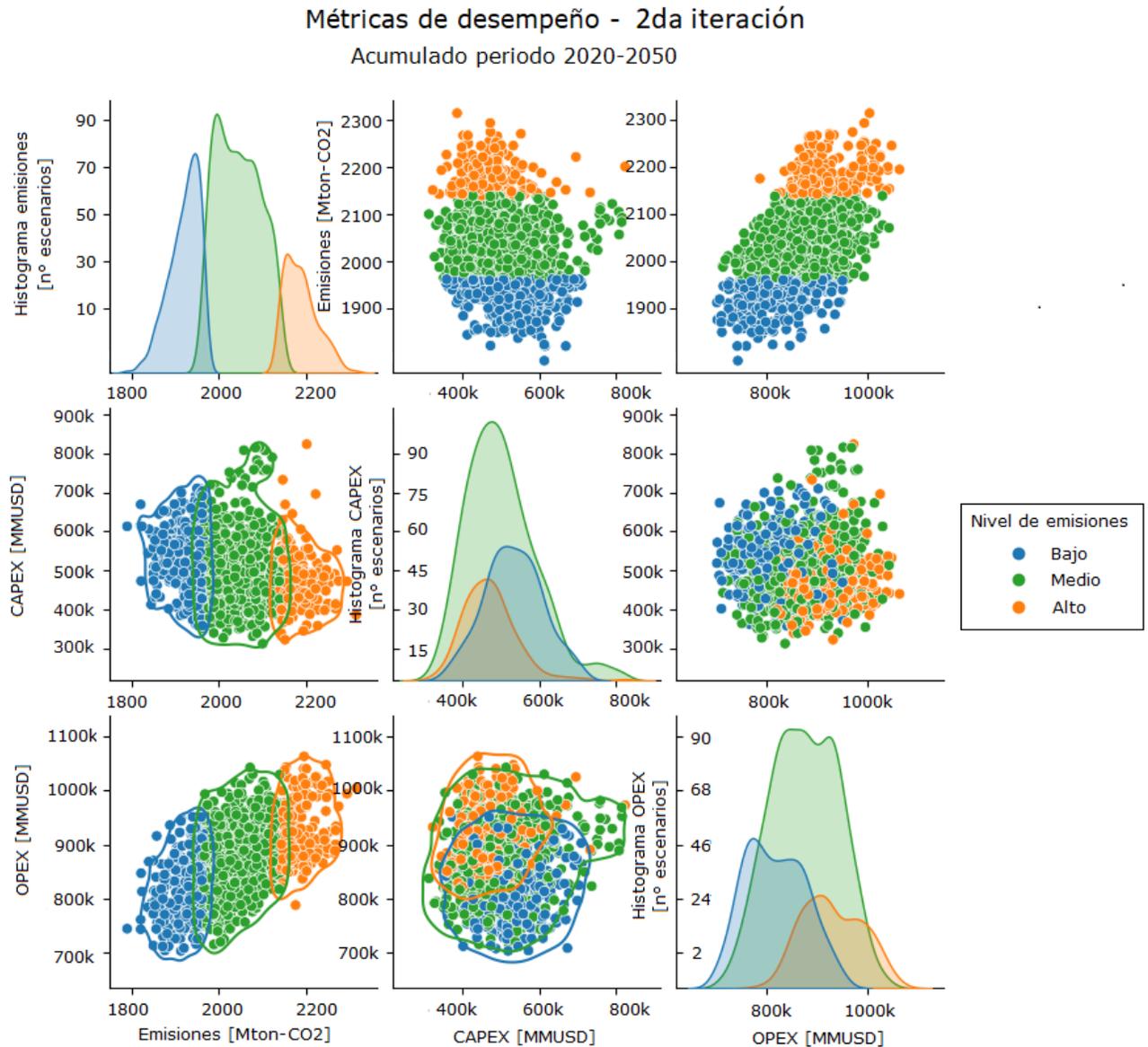


Figura 26: Métricas de desempeño totales o suma de sectores, obtenidas de 2da iteración, periodo 2020-2050.

Se puede observar que se intensifican las observaciones anteriores, en cuanto a las relaciones entre emisiones-CAPEX y emisiones OPEX.

Además, se tiene una cantidad similar de escenarios en los 3 niveles de emisiones, representado por los histogramas en la diagonal del conjunto de gráficos.

Cabe destacar que la función *pairplot* asigna los colores en función de la distribución de elementos, por lo que en este gráfico se tiene nivel de emisiones medio en naranja y nivel de emisiones alto en verde, contrario a los gráficos anteriores y posteriores.

#### 4.1.1.3. Tercera iteración

En Figura 27, se presentan las métricas de desempeño del modelo propuesto iterado 3 veces.

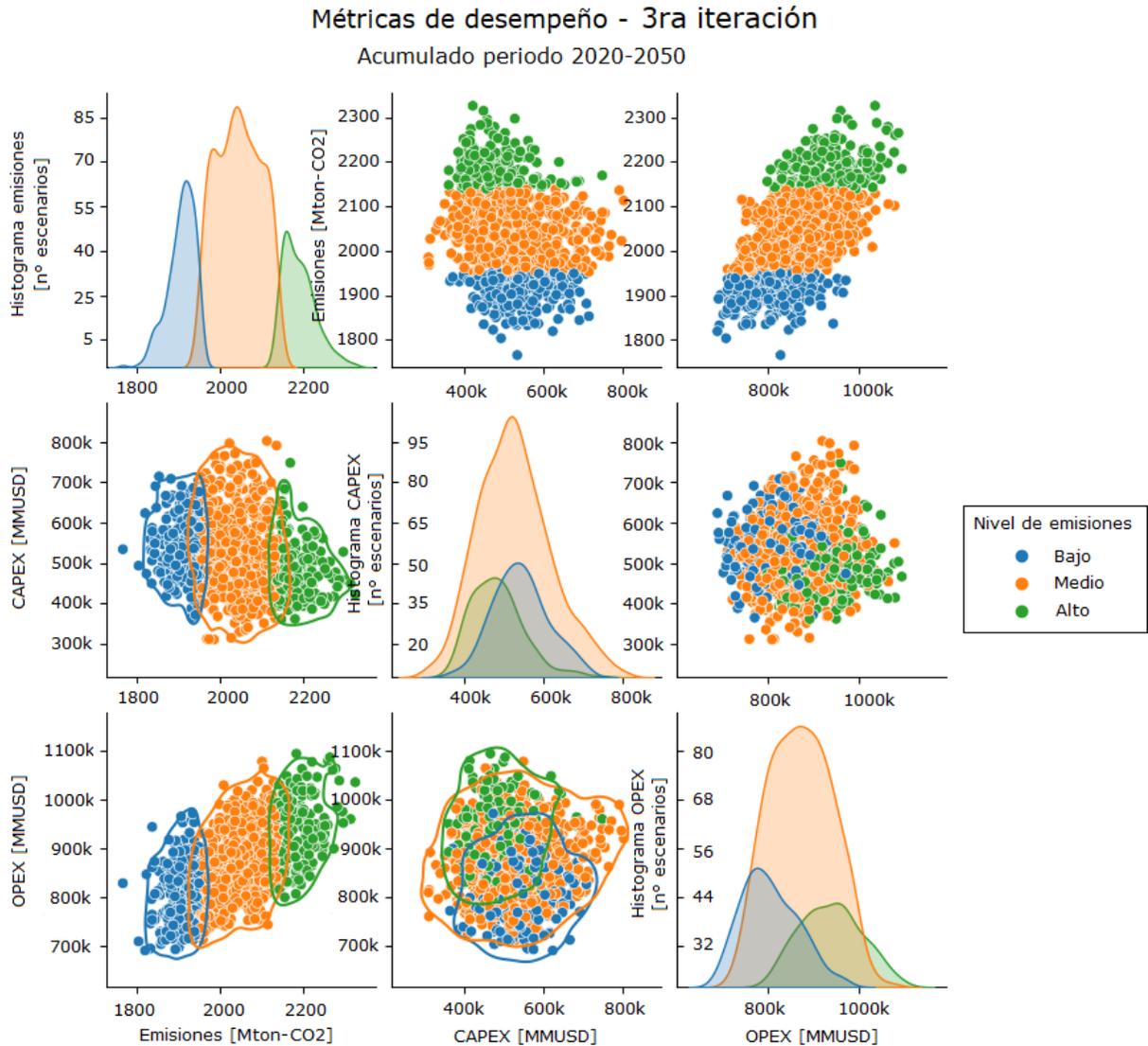


Figura 27: Métricas de desempeño totales o suma de sectores, obtenidas de 3ra iteración, periodo 2020-2050.

La observación más relevante es el aumento de escenarios con nivel bajo en emisiones, con mayor cantidad de escenarios que los niveles de emisiones medio y alto.

Cabe destacar que el nivel de emisiones bajo era el segmento de menor cantidad de escenarios en modelo original.

#### 4.1.1.4. Cuarta iteración

En Figura 28, se presentan las métricas de desempeño del modelo propuesto iterado 4 veces. Se puede observar una disminución de la cantidad de escenarios de emisiones acumuladas en su valor medio, a su vez que aumentan los escenarios de emisiones altas y bajas, comparado con las iteraciones anteriores.

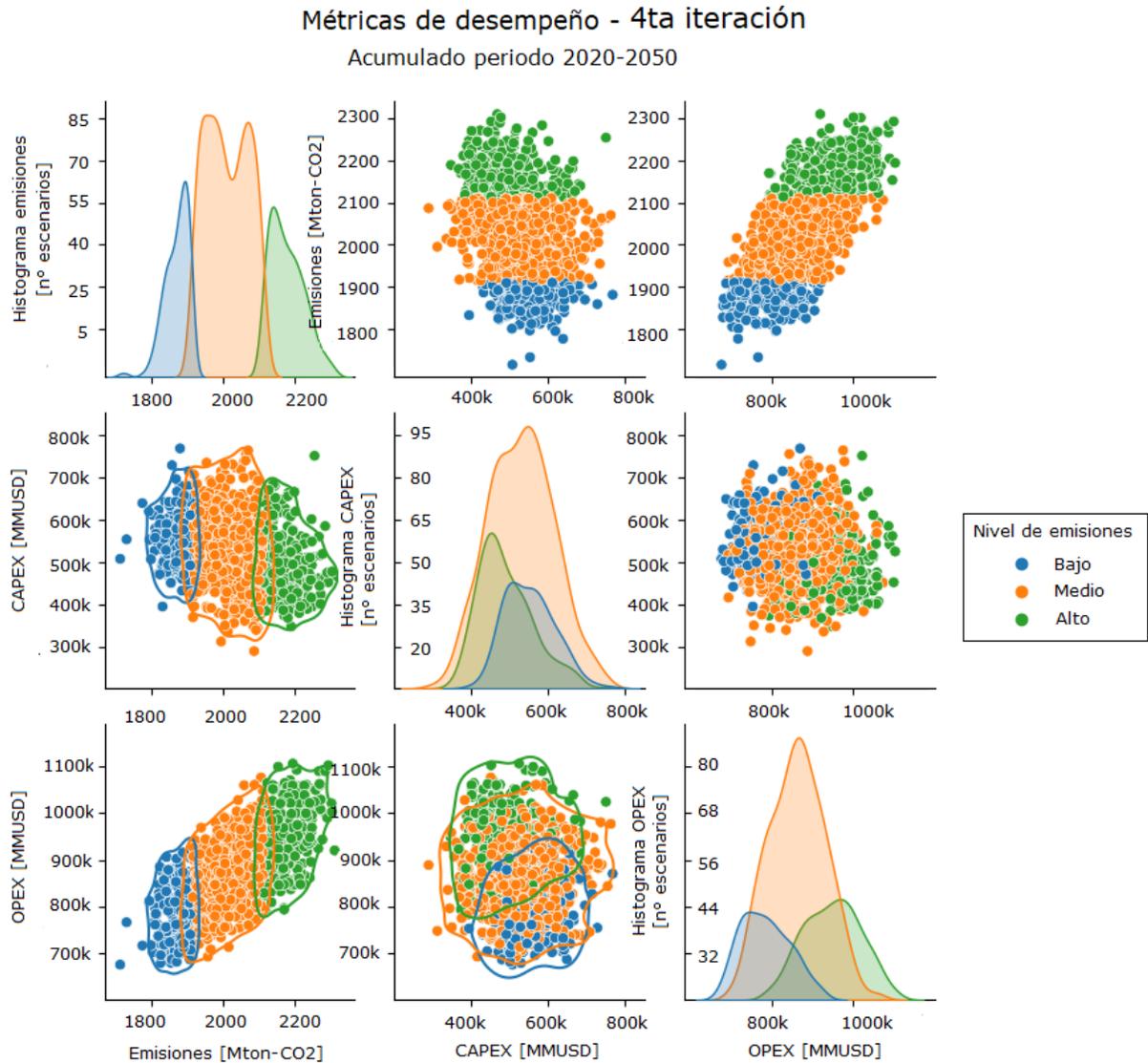


Figura 28: Métricas de desempeño totales o suma de sectores, obtenidas de 4ta iteración, periodo 2020-2050.

### 4.1.2. Dispersión y rangos de incertidumbre

Siguiendo el objetivo de comparar modelos original, propuesto y sus iteraciones, a continuación, se presentan gráficos de caja de los 1001 escenarios evaluados.

Un gráfico de caja entrega información respecto a los cuartiles de un conjunto de datos. Cada cuartil tiene la misma cantidad de escenarios, y se grafican por las 5 líneas horizontales dentro de cada figura.

De lo anterior se desprende que la línea dentro de cada caja es la mediana de los datos. Además, se tiene el promedio representado por una X, y datos *outliers* o datos de valor atípico, representados por círculos fuera de cada figura.<sup>12</sup>

En Figura 29 se presenta el gráfico de caja para las emisiones totales del periodo 2020-2050, comparando los modelos original, propuesto e iteraciones de la propuesta.

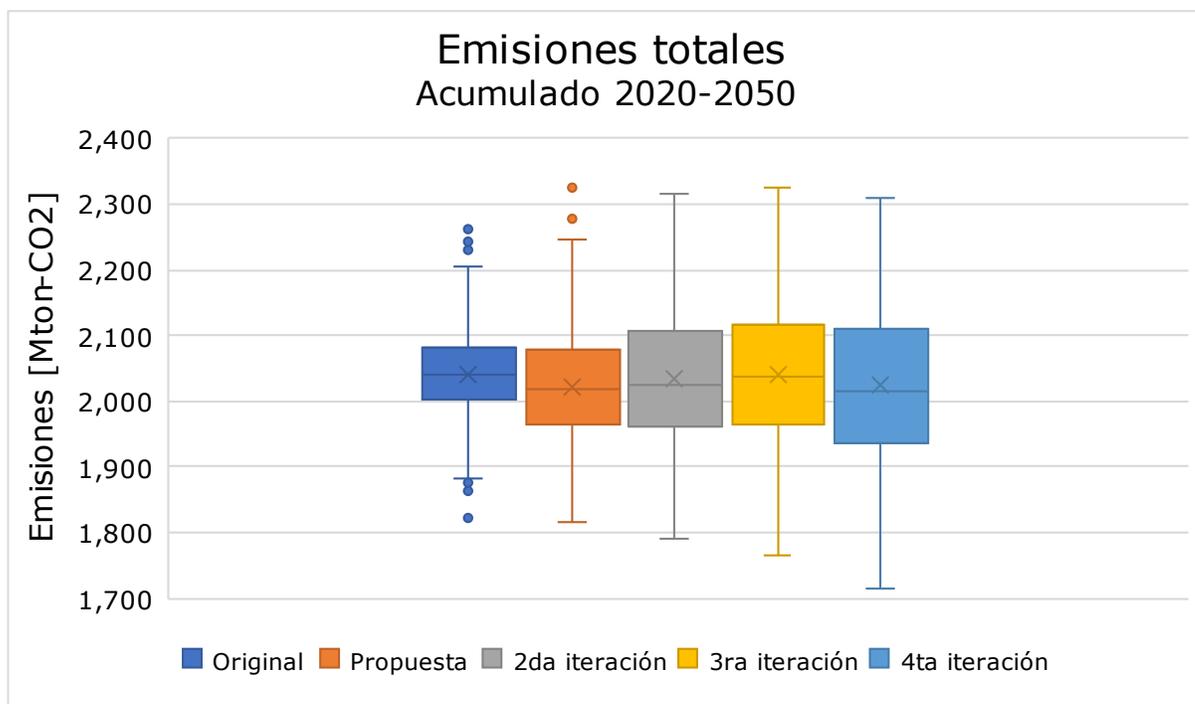


Figura 29: Gráfico de caja de emisiones totales o suma de sectores, periodo 2020-2050.

Se puede observar un aumento progresivo del rango de emisiones a medida que se itera el modelo propuesto, manteniendo un promedio similar entre un modelo y otro. Este aumento de rango va de la mano con un incremento de concentración de valores en los extremos, desapareciendo los escenarios de valores atípicos u *outliers*.

---

<sup>12</sup> Un dato se considera de valor atípico cuando supera la diferencia entre el tercer cuartil y el primer cuartil, por 1.5. Criterio conocido como Test de Tukey.

Se tiene además que aumenta el rango de los cuartiles medios respecto del modelo original. Es decir, con la metodología propuesta se aumentan la distancia entre escenarios de emisiones acumuladas de valor cercano a su promedio.

Comportamiento similar se puede observar en Figura 30, correspondiente a las emisiones totales en el año 2050.

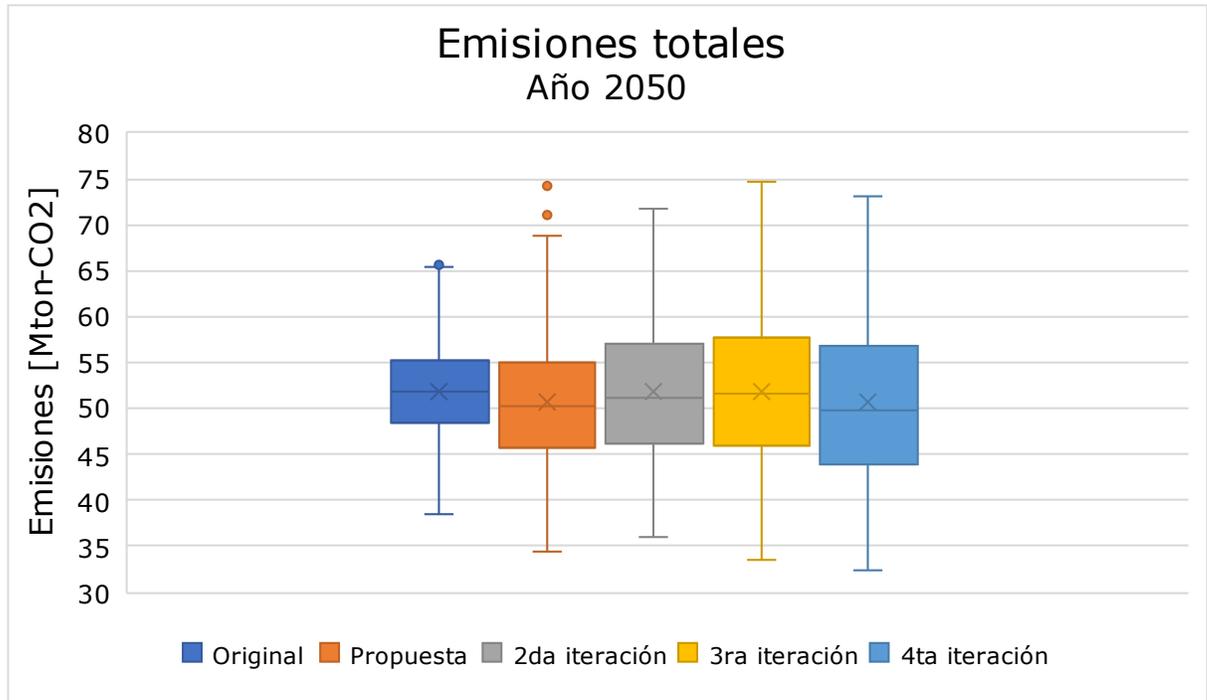


Figura 30: Gráfico de caja de emisiones totales o suma de sectores, año 2050.

En cuanto a costos financieros, presentes en Figura 31 y Figura 32, estos presentan un menor valor mínimo y máximo a medida que se itera la metodología propuesta, disminuyendo también la presencia de valores atípicos.

Nuevamente destacar un promedio similar entre modelos, y un aumento progresivo de los rangos de cada cuartil.

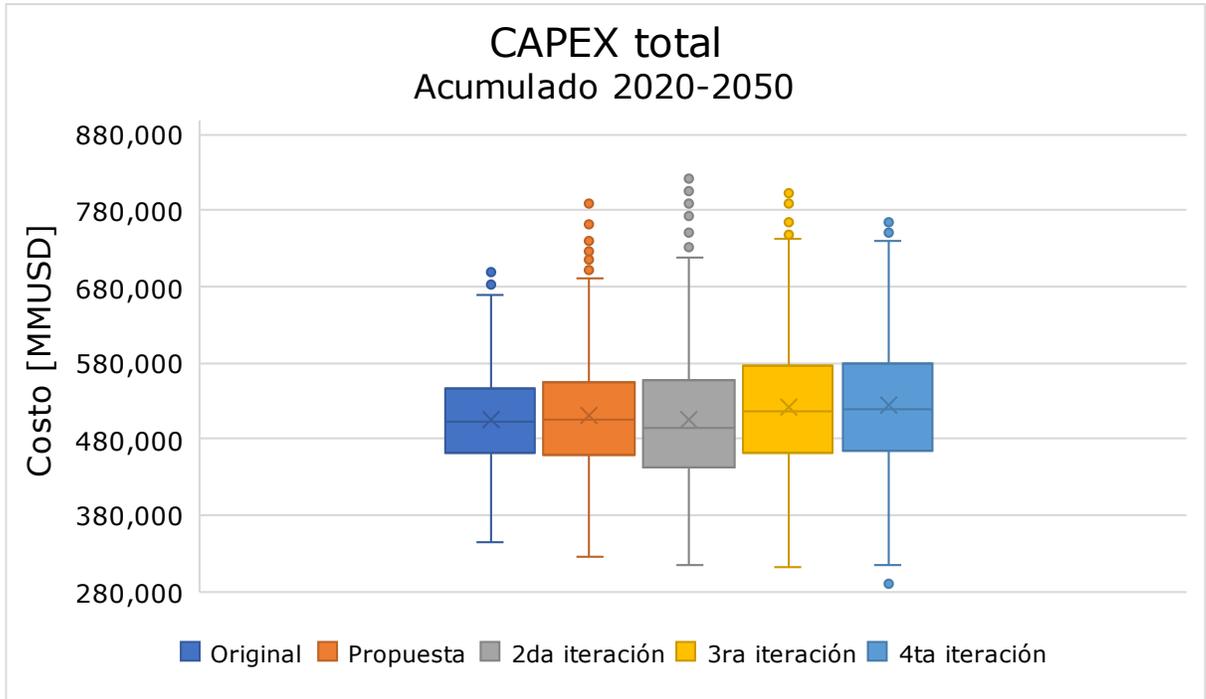


Figura 31: Gráfico de caja de CAPEX total o suma de sectores, periodo 2020-2050.

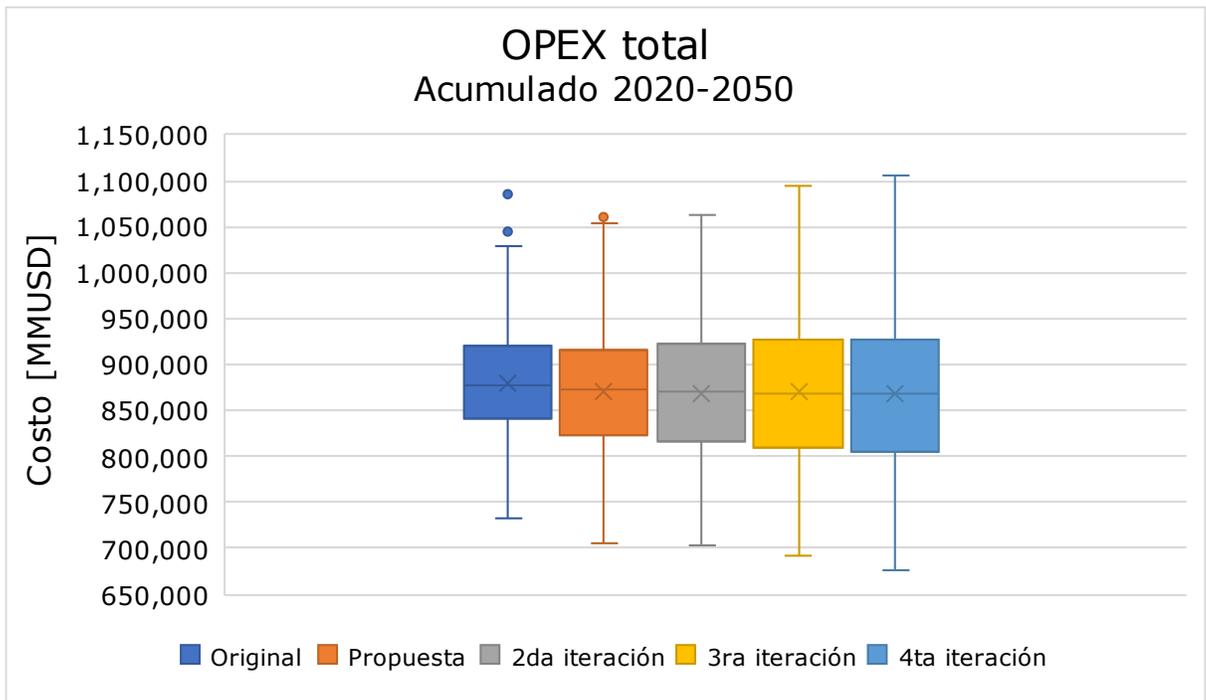


Figura 32: Gráfico de caja de OPEX total o suma de sectores, periodo 2020-2050.

### 4.1.3. Trayectorias de los escenarios

A modo de visualizar los escenarios obtenidos de la aplicación de la metodología propuesta, se tienen las trayectorias de las métricas de desempeño en el periodo comprendido entre los años 2020 y 2050.

Cabe destacar que en color azul se presentan los escenarios de referencia, y en color celeste los escenarios proyectados. Además, las escalas son las mismas entre cada grupo de gráficos.

#### 4.1.3.1. Emisiones GEI totales

En Figura 33, se presentan las emisiones totales o suma de sectores, en el periodo de años 2020-2050.

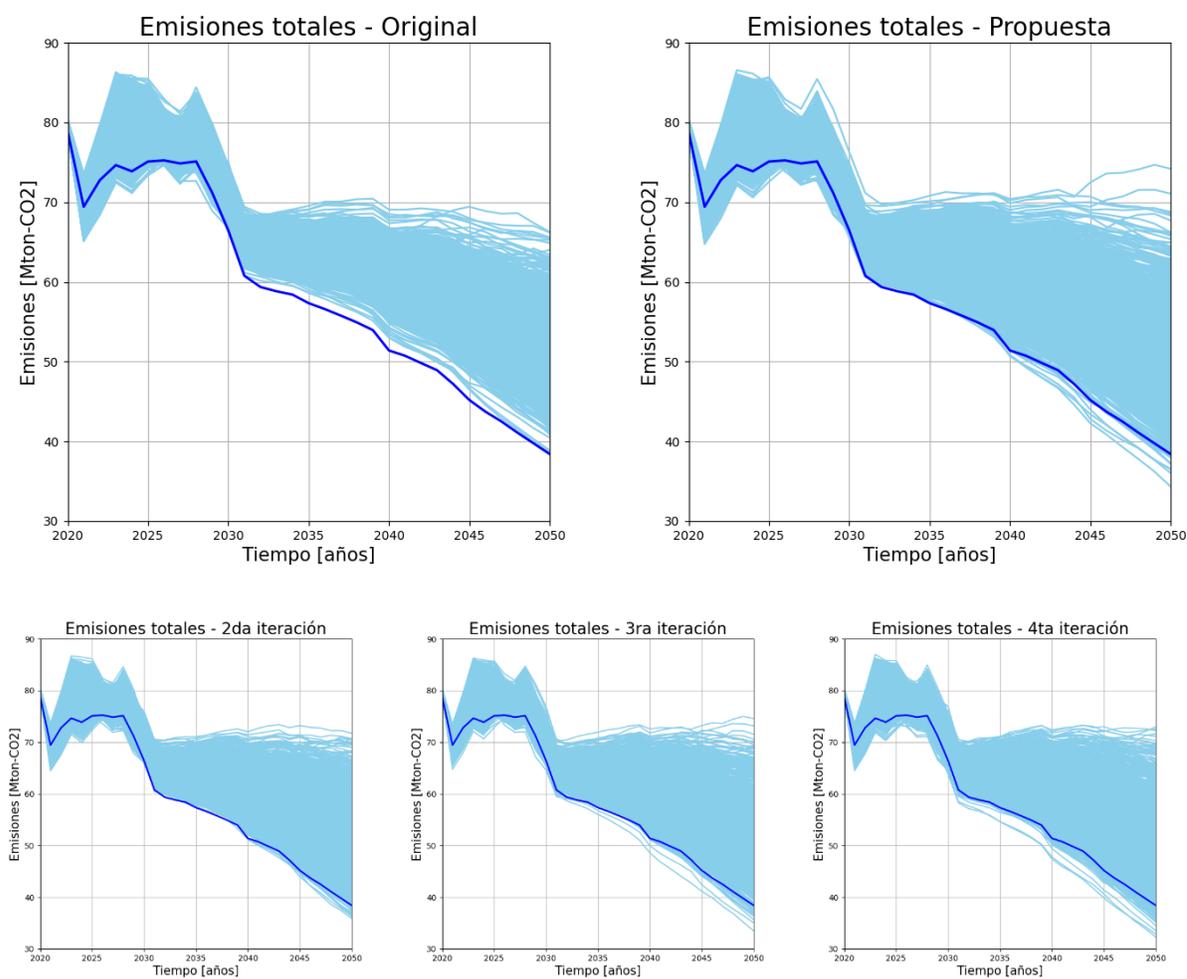


Figura 33: Trayectorias de emisiones totales o suma de sectores, para modelos original, propuesto e iteraciones. Periodo 2020-2050.

Se puede observar que en modelo original no se presentan trayectorias completamente bajo el escenario de referencia (solo existen tramos menores en el periodo 2020-2030). En cambio, al aplicar la metodología propuesta, se incrementa el rango de emisiones durante el periodo y en el año 2050, consiguiendo además presentar trayectorias bajo el escenario de referencia.

#### 4.1.3.2. CAPEX total

En Figura 34 se presentan las trayectorias de CAPEX total o suma de sectores.

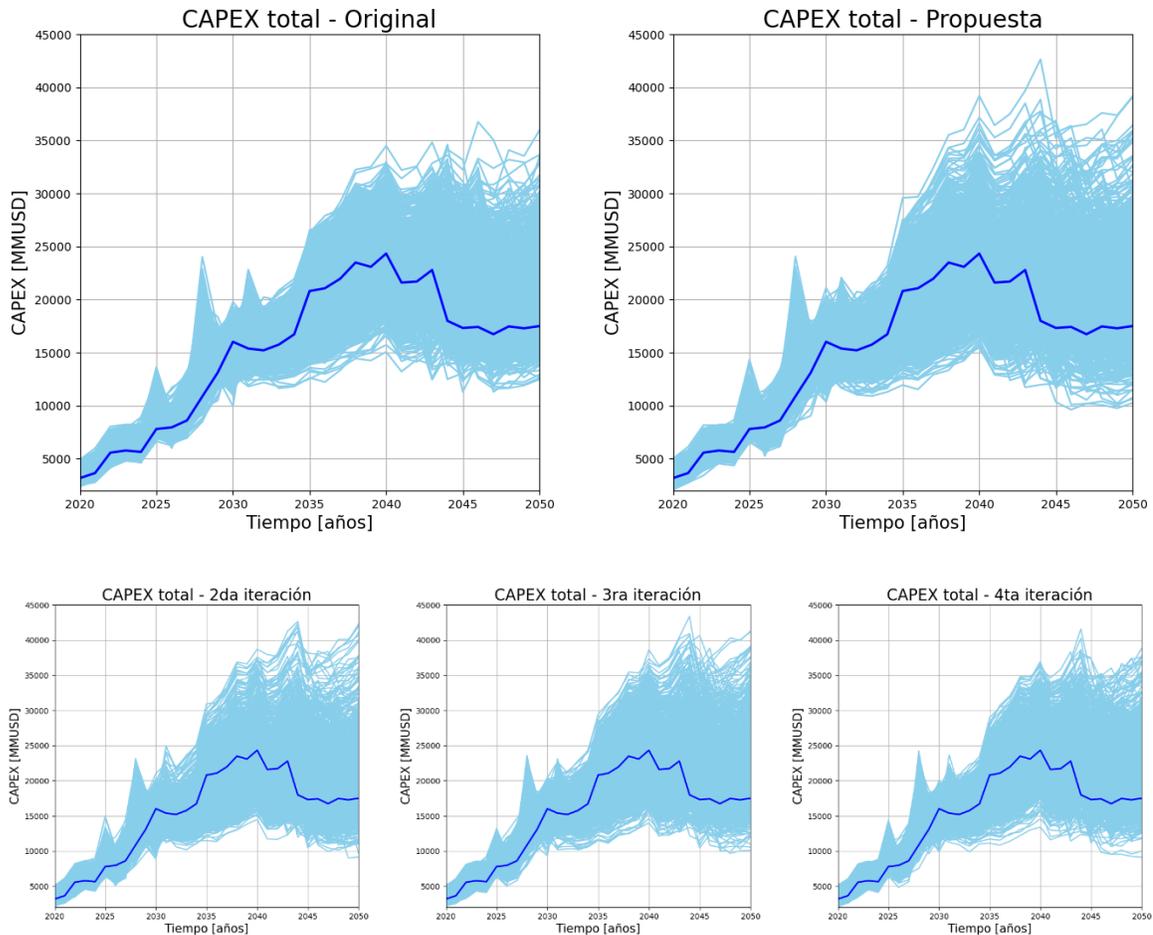


Figura 34: Trayectorias de CAPEX total o suma de sectores, para modelos original, propuesto e iteraciones. Periodo 2020-2050.

Los modelos presentan el mismo comportamiento, con la salvedad de que aumenta el rango máximo y mínimo a medida que se aplica la metodología propuesta. Efecto notorio principalmente en los picos de la trayectoria de referencia.

Por último, en Figura 35 se presentan las trayectorias de OPEX, donde se observa el aumento del rango de las trayectorias, sin cambios notorios de sus curvas.

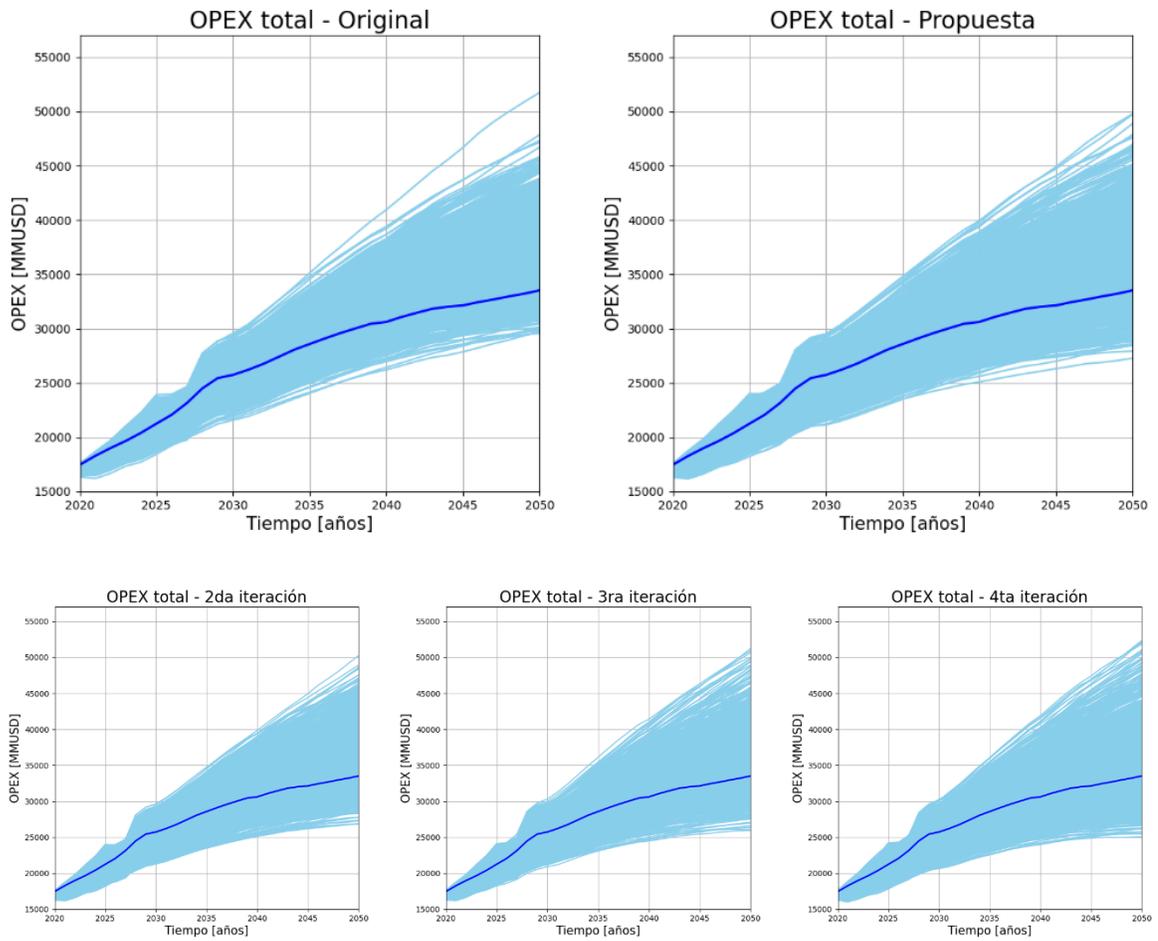


Figura 35: Trayectorias de OPEX total o suma de sectores, para modelos original, propuesto e iteraciones. Periodo 2020-2050.

## 4.2. Resultados por sectores

A continuación, se presentan los resultados principales por sector.

Se ejemplifica con resultados del modelo propuesto, en cuanto al manejo de los parámetros con PCA y *clustering k-means*.

Además, se incluye el comportamiento en cuanto a las métricas de desempeño (emisiones acumuladas y 2050, CAPEX y OPEX acumulados), en modelo original, propuesto y sus iteraciones.

Por último, cada sector concluye con las trayectorias de emisiones, CAPEX y OPEX, de los modelos original y propuesto, con la finalidad de comparar el rango y distribución de los distintos escenarios. Todo lo anterior, para el caso de estudio de medidas NDC y supuestos actuales.

### 4.2.1. Sector comercio

El sector comercio presenta solo un parámetro propio del sector con incertidumbre, junto a parámetros multisectoriales como PIB y población. Estos se presentan en sección Anexos, Tabla 8.

Luego de realizar el análisis de componentes principales (PCA), se obtiene la curvas=s de varianza explicada acumulada por componente presente en Figura 36, para escenarios de valores máximos en cuanto a métricas de desempeño (resultado similar en escenarios de valores mínimos).

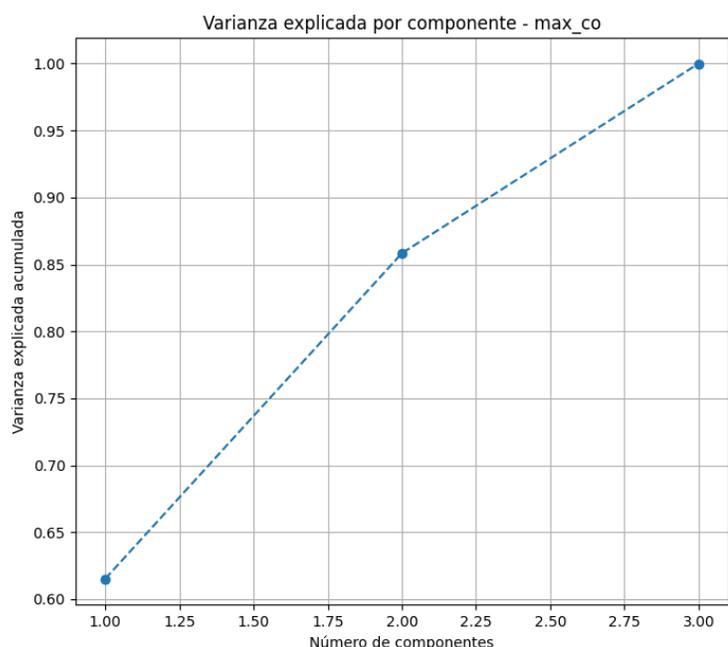


Figura 36: PCA y varianza explicada acumulada en sector comercio, para escenarios de valores máximos en cuanto a métricas de desempeño.

Se puede observar que es posible definir la matriz de datos en 3 componentes principales, misma cantidad de parámetros. Dado que con 2 componentes principales se puede obtener cerca del 85% de la información, se escoge esta cantidad a fin de visualizar el peso de cada parámetro y su influencia en el resultado del escenario.

Luego, en Figura 37 se presentan los pesos por parámetro en las dos componentes principales consideradas.

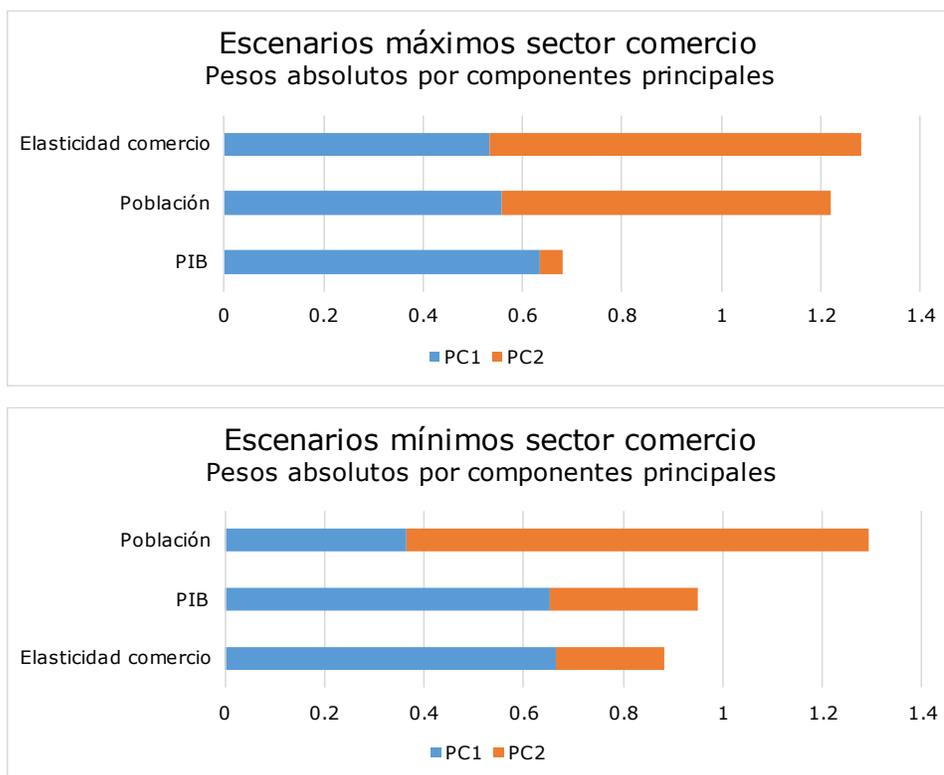


Figura 37: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos (arriba) y mínimos (abajo) del sector comercio.

Se puede observar que la elasticidad de comercio presenta resultados opuestos entre escenarios de valores máximos y mínimos. Se puede decir que, en escenarios de valores mínimos, tiene valores similares por lo que no explica mayor varianza dentro de la matriz de datos. Luego, este sería un parámetro cuyo valor predispone a escenarios de valores mínimos.

Por el contrario, en escenarios de valores máximos, presenta una mayor información de la variabilidad entre el conjunto de escenarios, por lo que no es un parámetro cuyo valor defina escenarios de valores máximos.

Por otra parte, el parámetro multisectorial población tiene mayor representación de la variabilidad de la matriz de datos que el parámetro PIB, en ambos casos.

Debido a la cantidad de parámetros del sector comercio con rango de incertidumbre definida distinto a cero, se consideran todos los parámetros en la clasificación de escenarios, por lo que no hay una reducción de parámetros.

Luego, se evalúa la cantidad de escenarios representativos, los que se obtienen de los centroides de las agrupaciones por *k-means*. En Figura 38 se presenta la curva de análisis, de la cual se obtiene una cantidad de 10 clústeres para escenarios máximos, y 4 para escenarios mínimos.

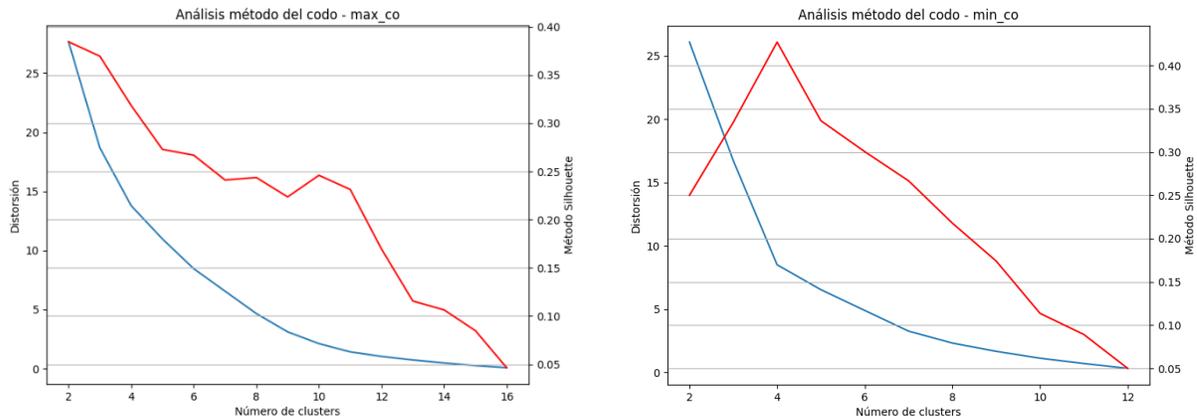


Figura 38: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector comercio, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul).

Finalmente, luego de la proyección de escenarios se tienen las métricas de desempeño para el sector comercio. Estas se representan en gráficos de caja, presentes en Figura 39 y Figura 40. También se presentan las trayectorias en Figura 41, Figura 42, y Figura 43.

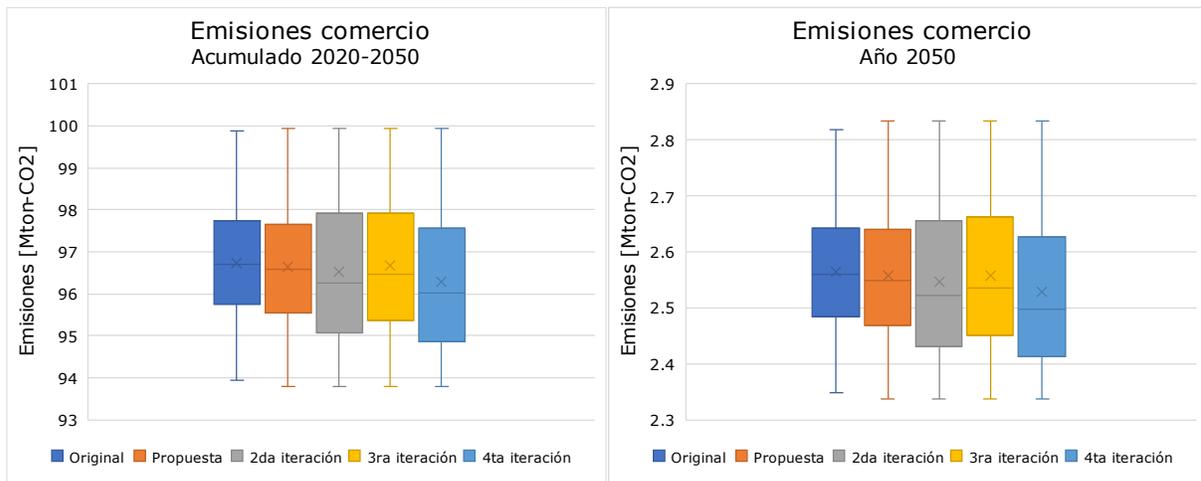


Figura 39: Gráficos de caja asociado a las emisiones del sector comercio.

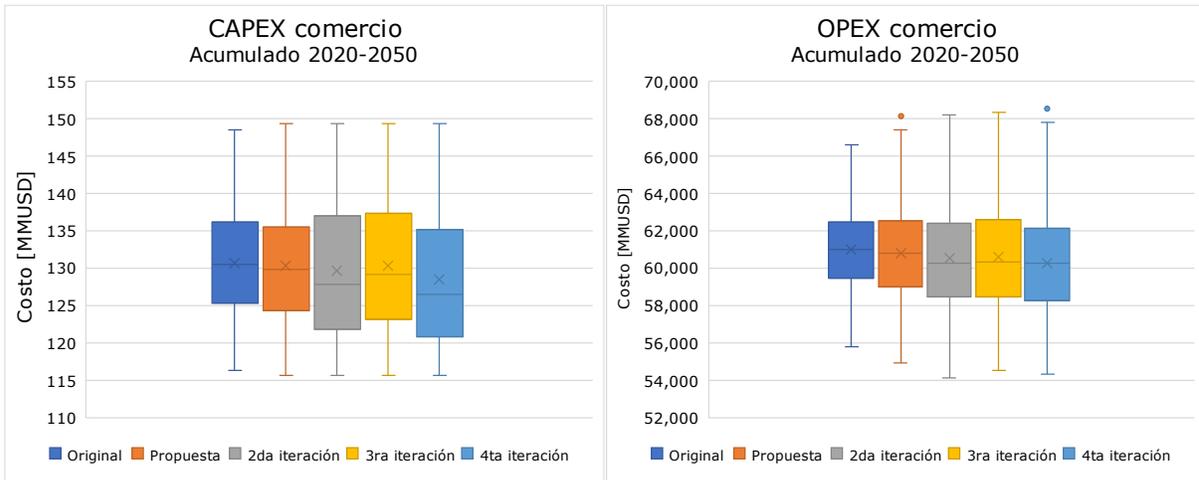


Figura 40: Gráficos de caja asociado a costos del sector comercio.

De los gráficos de caja, se observa un aumento nulo del rango de incertidumbre de emisiones acumuladas e indicador CAPEX.

No obstante, sí es posible observar un aumento en el rango de incertidumbre para las emisiones en el año 2050 al aplicar la metodología propuesta por primera vez, y para el OPEX acumulado en el periodo 2020-2050, en cada una de las iteraciones.

Nuevamente, se observa en las trayectorias de emisiones (Figura 41) y CAPEX (Figura 42), un comportamiento similar al de los gráficos de caja: se tienen valores prácticamente idénticos entre el modelo original y la aplicación de la metodología propuesta.

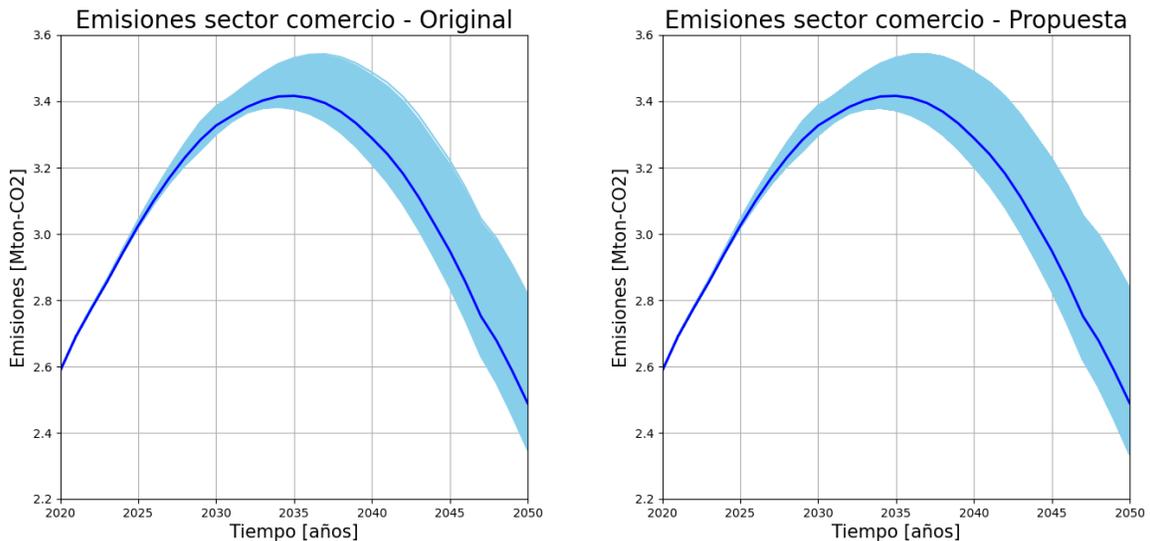


Figura 41: Trayectorias de emisiones sector comercio, modelo original y propuesto. Periodo 2020-2050.

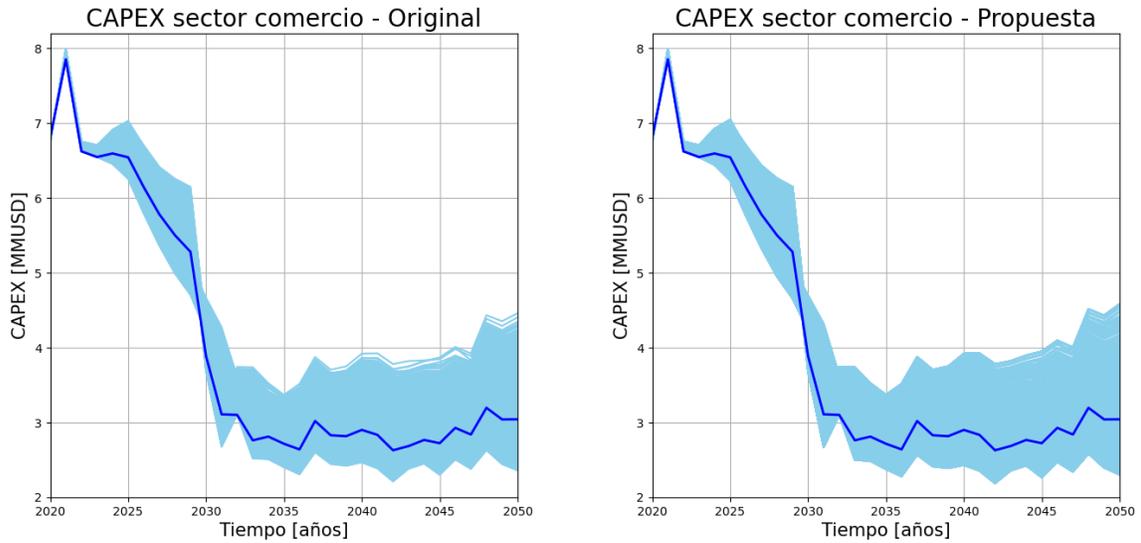


Figura 42: Trayectorias de CAPEX sector comercio, modelo original y propuesto. Periodo 2020-2050.

No obstante, en Figura 43 es posible observar la presencia de trayectorias de mayor y menor OPEX para el modelo propuesto, en comparación al modelo de referencia.

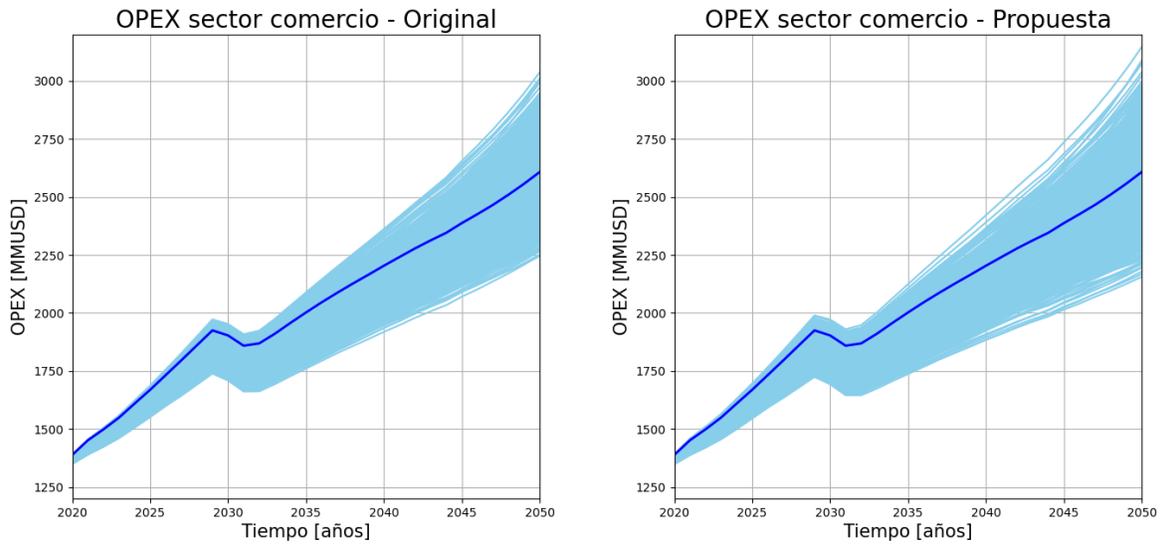


Figura 43: Trayectorias de OPEX sector comercio, modelo original y propuesto. Periodo 2020-2050.

## 4.2.2. Sector industria y minería

El sector industria y minería presenta múltiples parámetros con rango de incertidumbre definida. Estos se presentan en sección Anexos, Tabla 9.

Luego de realizar el análisis de componentes principales (PCA), se obtiene la curva de varianza explicada acumulada por número de componentes principales, presente en Figura 44. Este corresponde a los escenarios de valores máximos, de similar comportamiento al de escenarios de valores mínimos.

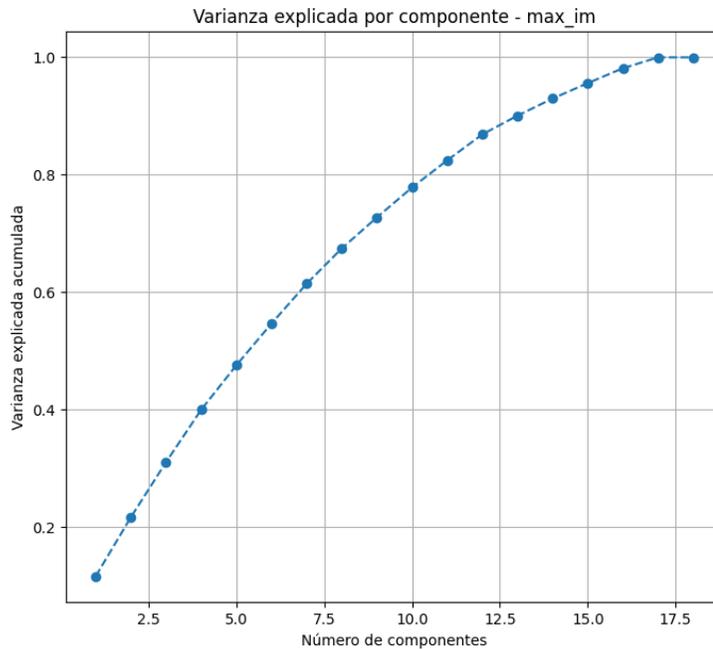


Figura 44: PCA y varianza explicada acumulada en sector industria y minería, para escenarios de valores máximos en cuanto a métricas de desempeño.

Se puede observar que, con 10 componentes principales, es posible explicar un 80% de la varianza de la matriz de datos, asociada a los escenarios de valores máximos del sector, siendo entonces la cantidad de componentes principales escogidas.

Para los escenarios de valores mínimos, se tienen 14 componentes principales, las cuales explican, también, el 80% de la varianza acumulada.

Luego, en Figura 45 y Figura 46 se presentan los pesos por parámetro para las componentes principales consideradas en los casos de escenarios de valores máximos y de valores mínimos, respectivamente.

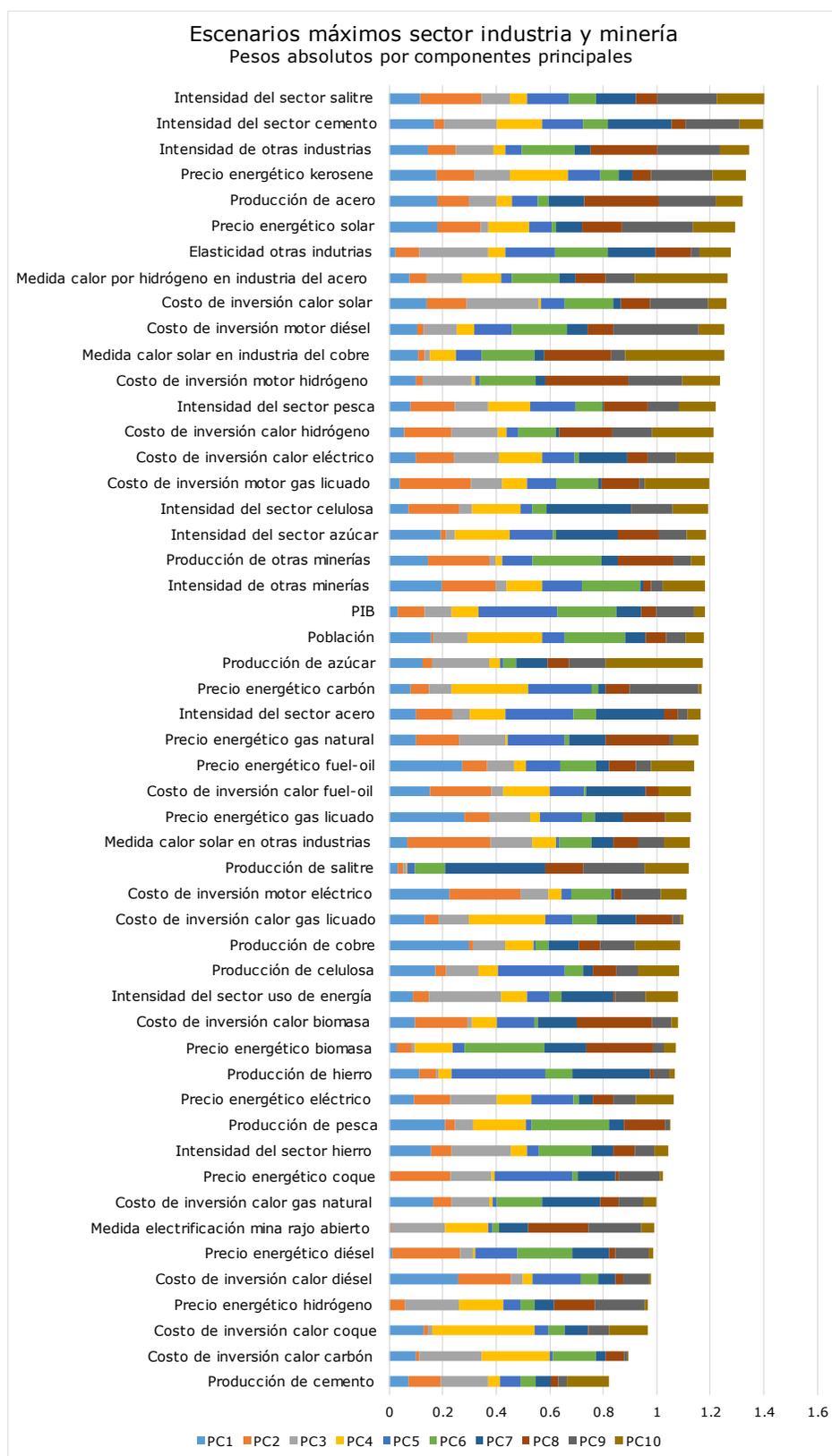


Figura 45: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos del sector industria y minería.

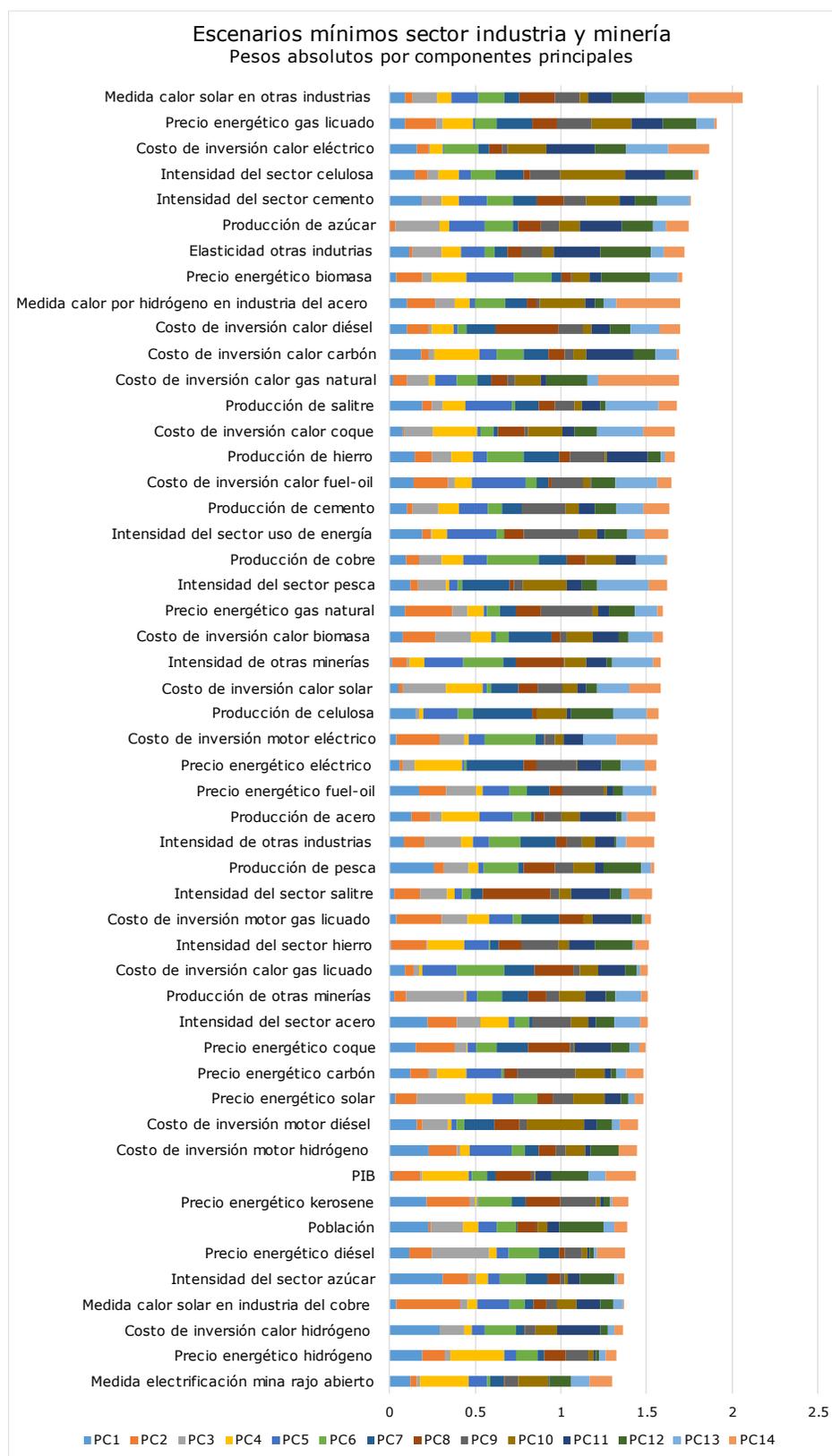


Figura 46: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores mínimos del sector industria y minería.

Para posteriormente aplicar el algoritmo de *clustering* a los parámetros de mayor representación de variabilidad, se escogen los siguientes parámetros para la matriz de escenarios de valores máximos:

- Intensidad del sector cemento.
- Precio energético kerosene.
- Costo de inversión en calor de fuente solar.
- Intensidad del sector salitre.
- Producción de acero.

Mientras que, para la matriz de escenarios de valores mínimos, se tienen los siguientes parámetros:

- Intensidad del sector cemento
- Precio energético gas licuado.
- Intensidad del sector celulosa.
- Medida de mitigación - calor de fuente solar en industrias varias.

Recordar de la sección Metodología, que la cantidad de parámetros se escoge en función de identificar curvas pronunciadas en las métricas de validación del algoritmo de *clustering*. A su vez, se escogen los parámetros de mayor peso ponderado según la varianza explicada de sus componentes principales (a mayor varianza explicada, mayor peso ponderado).

Luego, con estos parámetros se analiza la cantidad de escenarios representativos por *clustering k-means*. En Figura 47 se presenta la curva de análisis, de la cual se obtiene una cantidad de 9 clústeres para escenarios máximos, y 13 para escenarios mínimos.

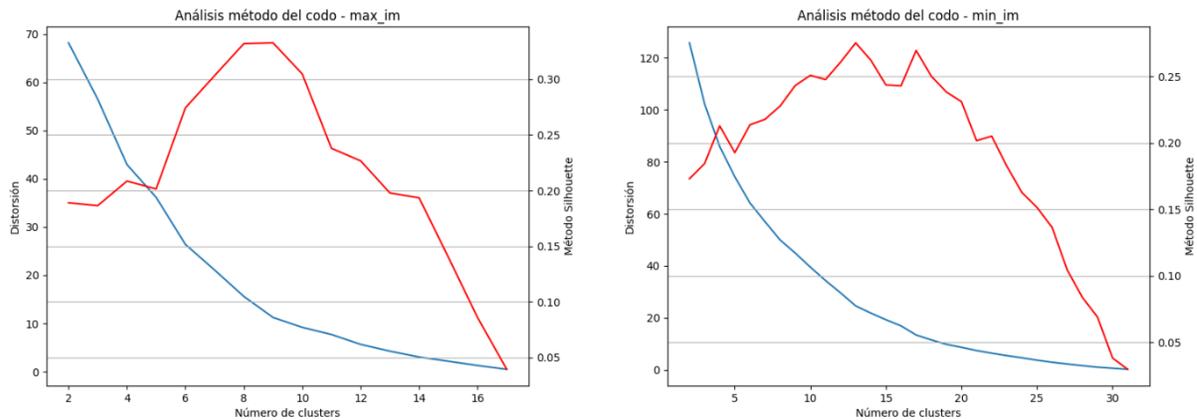


Figura 47: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector industria y minería, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul).

Finalmente, luego de la proyección de escenarios se tienen las métricas de desempeño para el sector industria y minería, presentes en Figura 48 y Figura 49. También se presentan las trayectorias en Figura 50, Figura 51 y Figura 52.

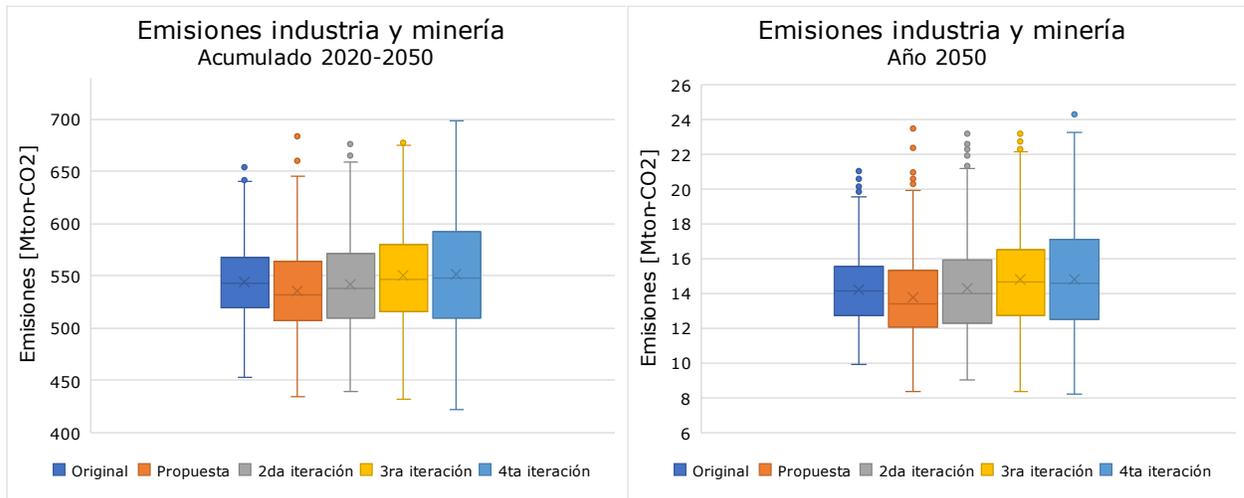


Figura 48: Gráficos de caja asociado a emisiones del sector industria y minería.

En cuanto a las emisiones acumuladas y emisiones del año 2050, se puede observar un aumento progresivo del límite superior sin considerar escenarios atípicos (puntos exteriores del rango). Mientras que, para el límite inferior, se tiene una disminución progresiva a excepción de la primera aplicación de la metodología propuesta, donde se tienen escenarios de bajas emisiones similares al de una 4ta iteración.

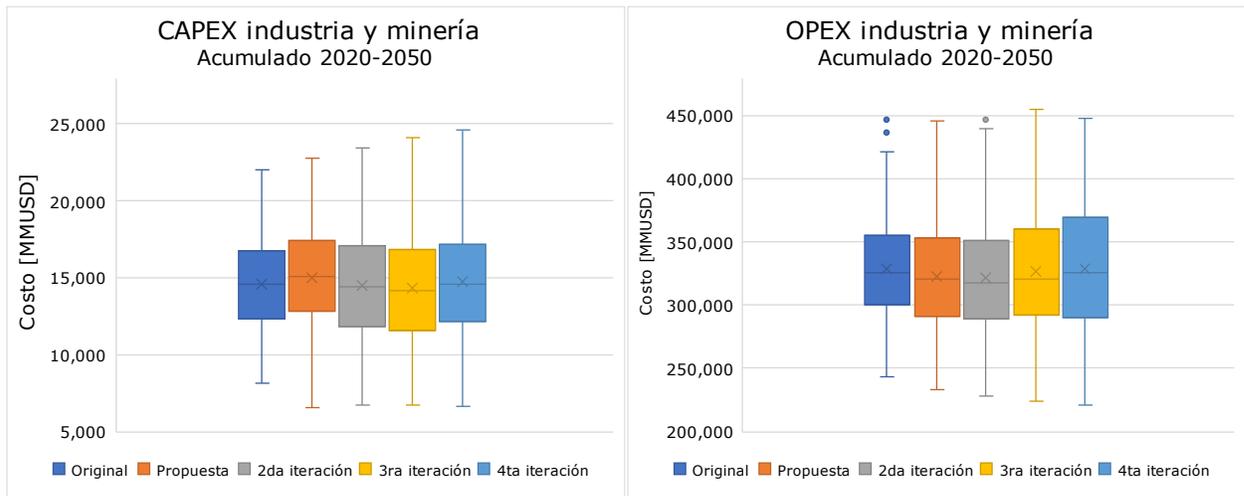


Figura 49: Gráficos de caja asociado a costos del sector industria y minería.

Mientras que para los indicadores de CAPEX y OPEX acumulados en el periodo, se tiene un aumento de rango para el modelo propuesto en comparación al modelo original o de referencia.

Para el indicador CAPEX, se tiene un aumento del límite superior en cada iteración, mientras que el límite inferior presenta un comportamiento constante.

Para el indicador OPEX, ambos límites tienen comportamiento constante con un aumento leve del rango en cada iteración.

Las observaciones realizadas a los gráficos de caja se reflejan en el comportamiento de las trayectorias de las métricas de desempeño.

En cuanto a las emisiones, presentes en Figura 50, se tiene un comportamiento similar entre el modelo de original y propuesto los primeros 10 años. Luego, las trayectorias del modelo propuesto se expanden respecto a su referencia, logrando un rango mayor de incertidumbre en el año 2050.

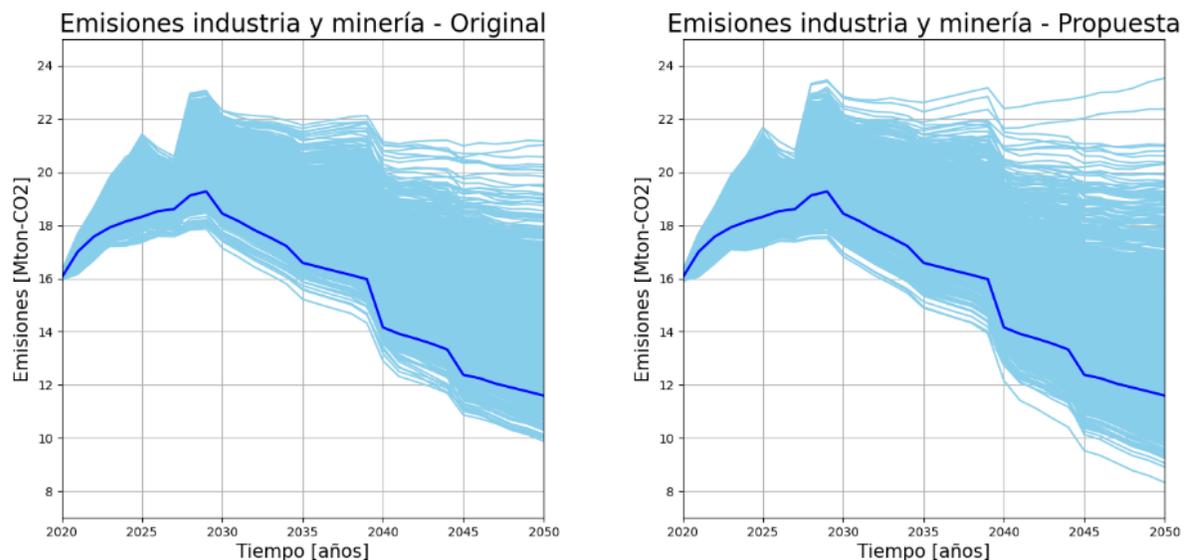


Figura 50: Trayectorias de emisiones sector industria y minería, modelo original y propuesto. Periodo 2020-2050.

Respecto a los indicadores de costos, se tienen trayectorias similares que difieren principalmente en los últimos años, con mayor notoriedad en los picos de la trayectoria de referencia.

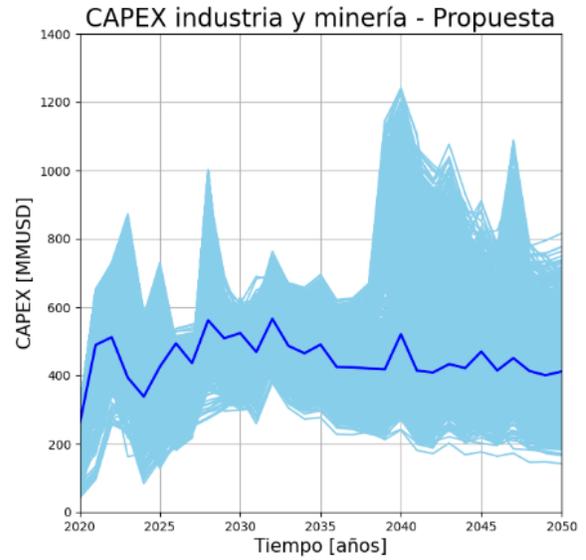
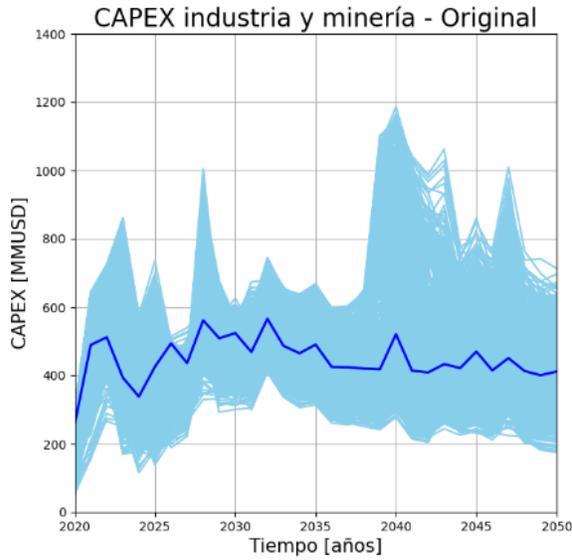


Figura 51: Trayectorias de CAPEX sector industria y minería modelo original y propuesto. Periodo 2020-2050.

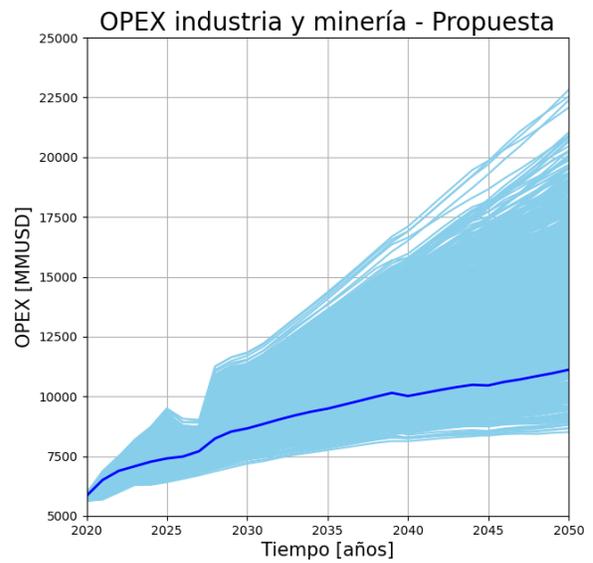
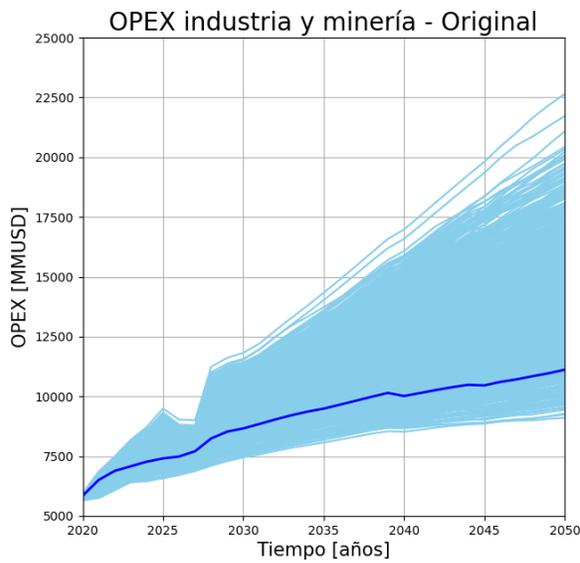


Figura 52: Trayectorias de OPEX sector industria y minería, modelo original y propuesto. Periodo 2020-2050.

### 4.2.3. Sector público

El sector público presenta solo un parámetro propio del sector con incertidumbre, de igual forma que sector comercio, junto a parámetros multisectoriales. Estos se presentan en sección Anexos, Tabla 10.

Luego de realizar el análisis de componentes principales (PCA), se obtiene la curva de varianza explicada acumulada por componente presente en Figura 53, para escenarios de valores máximos en cuanto a métricas de desempeño. Para el caso de valores mínimos, se tiene un comportamiento similar.

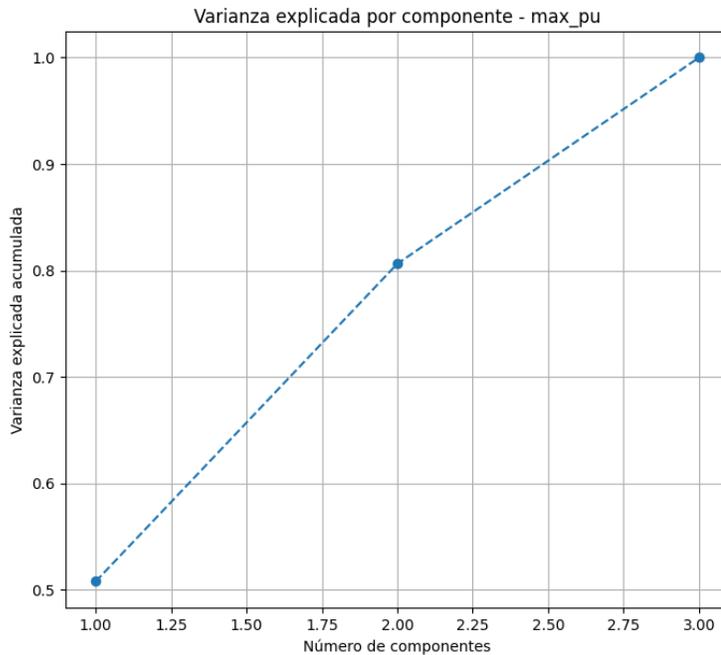


Figura 53: PCA y varianza explicada acumulada en sector público, para escenarios de valores máximos en cuanto a métricas de desempeño.

Dado que con dos componentes principales se puede obtener cerca de un 80% de varianza explicada acumulada, se escoge esta cantidad a fin de visualizar el peso de cada parámetro y su influencia en el resultado del escenario.

Luego, en Figura 54 se presentan los pesos por parámetro en las dos componentes principales consideradas.

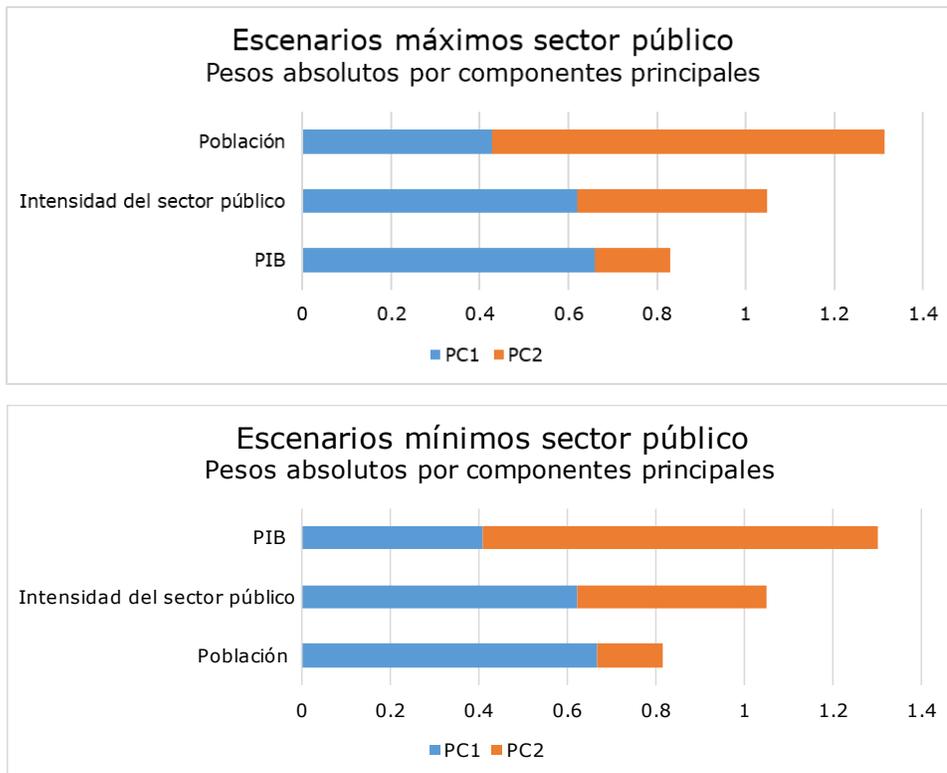


Figura 54: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos (arriba) y mínimos (abajo) del sector público.

Contrario a los resultados del sector comercio, en el sector público los parámetros multisectoriales PIB y población toman resultados opuestos dependiendo si se trata de escenarios de valores máximos o mínimos. Mientras que el parámetro de intensidad del sector público toma un rol similar entre ambos casos, en cuanto a al peso o importancia que presenta entre las componentes principales.

Nuevamente, debido a la cantidad de parámetros del sector público con rango de incertidumbre definida distinto a cero, se consideran todos los parámetros en la clasificación de escenarios. Es decir, no hay una reducción de parámetros.

Luego, se evalúa la cantidad de escenarios representativos, los que se obtienen de los centroides de las agrupaciones por *k-means*. En Figura 55 se presenta la curva de análisis, de la cual se obtiene una cantidad de 4 clústeres para escenarios de valores máximos, y 5 para escenarios de valores mínimos.

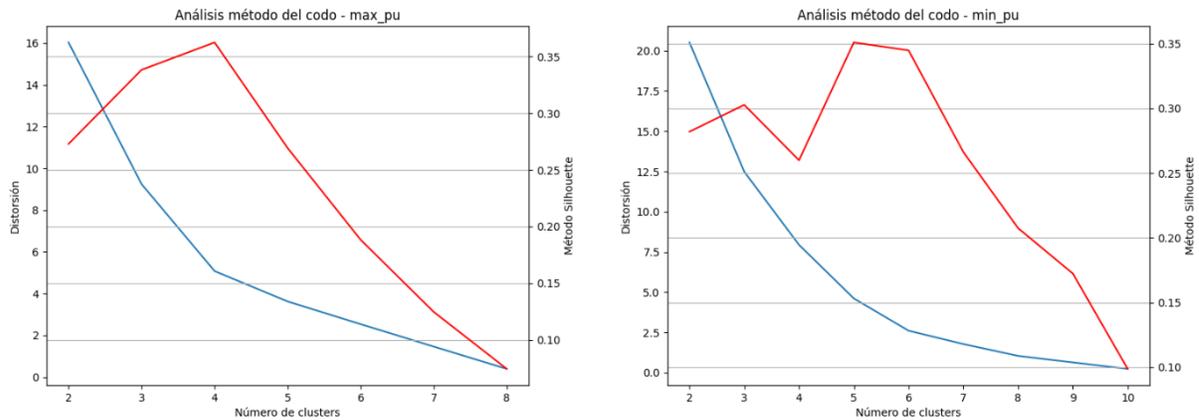


Figura 55: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector público, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul).

Finalmente, luego de la proyección de escenarios se tienen las métricas de desempeño para el sector público, presentes en Figura 56. También se presentan las trayectorias en Figura 57.

Recordar que el sector público no presenta resultados asociados a costos dado que no presenta medidas de mitigación.

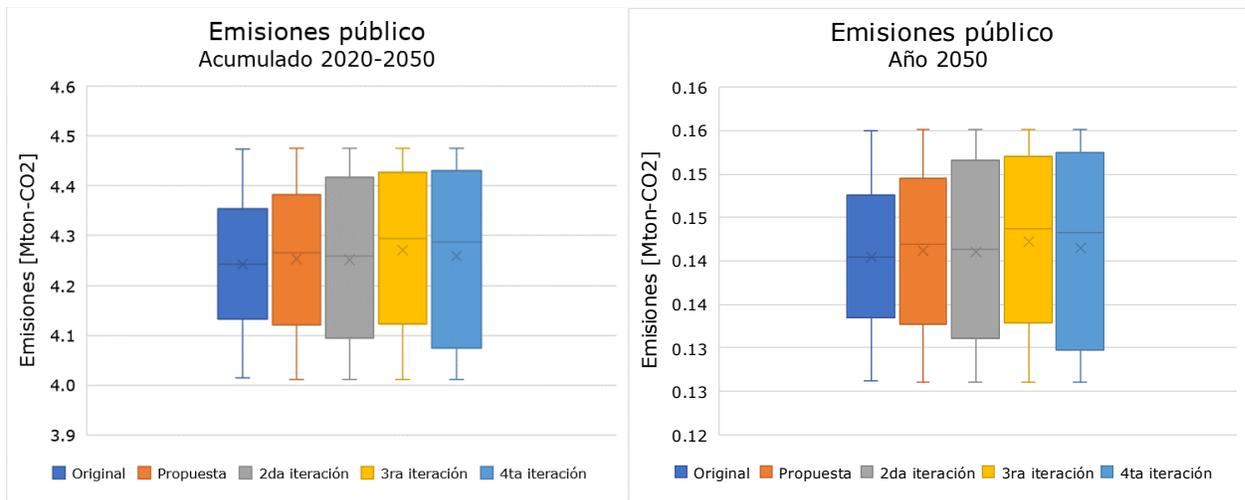


Figura 56: Gráficos de caja asociado a las emisiones del sector público.

De los gráficos de caja se puede observar que los rangos de incertidumbre permanecen constantes, mientras que los límites entre cuartiles se desplazan hacia los extremos. Recordando que se tiene la misma cantidad de escenarios en cada cuartil, se tiene un aumento de concentración de escenarios en los extremos.

Luego, comparando las trayectorias entre modelo de referencia y propuesto, se tiene un rango de incertidumbre idéntico. La principal diferencia, es que en las trayectorias del modelo propuesto se presentan espacios entre trayectorias más

pronunciadas que el modelo original, lo que se condice con la concentración en los extremos presente en los gráficos de caja.

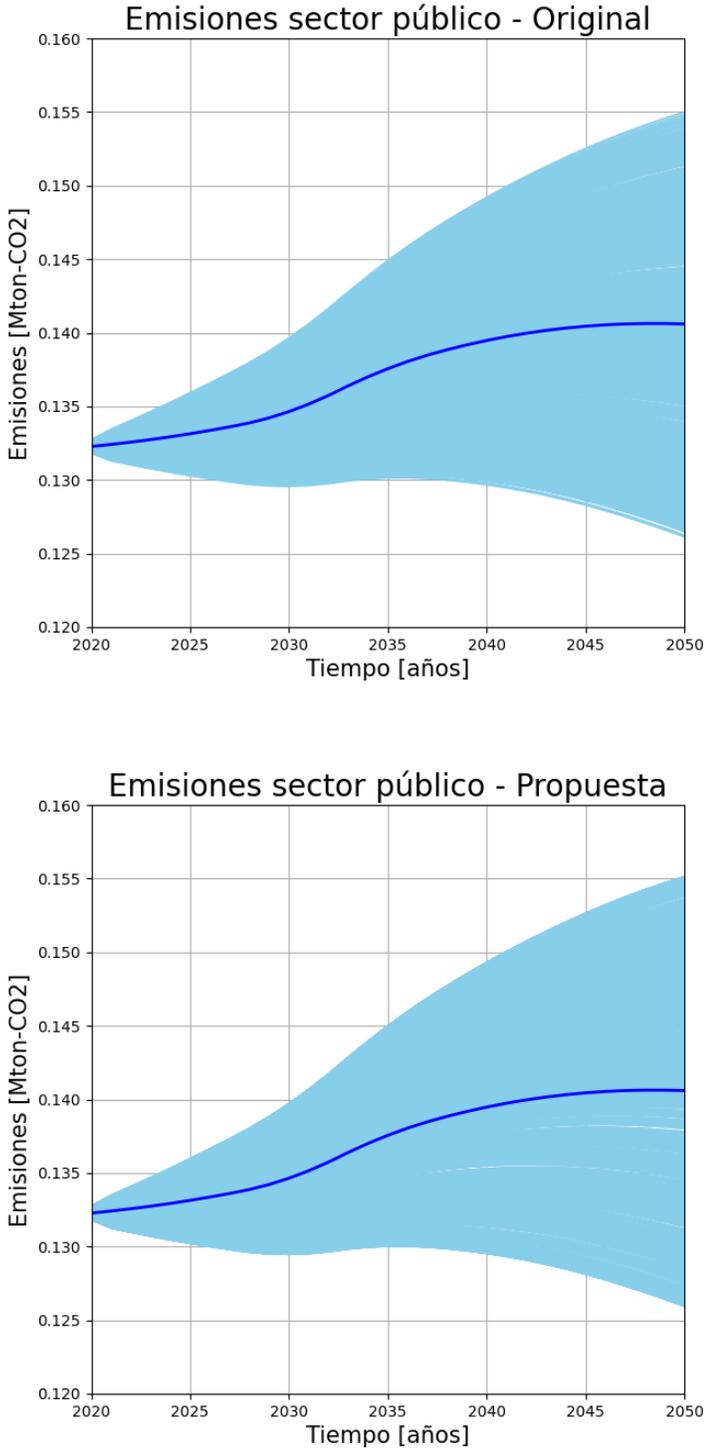


Figura 57: Trayectorias de emisiones sector público, modelo original y propuesto. Periodo 2020-2050.

#### 4.2.4. Sector residencial

El sector residencial presenta múltiples parámetros con rango de incertidumbre, los cuales se presentan en sección Anexos, Tabla 11.

A partir del análisis de componentes principales (PCA), se obtiene la curva de varianza explicada acumulada por componente presente en Figura 58, correspondiente a los escenarios de valores máximos, de similar comportamiento al de escenarios de valores mínimos.

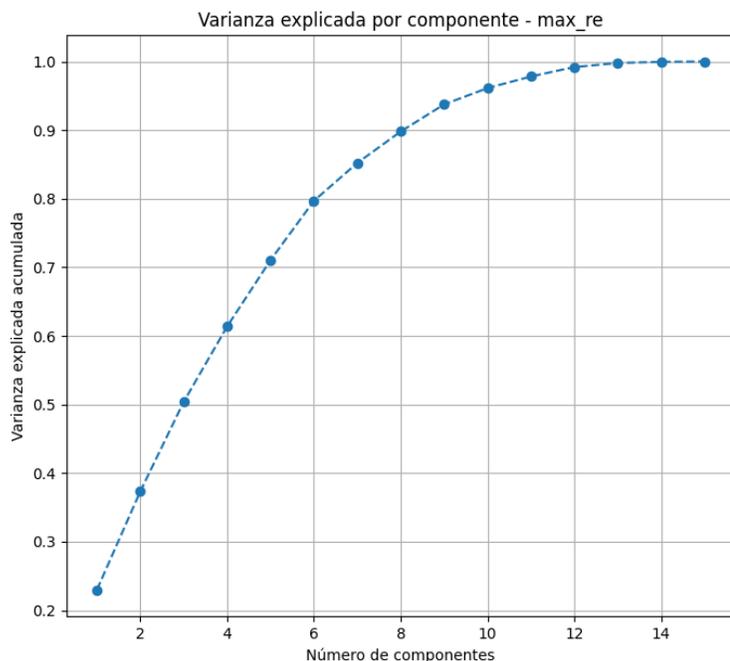


Figura 58: PCA y varianza explicada acumulada en sector residencial, para escenarios de valores máximos en cuanto a métricas de desempeño.

Se puede observar que, con 8 componentes principales, se explica el 90% de la varianza acumulada, y coincide con la disminución de la pendiente, es decir, que de considerar una componente principal más, no se vería un aumento considerable.

Luego, se consideran 8 componentes principales en ambos casos, de escenarios de valores máximos y mínimos, ambos con un 90% de varianza explicada acumulada.

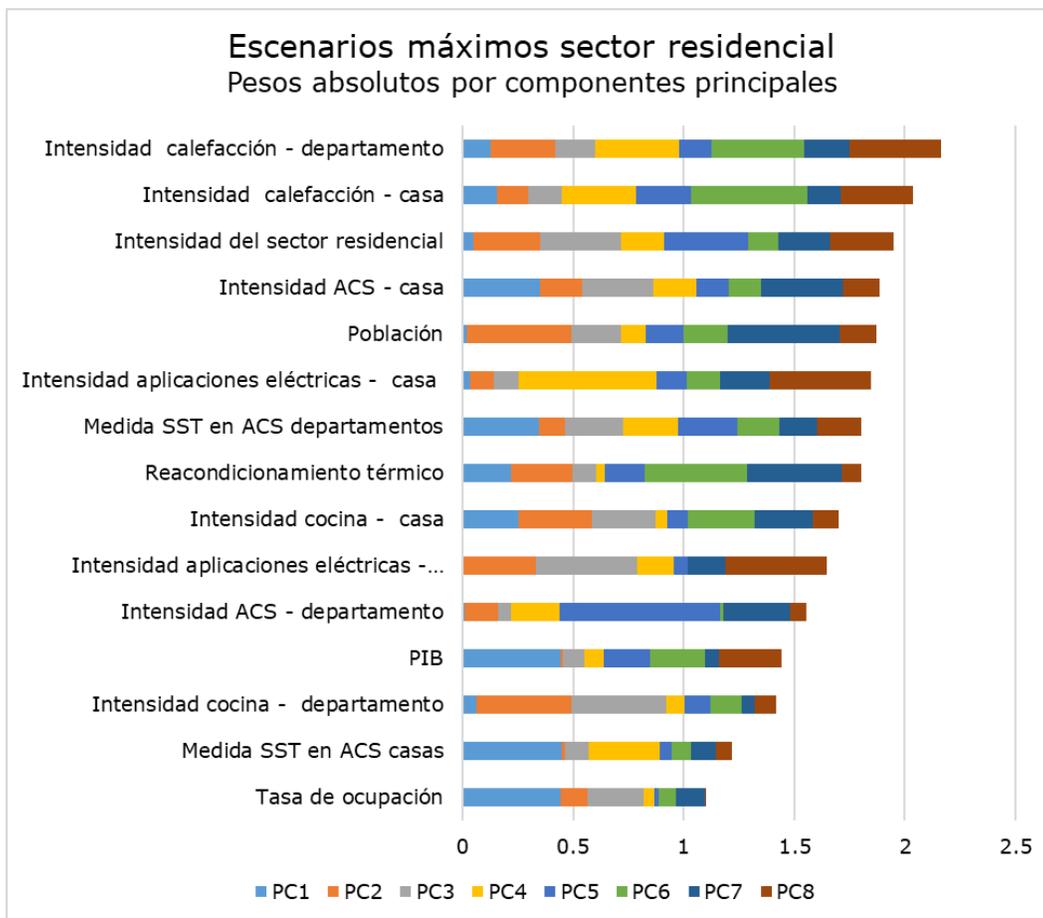


Figura 59: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos del sector residencial.

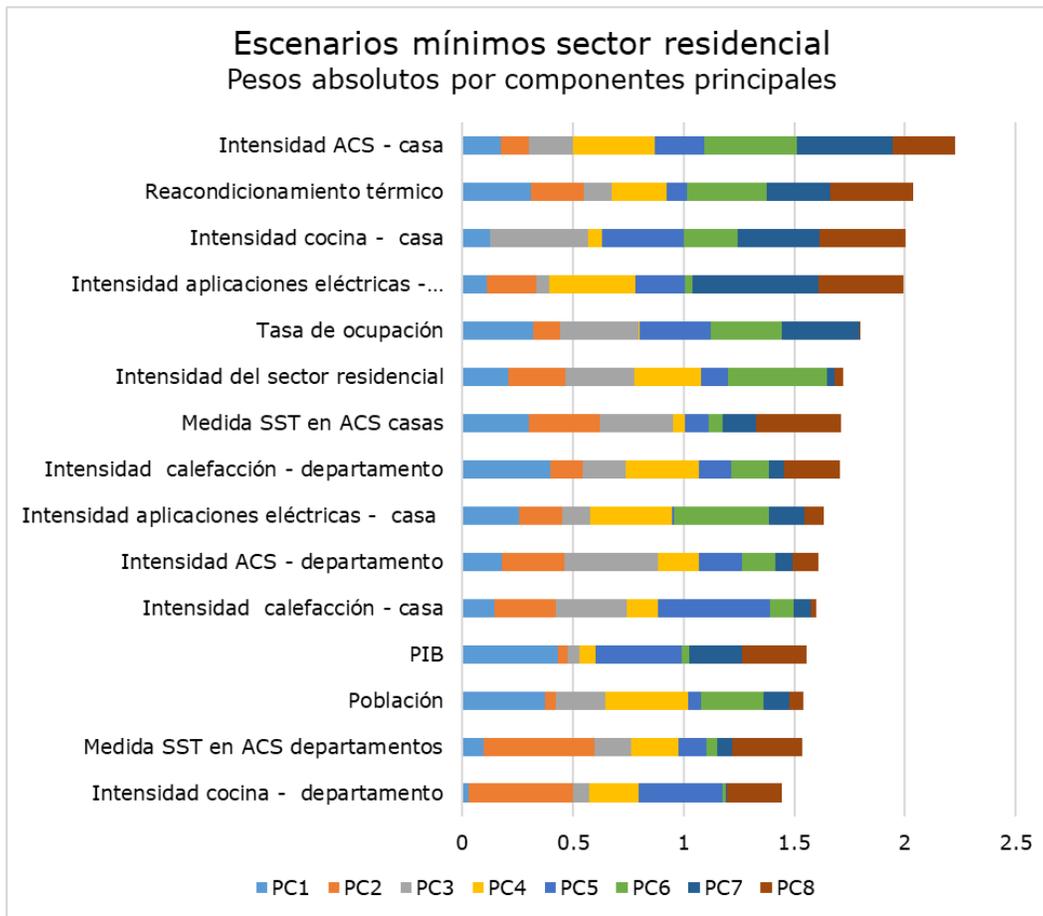


Figura 60: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores mínimos del sector residencial.

Para luego aplicar el algoritmo de *clustering* a los parámetros de mayor representación de variabilidad, se escogen los siguientes parámetros de la matriz de escenarios de valores máximos:

- Intensidad de calefacción – departamento.
- Intensidad de ACS – casa.
- Intensidad de cocina – casa.
- Intensidad de calefacción – casa.
- Intensidad del sector residencial.
- Medida de sistemas solares térmicos en ACS – departamento.

Mientras que, para la matriz de escenarios de valores mínimos, se tienen los siguientes parámetros:

- Intensidad de cocina – departamento.
- Intensidad de ACS – casa.
- Intensidad del sector residencial.
- Tasa de ocupación.
- Medida de sistemas solares térmicos en ACS – casa.
- Reacondicionamiento térmico de hogar.

Luego, con estos parámetros representativos se evalúa la cantidad de escenarios representativos, los que se obtienen de los centroides de las agrupaciones por *k-means*. En Figura 61 se presentan las curvas de análisis, de las cuales se obtiene una cantidad de 9 clústeres para escenarios máximos, y 7 para escenarios mínimos.

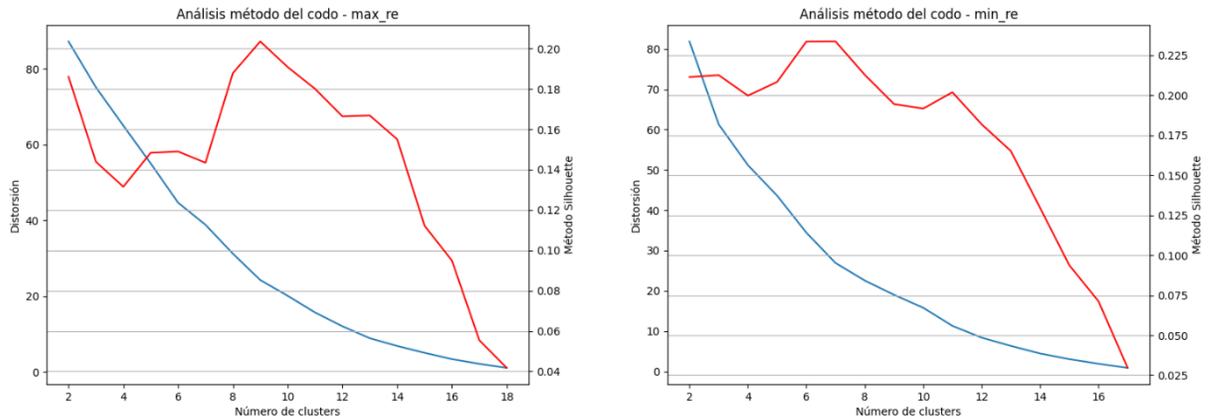


Figura 61: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector residencial, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul).

Finalmente, luego de la proyección de escenarios se tienen las métricas de desempeño para el sector residencial, presentes en Figura 62 y Figura 63.

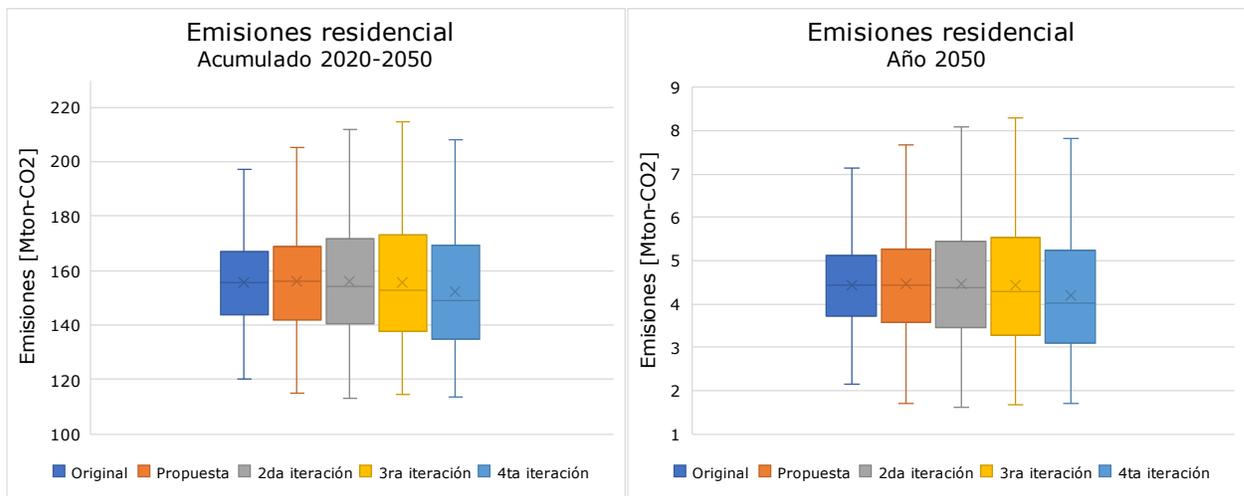


Figura 62: Gráficos de caja asociado a emisiones del sector residencial.

En cuanto a las emisiones acumuladas en el periodo y del año 2050, se tiene un comportamiento similar en la comparación entre modelos: un aumento del rango de incertidumbre entre un modelo y otro, excepto por la 4ta iteración de la metodología propuesta.

Se puede observar también una disminución del promedio a medida que se itera la metodología propuesta, así como una disminución del rango del cuartil inferior.

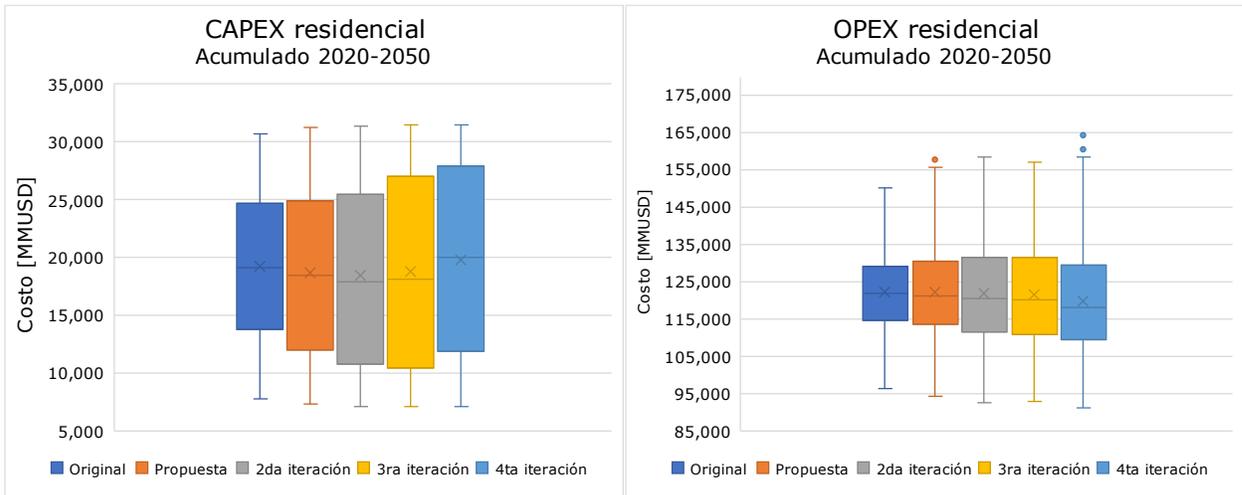


Figura 63: Gráficos de caja asociado a costos del sector residencial.

Respecto a los costos asociados a los indicadores CAPEX y OPEX, se tiene un aumento del rango de incertidumbre entre el modelo de referencia u original, y el modelo propuesto o de referencia con metodología aplicada.

Luego, las iteraciones demuestran una concentración mayor de trayectorias en los extremos, recordando que cada uno de los cuartiles presenta la misma cantidad de escenarios que el resto.

También se presentan las trayectorias en Figura 64, Figura 65 y Figura 66.

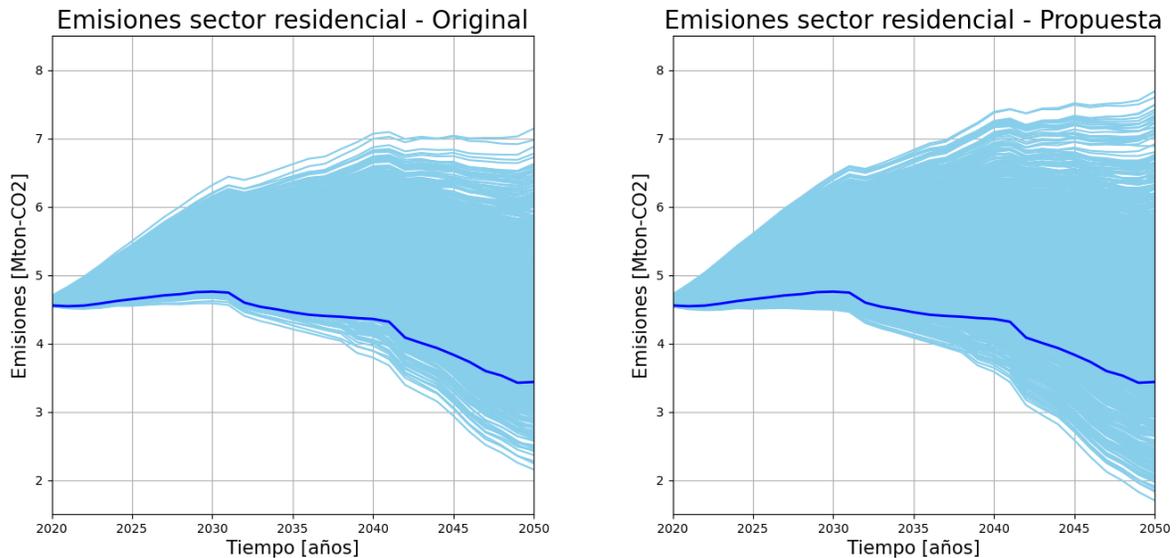


Figura 64: Trayectorias de emisiones sector residencial, modelo original y propuesto. Periodo 2020-2050.

Las trayectorias de emisiones del sector residencial demuestran el aumento del rango de incertidumbre, así como una mayor concentración de escenarios en los extremos, comparando el modelo propuesto respecto al modelo de referencia.

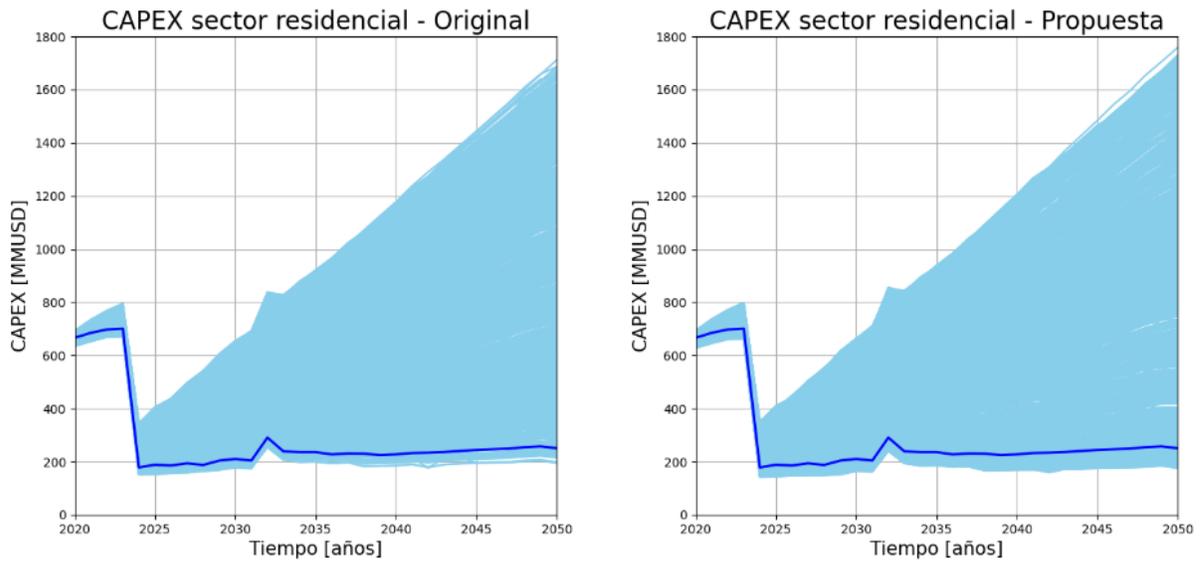


Figura 65: Trayectorias de CAPEX residencial, modelo original y propuesto. Periodo 2020-2050.

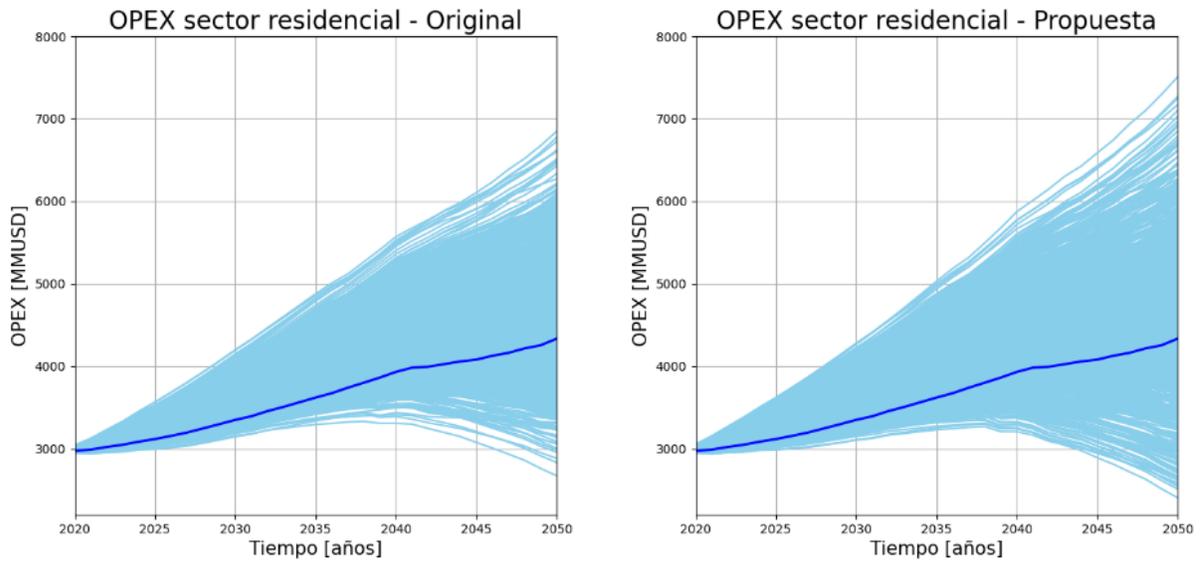


Figura 66: Trayectorias de OPEX sector residencial, modelo original y propuesto. Periodo 2020-2050.

Finalmente, destacar el aumento de trayectorias proyectadas bajo la trayectoria de referencia para ambos indicadores de costos del tipo CAPEX y OPEX, así como un aumento de la trayectoria de máximos valores.

#### 4.2.5. Sector transporte

El sector transporte presenta múltiples parámetros con rango de incertidumbre definida, los cuales se presentan en sección Anexos, Tabla 12.

Luego de aplicar el análisis de componentes principales (PCA), se obtiene la curva de varianza explicada acumulada por componente presente en Figura 67.

Considerando escoger la cantidad de componentes principales desde donde la curva de varianza explicada acumulada se encuentra disminuyendo su pendiente, para los escenarios de valores máximos se tienen 12 componentes principales, y para valores mínimos se tienen 25 componentes principales. Lo anterior equivale a un 80% y 90% de varianza explicada acumulada, respectivamente.

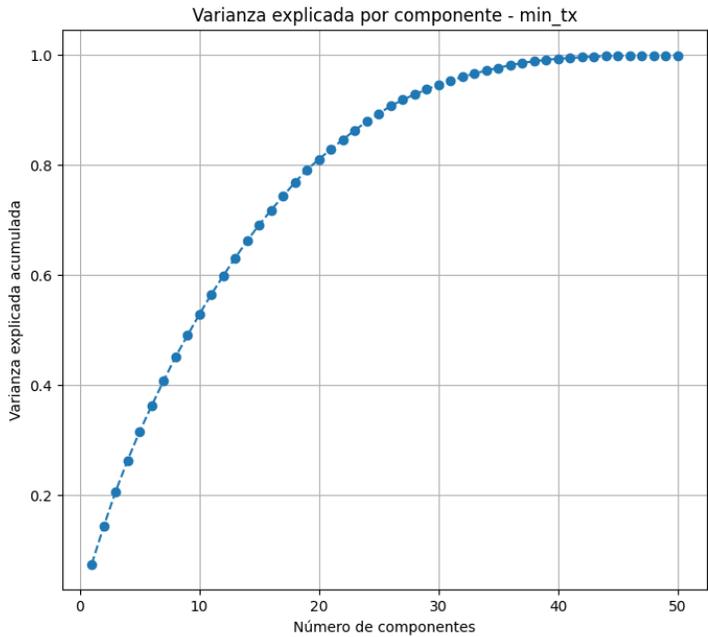
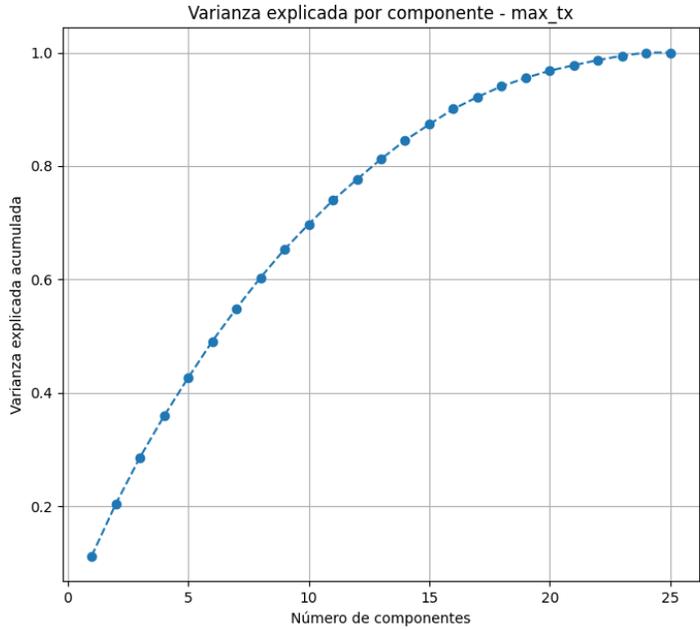


Figura 67: PCA y varianza explicada acumulada en sector transporte, para escenarios de valores máximos (arriba) y mínimos (abajo) en cuanto a métricas de desempeño.

Luego, se presentan los pesos por parámetro para la cantidad de componentes principales consideradas, en el caso de escenarios de valores máximos (ver Figura 68) y en el de escenarios de valores mínimos (ver Figura 49).

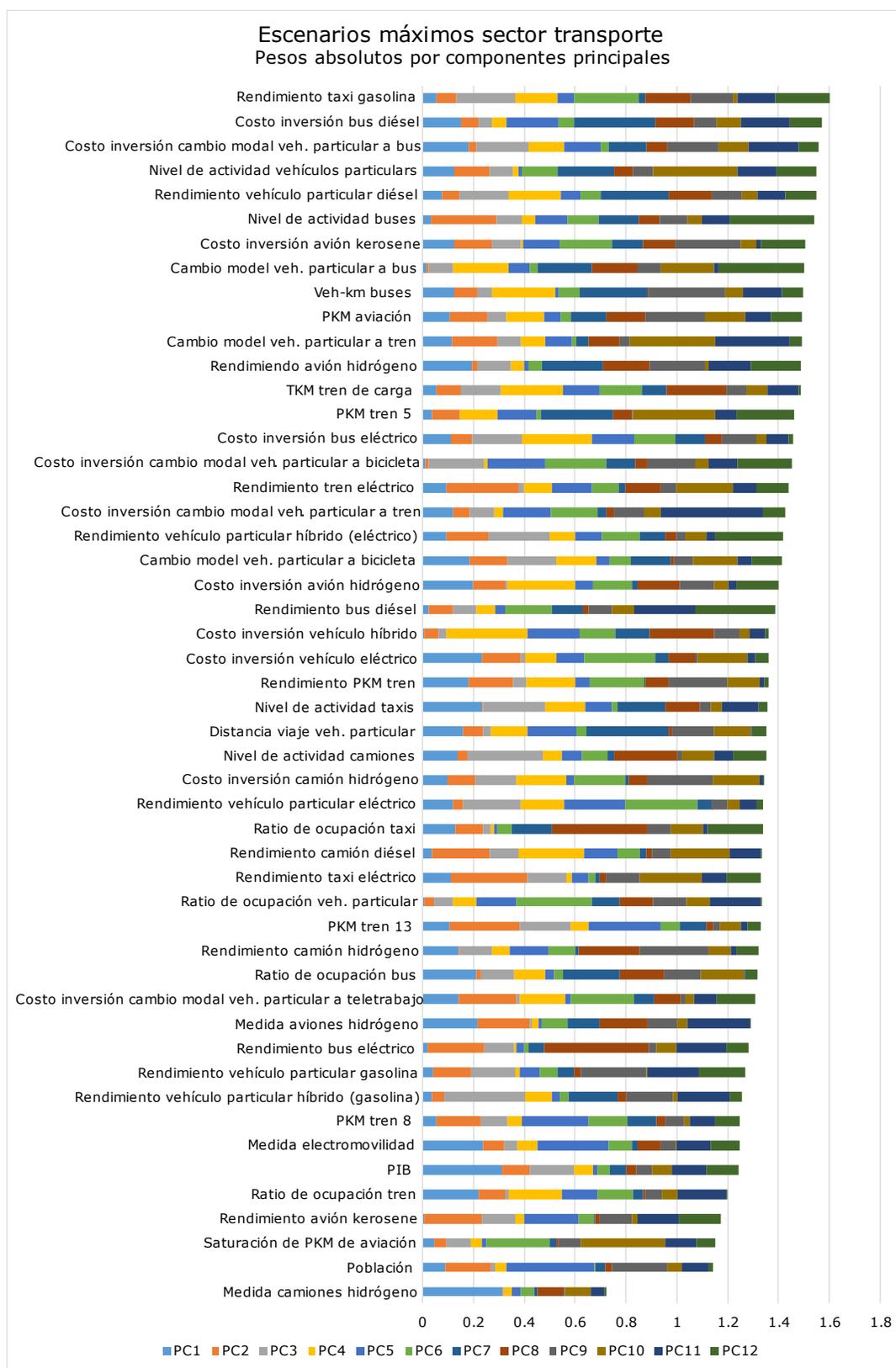


Figura 68: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos del sector transporte.



Con la finalidad de aplicar el algoritmo de *clustering k-means* a los parámetros de mayor variabilidad en la matriz de datos, se escogen los siguientes parámetros para el caso de escenarios de valores máximos:

- Costo de inversión bus eléctrico.
- Rendimiento de vehículo particular diésel.
- Rendimiento de taxi a gasolina.
- Costo de inversión cambio modal vehículo particular a bus.
- Costo de inversión cambio modal vehículo particular a bicicleta.

Mientras que, para los escenarios de valores mínimos, se tienen los siguientes parámetros:

- Medida de electromovilidad en vehículos particulares y comerciales.
- Población.
- Nivel de actividad vehículos particulares.
- Rendimiento PKM trenes.
- PKM tren de 13 vagones.

Luego, con los parámetros representativos se evalúa la cantidad de escenarios representativos, los que se obtienen de los centroides de las agrupaciones por *k-means*. En Figura 70 se presentan las curvas de análisis, de las cuales se obtiene una cantidad de 13 clústeres para escenarios de valores máximos y 13 para escenarios de valores mínimos.

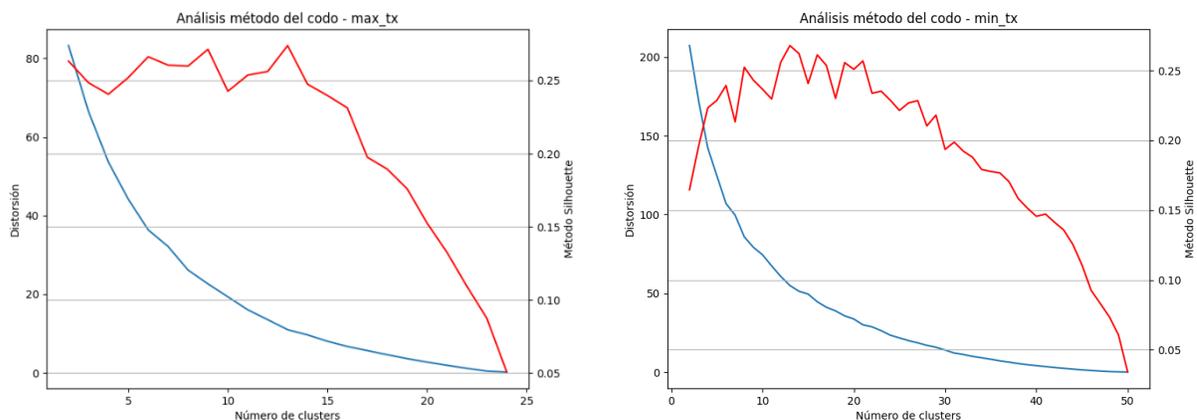


Figura 70: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector transporte, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul).

Finalmente, luego de la proyección de escenarios se tienen las métricas de desempeño para el sector transporte, presentes en Figura 71 y Figura 72.

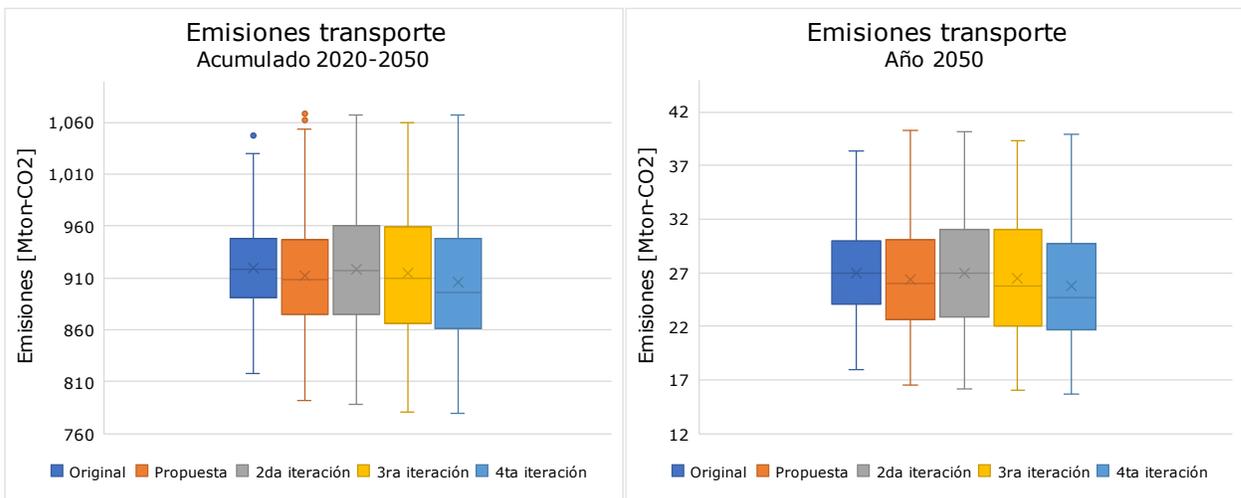


Figura 71: Gráficos de caja asociado a emisiones del sector transporte.

De los gráficos de emisiones, se destaca que el mayor aumento del rango de incertidumbre se produce la primera vez que se aplica la metodología propuesta. Incluso, para las emisiones del año 2050, se tiene que el mayor rango lo tiene el modelo propuesto, es decir, el modelo de referencia con metodología aplicada una sola vez.

Como es la tónica en los distintos sectores, la metodología propuesta también aumenta la concentración de escenarios en los extremos, ya que además de aumentar los límites superiores e inferiores, expande los límites entre cuartiles desde el centro.

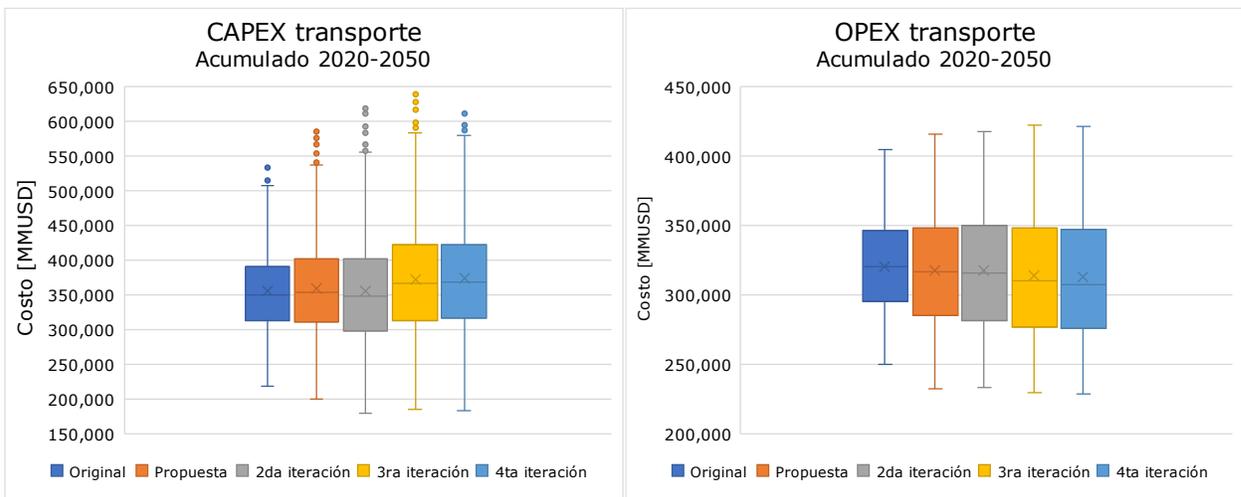


Figura 72: Gráficos de caja asociado a costos del sector transporte.

En cuanto al indicador CAPEX, se tiene un aumento del rango de incertidumbre tanto de los extremos del gráfico de caja, como de los valores atípicos, desde el modelo de referencia hasta la 3ra iteración. Nuevamente, la 4ta iteración presenta una contracción del rango de incertidumbre, fenómeno visto en distintos sectores.

Mientras que el indicador OPEX presenta un aumento de sus límites superiores e inferiores cada vez que se itera la aplicación de la metodología propuesta.

En Figura 73, Figura 74 y Figura 75, se presentan las trayectorias de las métricas de desempeño del sector transporte.

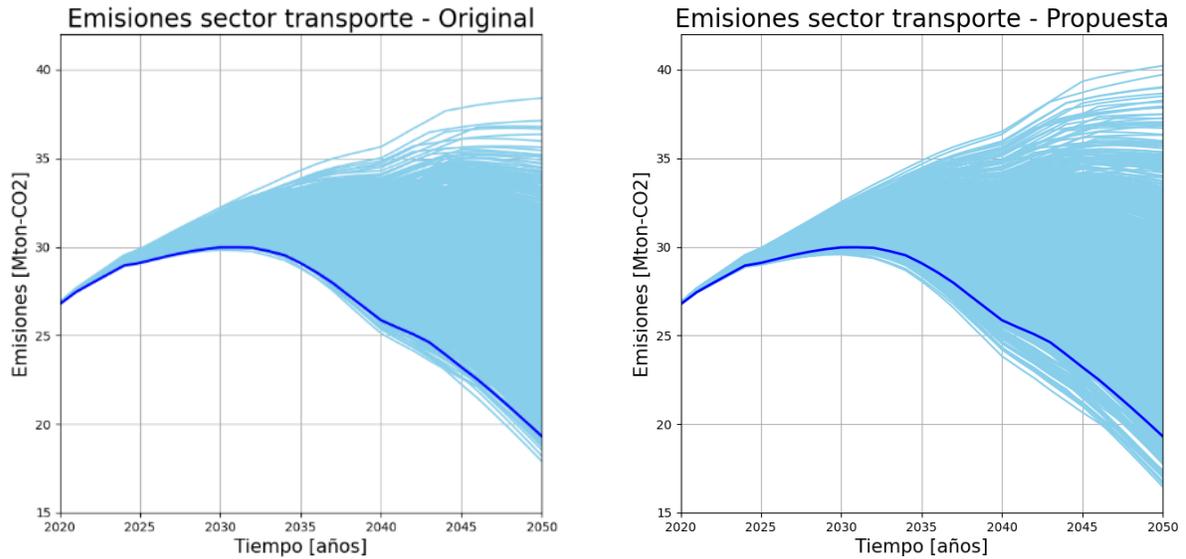


Figura 73: Trayectorias de emisiones sector transporte, modelo original y propuesto. Periodo 2020-2050.

En cuanto a las emisiones, destacar la presencia de una mayor cantidad de trayectorias proyectadas bajo la trayectoria de referencia, lo que va de la mano con un aumento de los límites superior e inferior.

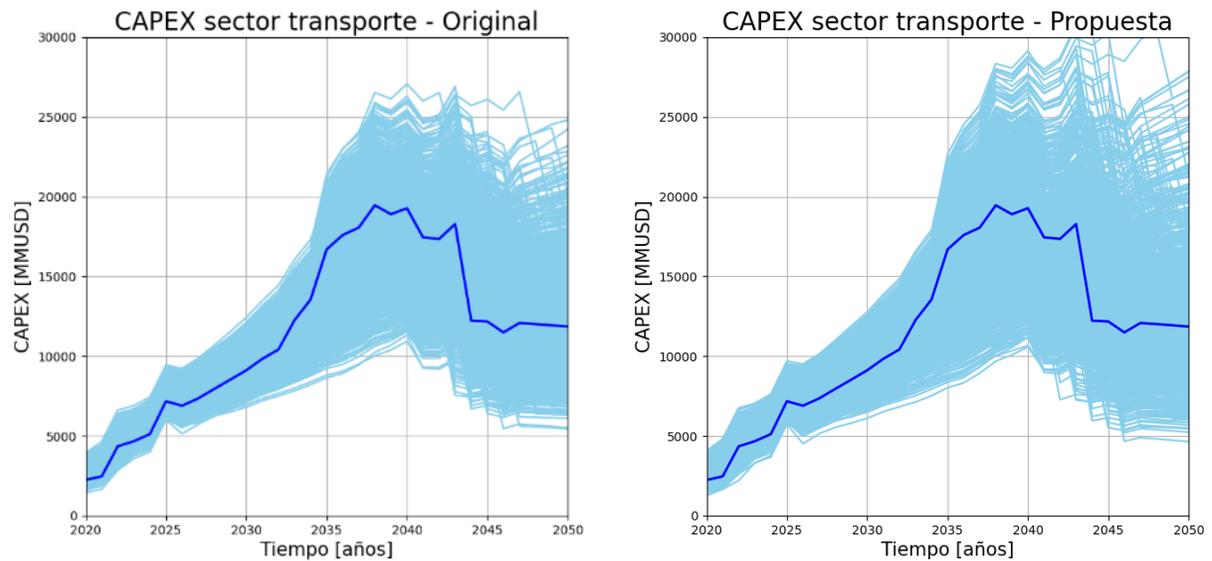


Figura 74: Trayectorias de CAPEX transporte, modelo original y propuesto. Periodo 2020-2050.

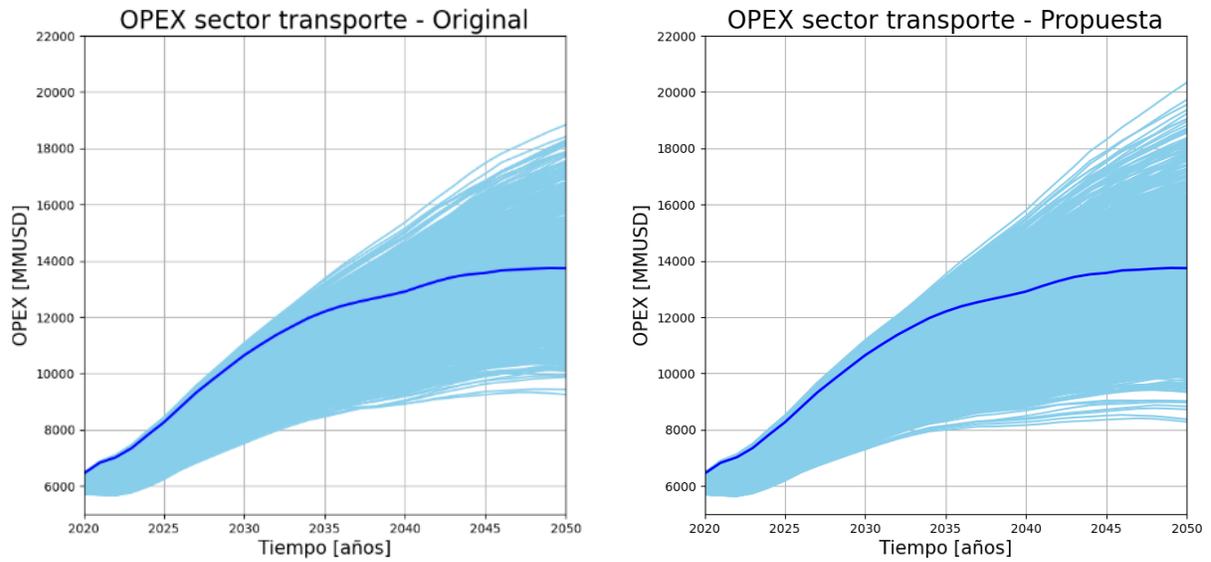


Figura 75: Trayectorias de OPEX sector transporte, modelo original y propuesto. Periodo 2020-2050.

Como se puede observar, en los indicadores económicos CAPEX y OPEX también se tiene un aumento del rango de incertidumbre, lo cual es posible percibir visualmente desde el año 2030.

## 4.2.6. Sector generación eléctrica

El sector generación eléctrica presenta múltiples parámetros con rango de incertidumbre, los cuales se presentan en sección Anexos, Tabla 13.

A partir del análisis de componentes principales (PCA), se obtiene la curva de varianza explicada por componente presente en Figura 76.

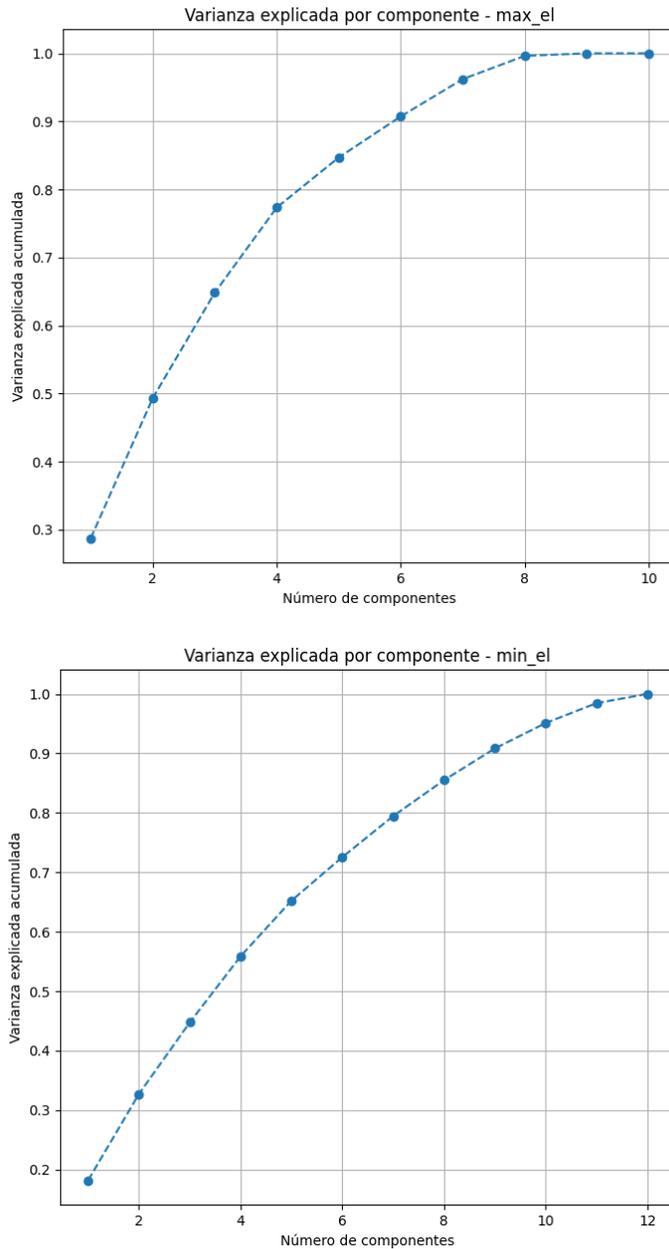


Figura 76: PCA y varianza explicada acumulada en sector generación eléctrica, para escenarios de valores máximos (arriba) y mínimos (abajo) en cuanto a métricas de desempeño.

Para un 80% de varianza explicada acumulada se tienen 4 componentes principales para los escenarios de valores máximos, y 7 componentes principales para los escenarios de valores mínimos. Estos valores fueron seleccionados considerando la disminución de la pendiente de la curva, desde donde la ganancia (en términos de información) de agregar una componente principal más, es poco atractiva, considerando el objetivo principal de reducir dimensionalidad de la matriz de datos.

Luego, en Figura 77 y en Figura 78 se presentan los pesos por parámetro para las componentes principales consideradas en los casos de escenario de valores máximos y de valores mínimos, respectivamente.

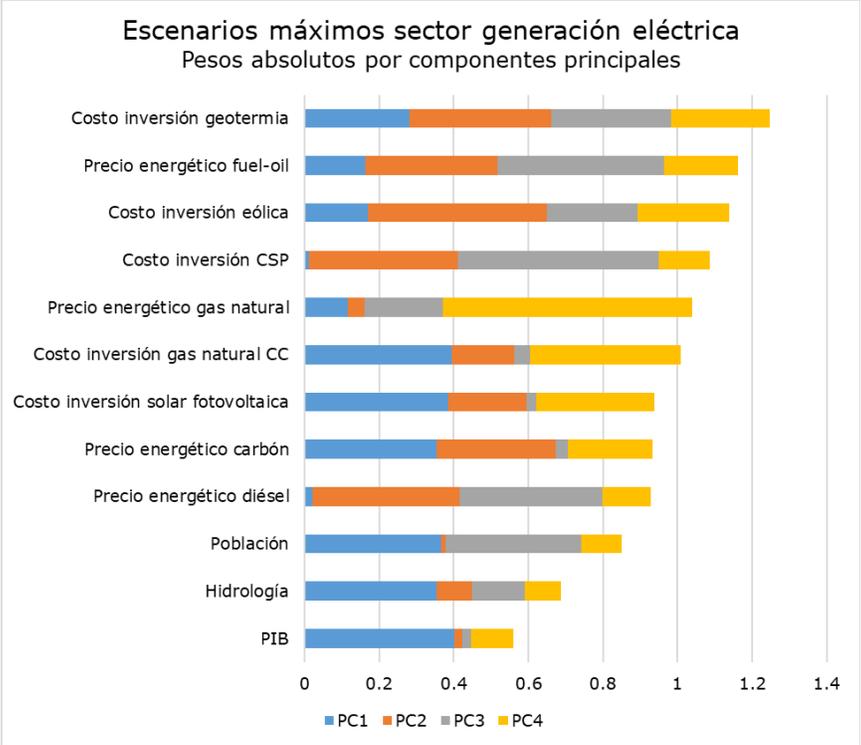


Figura 77: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores máximos del sector generación eléctrica.

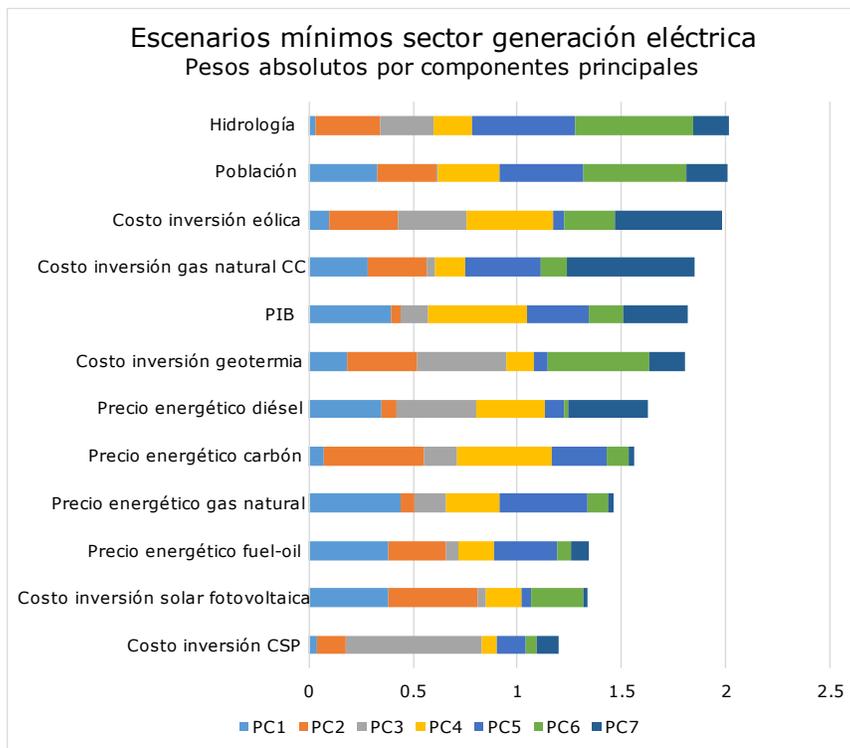


Figura 78: Peso absoluto de cada parámetro en las componentes principales, para escenarios de valores mínimos del sector generación eléctrica.

Siguiendo los pasos de la metodología propuesta, se escogen los parámetros que representan mayor peso en la variabilidad de los datos. Para la matriz de escenarios de valores máximos, se tienen:

- Precio energético diésel.
- Costo de inversión planta geotérmica.
- Costo de inversión planta eólica.

Mientras que para la matriz de escenarios de valores máximos, se tienen los siguientes parámetros:

- Hidrología.
- Costo de inversión planta eólica.
- PIB
- Población.

Luego, con los parámetros representativos se evalúa la cantidad de escenarios representativos, los que se obtienen de los centroides de las agrupaciones por *k-means*. En Figura 79 se presentan las curvas de análisis, de las cuales se obtiene una cantidad de 5 clústeres para escenarios máximos, y 14 para escenarios mínimos.

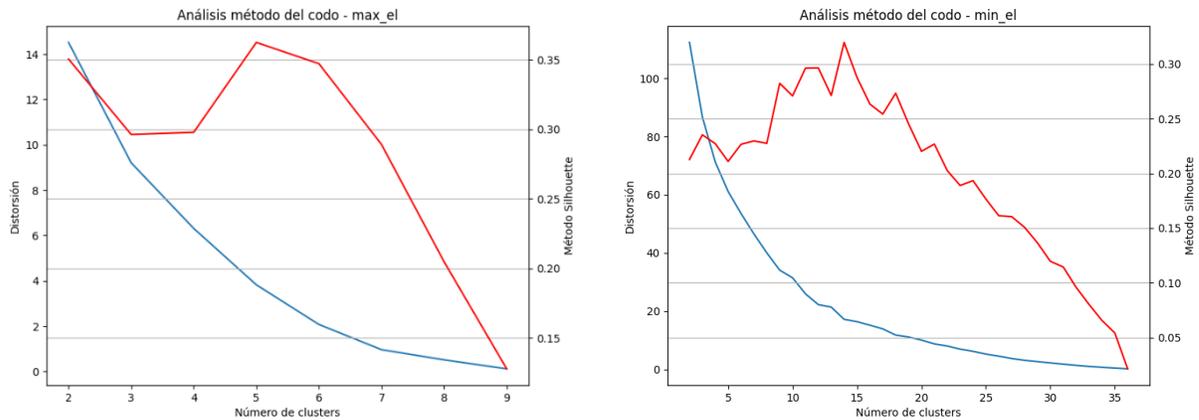


Figura 79: Análisis por método del codo para parámetros de mayor variabilidad de los escenarios de valores máximos (izquierda) y mínimos (derecha) del sector generación eléctrica, considerando las métricas de validación método Silhouette (curva roja) y distorsión (curva azul).

Finalmente, luego de la proyección de escenarios se tienen las métricas de desempeño para el sector generación eléctrica, presentes en Figura 80 y Figura 81.

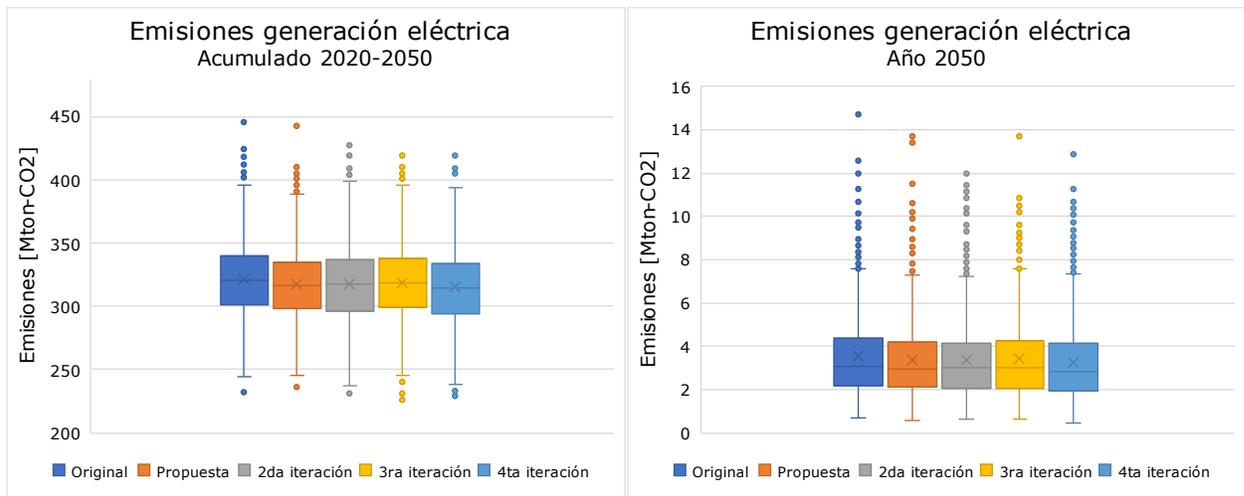


Figura 80: Gráficos de caja asociado a emisiones del sector generación eléctrica.

Se puede observar de los gráficos de emisiones acumuladas y del año 2050, una disminución del rango de incertidumbre en cada iteración de la metodología propuesta. Además, se tiene una presencia considerable de escenarios atípicos, representados por los puntos externos a cada gráfico de caja.

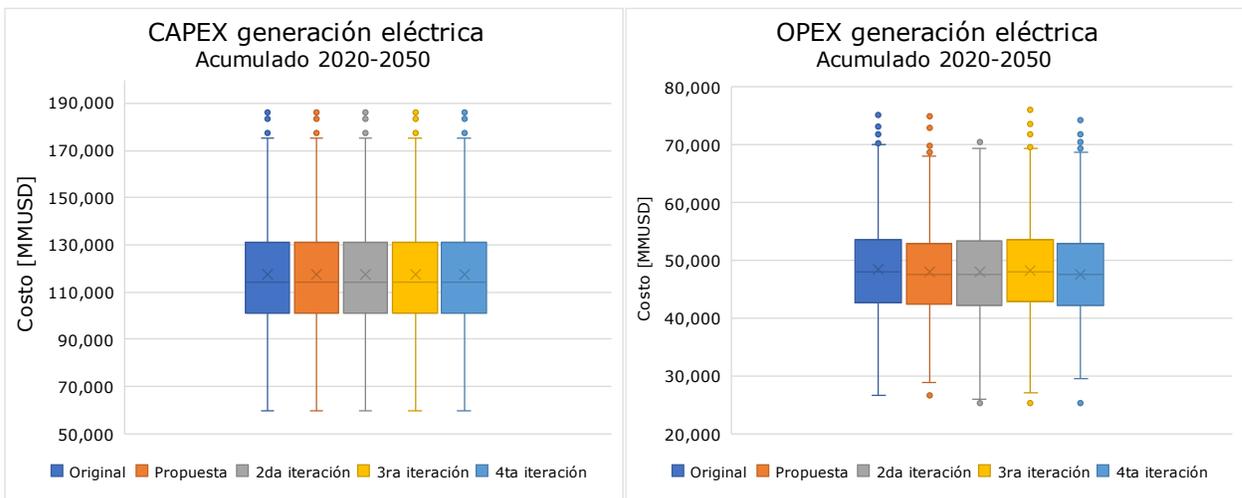


Figura 81: Gráficos de caja asociado a costos del sector generación eléctrica.

Para los gráficos de caja del indicador CAPEX, se tiene un valor constante entre el modelo original o de referencia, el modelo propuesto y sus iteraciones.

Mientras que para el indicador OPEX, se tiene una disminución del rango de incertidumbre al pasar del modelo de referencia al modelo propuesto, así también en la 4ta iteración. Por el contrario, la 2da y 3ra iteración presentan valores iguales o mayores de rango de incertidumbre, respecto del modelo original.

Finalmente, se presentan las trayectorias de las métricas de desempeño emisiones (Figura 82), CAPEX (Figura 83) y OPEX (Figura 84). De estas se observa un comportamiento idéntico de las métricas CAPEX y OPEX entre el modelo original y propuesto, mientras que la métrica de emisiones presenta una disminución del rango de incertidumbre, al aplicar la metodología propuesta.

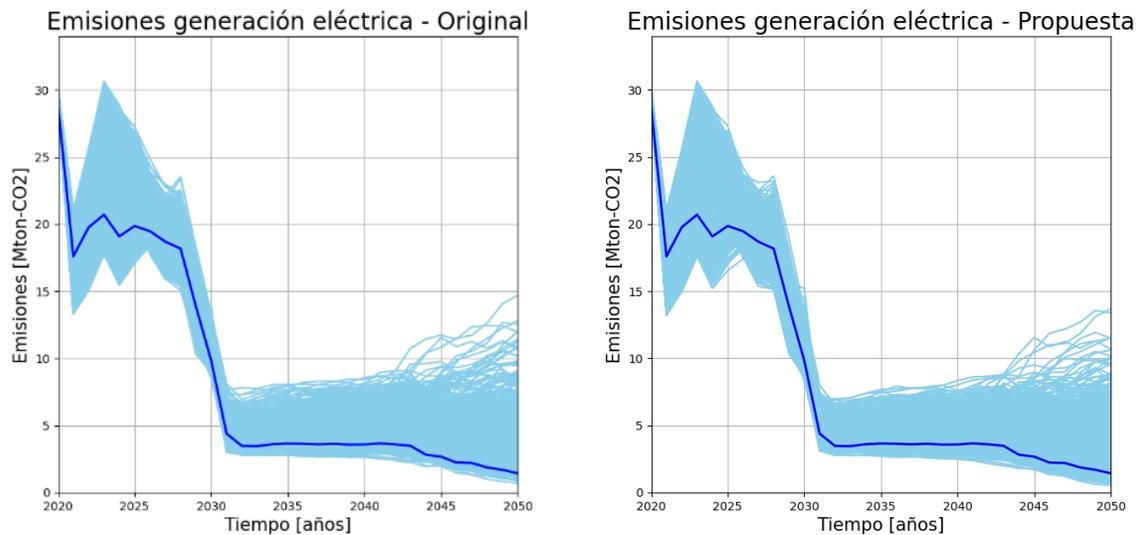


Figura 82: Trayectorias de emisiones sector generación eléctrica, modelo original y propuesto. Periodo 2020-2050.

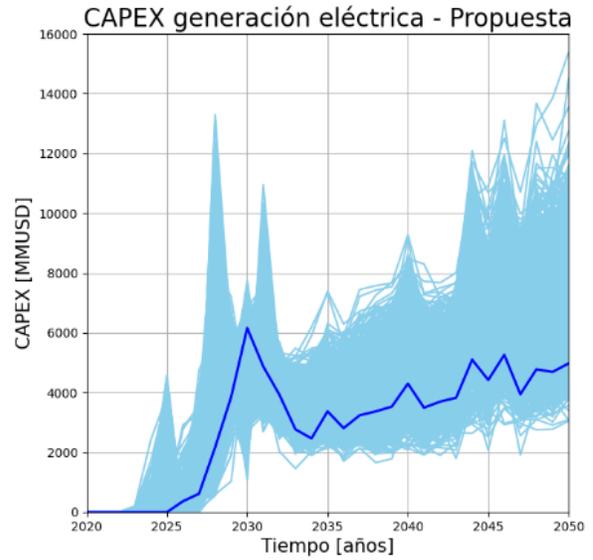
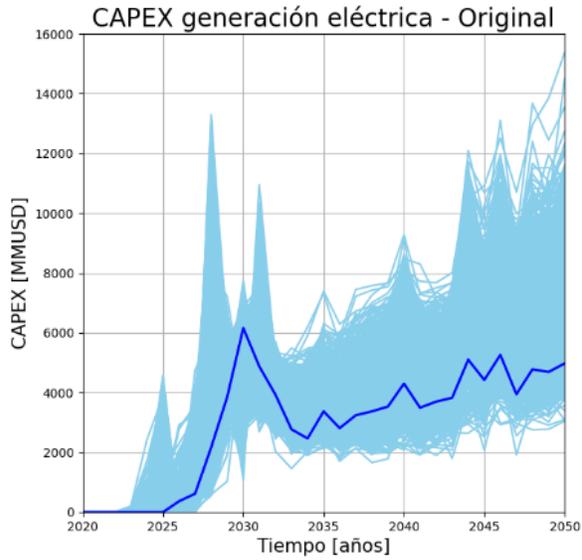


Figura 83: Trayectorias de CAPEX generación eléctrica, modelo original y propuesto. Periodo 2020-2050.

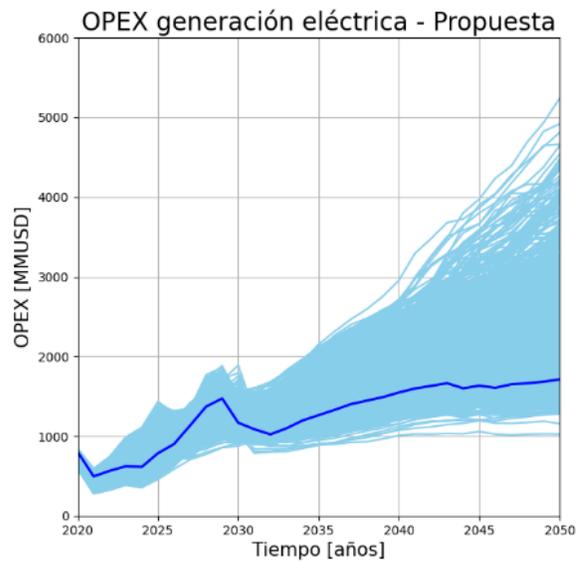
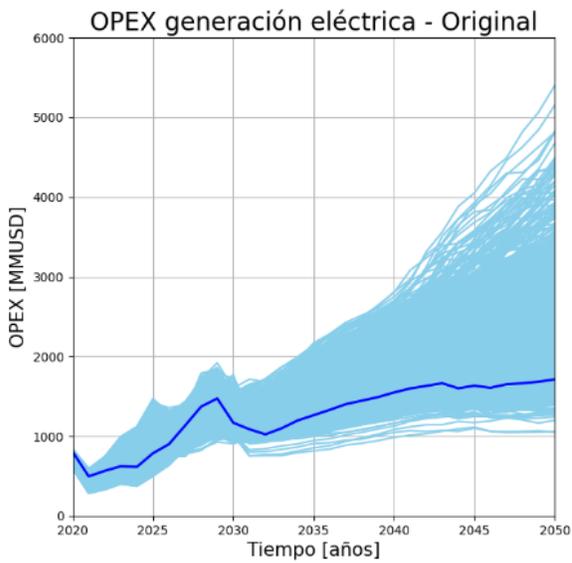


Figura 84: Trayectorias de OPEX sector generación eléctrica, modelo original y propuesto. Periodo 2020-2050.

# Capítulo 5

## Análisis de resultados

A continuación, se realizará un análisis de los resultados bajo la misma lógica en que fueron presentados, más dos tópicos de análisis crítico en contexto de modelos de proyección de emisiones GEI.

### 5.1. Modelo de referencia y metodología propuesta

La metodología propuesta presenta un aumento de los rangos de incertidumbre de las métricas de desempeño, tanto en sus valores máximos y mínimos, como en el rango total de cada métrica. Esta comparación se sustenta de lo presente en Tabla 6 y Tabla 7.

Tabla 6: Comparación de valores máximos y mínimos de las métricas de desempeño, para modelo de referencia u original y modelo propuesto.

<b>Métrica de desempeño</b>	<b>Máximo valor</b>		<b>Mínimo valor</b>	
	Original	Propuesto	Original	Propuesto
Emisiones 2050 [Mton-CO2]	66.3	74.2	38.4	34.4
Emisiones [Mton-CO2]	2,262	2,327	1,822	1,816
CAPEX [MMUSD]	699,446	788,787	343,211	323,910
OPEX [MMUSD]	1,143,599	1,163,299	732,105	706,372

Tabla 7: Comparación de rango de las métricas de desempeño y aumento porcentual, para modelo de referencia u original y modelo propuesto.

<b>Métrica de desempeño</b>	<b>Rango</b>		<b>Aumento rango</b>
	Original	Propuesto	
Emisiones 2050 [Mton-CO2]	27.8	39.8	43%
Emisiones [Mton-CO2]	439.8	510.4	16%
CAPEX [MMUSD]	356,235	464,877	30%
OPEX [MMUSD]	411,493	456,926	11%

A su vez, se tiene que la distribución de los escenarios cubre todo el rango, representado por los gráficos de caja y las trayectorias. Esto es importante ya que permite evaluar sensibilidades de los escenarios respecto a los valores que toman sus parámetros.

El promedio de las métricas presentes en los gráficos de caja es relativamente constante entre el modelo original, el modelo propuesto e iteraciones de este, por lo que se tiene una proyección de escenarios similar a la del modelo original.

Respecto a la presencia de escenarios críticos, definidos por la situación en que todos los sectores se encuentren en un valor máximo (o todos en un valor mínimo) de emisiones, en un mismo escenario.

Por lo tanto, un escenario crítico se da cuando el total de emisiones en un escenario es igual a la suma de los valores máximos de emisiones por sector, en los distintos escenarios. Así también para el caso de valores mínimos.

La metodología propuesta logra disminuir esta brecha, a su vez que aumenta los rangos de incertidumbre. No obstante, no logra un escenario crítico propiamente tal. En Figura 85, se presentan estos casos, incluyendo líneas de tendencia que demuestran un aumento de los valores máximos y una disminución de los valores mínimos.

Mientras que en Figura 86, se presenta la diferencia entre el escenario de valor máximo de emisiones de todos los sectores (denominado "total"), y la suma de los valores máximos de emisiones de cada sector en distintos escenarios (denominado "sectores"). Se observa una disminución de la brecha que representaría un escenario crítico en caso de llegar a valor nulo. Esta disminución ocurre tanto en valores máximos como mínimos de emisiones al año 2050.

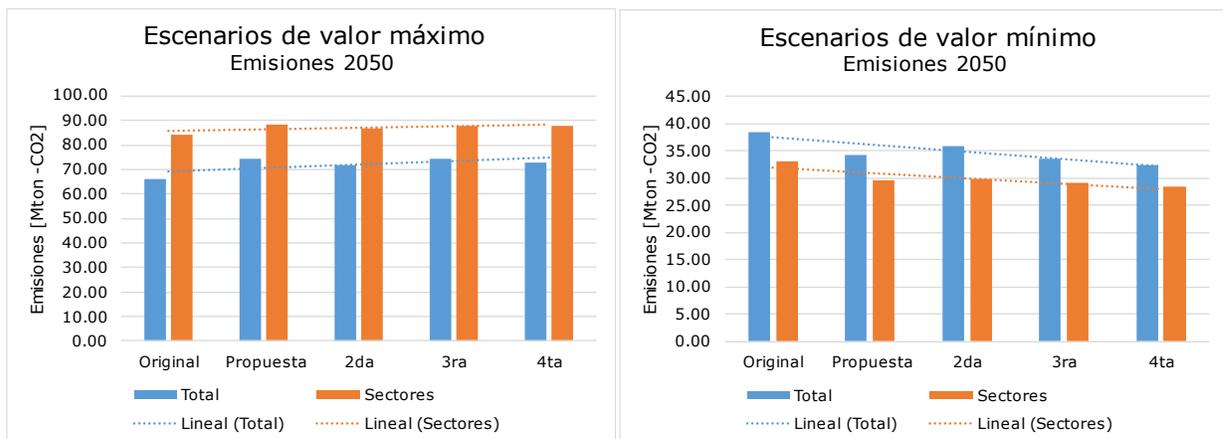


Figura 85: Comparación entre valores máximos (izquierda) y mínimos (derecha) para el total de emisiones del año 2050 en un mismo escenario, y la suma de sectores en distintos escenarios.

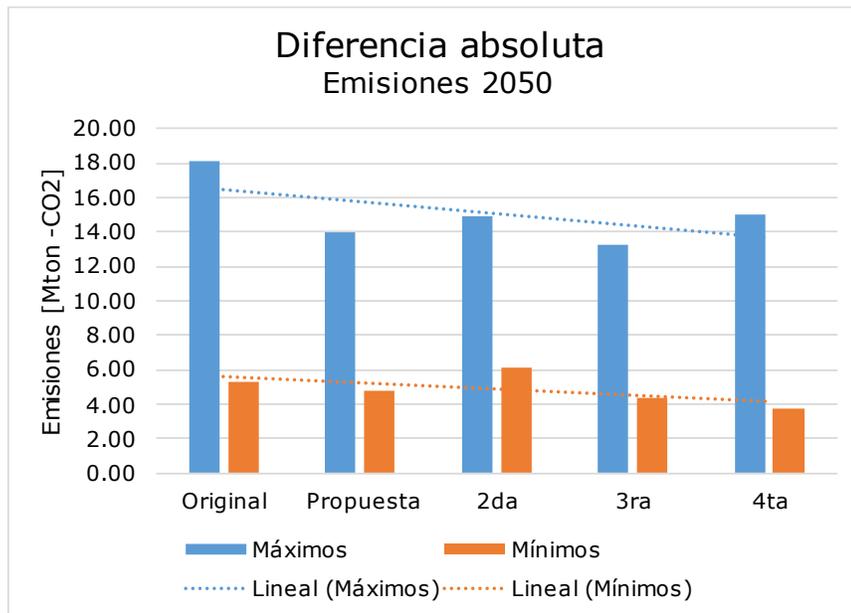


Figura 86: Diferencia entre máximos y mínimos totales y por suma de sectores, en los distintos modelos e iteraciones.

Es importante mencionar que el éxito de la metodología depende en gran parte de las decisiones que tome la persona que programe. Al utilizar aprendizaje no supervisado, no hay soluciones con etiquetas de bueno o malo.

Decisiones como el número de clústeres en función de las métricas de validación, el número de componentes principales según la curva de varianza explicada acumulada, o el número de proyecciones por cada sector o caso en estudio.

Luego, se recomienda ajustar el modelo en función de los resultados, realizando tantas simulaciones hasta que se logre identificar una solución que aumente el rango de incertidumbre de todas las métricas. En la presente memoria, esto no se realizó, de forma de poder contemplar el efecto de realizar varias iteraciones sin repetición o ajuste de sus decisiones.

Finalmente, si se utiliza la metodología para identificar escenarios de valores máximos y mínimos, sería conveniente aplicar la metodología propuesta hasta que el rango de incertidumbre no aumente, o los parámetros de los escenarios de valores máximos o mínimos converjan a un valor en específico (por ejemplo, con valor de la muestra de LHS igual a 1 o a 0).

## 5.2. Análisis por sectores

En esta sección, se presenta el análisis por sectores con enfoque en la comparación del modelo de referencia y el modelo propuesto, así como la identificación de parámetros por sector de mayor incidencia en la definición de un escenario como de valor máximo o mínimo en emisiones.

### 5.2.1. Sector comercio

De lo visto en los resultados del sector comercio, se tienen rangos de incertidumbre constantes en cuanto a emisiones acumuladas y en el año 2050. Para explicar esta situación, se identifica el promedio de los 10 escenarios de valor máximo y 10 de valor mínimo en emisiones del año 2050, para el modelo propuesto o de metodología aplicada. Estos promedios se presentan en Figura 87.

Estos promedios se presentan en Figura 87. Es posible observar que los parámetros PIB y elasticidad de comercio presentan promedios totalmente distintos según se trate de valores máximos o mínimos. Mientras que el parámetro población, tiene valores promedio cercano a 0.5, por lo que los escenarios no presentan tendencia clara hacia alguno de los extremos.

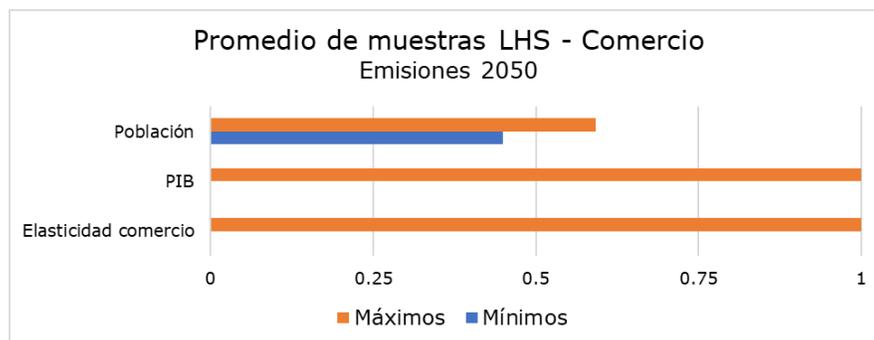


Figura 87: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector comercio, modelo propuesto.

Lo anterior se justifica con los gráficos de peso absoluto en las componentes principales, presentes en Figura 37, donde se tiene que el parámetro población representa una gran variabilidad dentro de ambos conjuntos de escenarios (de valores máximos y valores mínimos).

En cuanto a las métricas de CAPEX y OPEX, ambas presentan un leve crecimiento de su rango de incertidumbre (ver Figura 40).

## 5.2.2. Sector industria y minería

El sector industria y minería presenta un aumento del rango de incertidumbre en todas sus métricas de desempeño, tal como se presenta en sección Resultados, Figura 48 y Figura 49.

Para observar la incidencia de los parámetros en los escenarios de valores máximos y mínimos, en Figura 88 se presenta el promedio de muestras de LHS para los 10 escenarios de valores máximos de emisiones en el año 2050, y para 10 escenarios de valores mínimos.

Se puede observar que la principal diferencia de estos promedios entre máximos y mínimos se encuentra en los parámetros asociados a medidas de mitigación e intensidad del uso de energía. Dado que una muestra LHS de valor cercano a cero se relaciona a una trayectoria de valores menores a la trayectoria de referencia, entonces, se tiene que, para los parámetros de medidas de mitigación, valores cercanos a 1 implican mayor cumplimiento de los objetivos, y menor cumplimiento en valores cercanos a cero.

Además, los parámetros asociados a la producción, precios energéticos y costos de inversión también presentan polaridades en cuanto al valor que toman las muestras de LHS si se trata de escenarios de valores máximos o de valores mínimos.

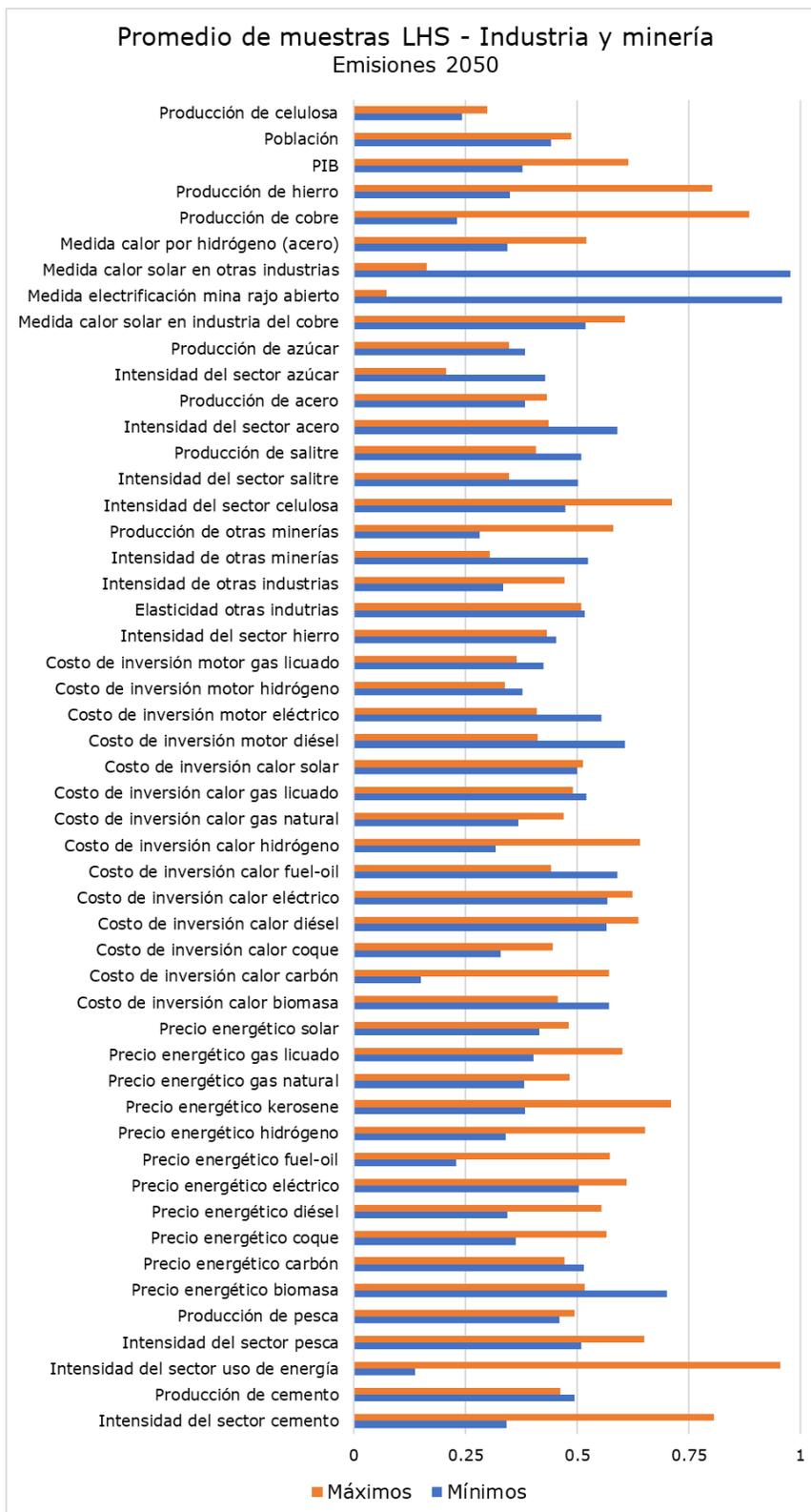


Figura 88: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector industria y minería, modelo propuesto.

### 5.2.3. Sector público

El sector público presenta el mismo análisis que el sector comercio. Se tienen parámetros totalmente polarizados (de muestras LHS iguales o muy cercanas a 0 en escenarios de valores mínimos, e iguales a 1 en escenarios de valores máximos). Estos parámetros son la intensidad del sector público y la población. Lo anterior se puede visualizar en Figura 89.

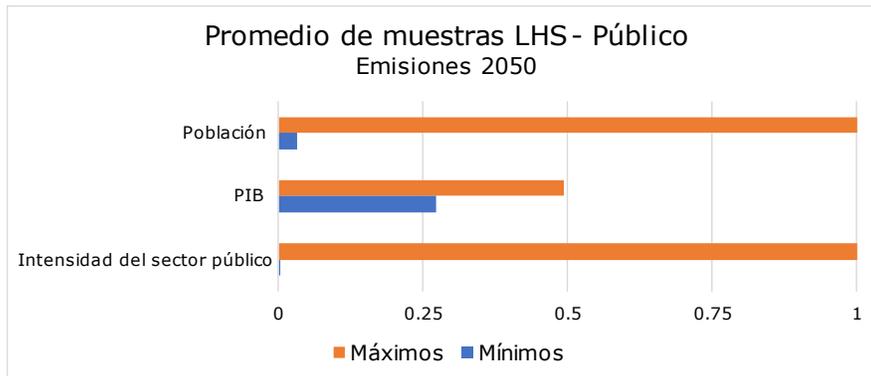


Figura 89: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector público, modelo propuesto.

Mientras que el sector público tiene como parámetros más incidentes la intensidad y la población, el sector comercio lo tiene en parámetros como elasticidad y PIB. Es decir, el sector comercio es altamente dependiente de variables económicas, mientras que el sector público depende de variables demográficas.

Además, mencionar que los promedios cercanos a 0 y a 1 concluyen que el sector presente los mismos valores máximos y mínimos en el modelo de referencia como en el modelo propuesto e iteraciones. Por tanto, es un sector que ya ha alcanzado su máximo rango de incertidumbre.

### 5.2.4. Sector residencial

Siguiendo la línea de los análisis anteriores, para el sector residencial se tiene un aumento del rango de incertidumbre en todas sus métricas de desempeño, cada vez que se aplica la metodología propuesta (es decir, modelo propuesto e iteraciones), según los visto en sección Resultados, principalmente en Figura 62 y Figura 63.

En cuanto a los valores que toman las muestras de LHS en los escenarios de valores máximos y mínimos de emisiones en el año 2050, se tienen parámetros con valores claramente opuestos, tales como las medidas de uso de sistemas solares térmicos (SST) en el agua caliente sanitaria (ACS), en casas y departamentos, la tasa de ocupación y el reacondicionamiento térmico. Lo anterior se presenta en Figura 90.

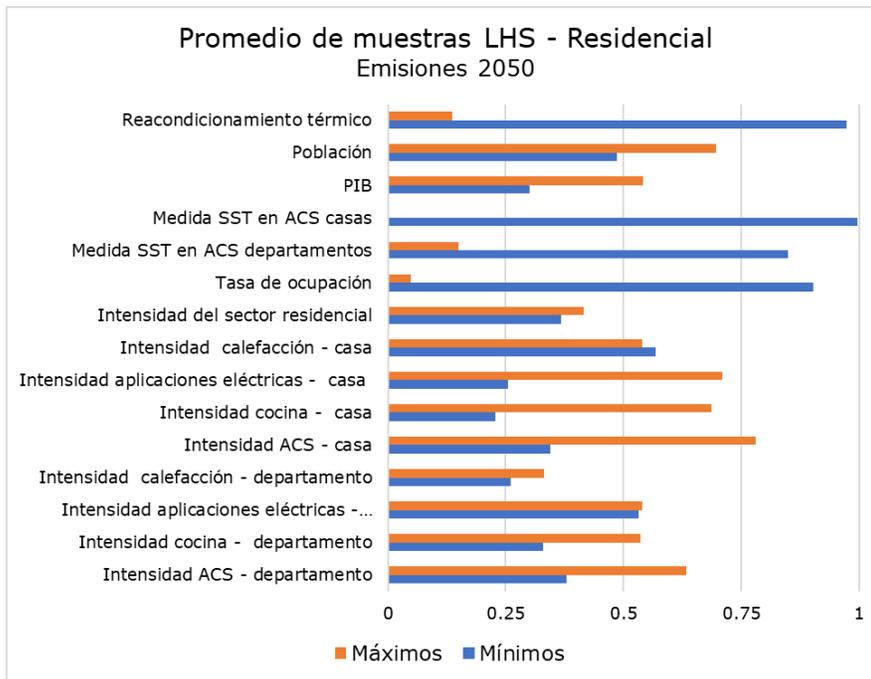


Figura 90: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector residencial, modelo propuesto.

### 5.2.5. Sector transporte

Respecto de los resultados del sector transporte, estos presentan un aumento del rango de incertidumbre en todas las métricas de desempeño, según lo visto en Figura 71 y Figura 72.

En cuanto a los parámetros presentes gráficamente en Figura 91, para los escenarios de valores máximos y mínimos en emisiones del año 2050, se tienen muestras de LHS de valores considerablemente opuestos en parámetros como medida de electromovilidad, medida de camiones a hidrógeno, PIB y cambios modales desde vehículo particular a tren y bus, y el ratio de ocupación de este último.

Por lo tanto, se desprende que los parámetros que más influyen en las emisiones del sector transporte corresponden a los parámetros asociados al cambio de participación de las tecnologías (de combustible fósil a electricidad o hidrógeno), al modo de transporte de pasajeros, y del parámetro PIB, fuertemente asociado a la posibilidad de invertir en las nuevas tecnologías y en el recambio de vehículos.



Figura 91: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector transporte, modelo propuesto.

## 5.2.6. Sector generación eléctrica

El sector generación eléctrica presenta un no cumplimiento respecto a las expectativas de aumento del rango de incertidumbre. Como se puede observar en Figura 80 y Figura 81, presentes en sección Resultados, se tiene una alta presencia de escenarios con valores atípicos (fuera del gráfico de caja), y una disminución del rango de emisiones acumuladas y del año 2050, una vez aplicada la metodología propuesta.

Recordar que este sector presenta un modelo de optimización de mínimo costo, recibiendo como entradas la demanda eléctrica asociada a cada uno de los escenarios y de los sectores del modelo. Además, sus parámetros con rango de incertidumbre definida mayor a cero corresponden a costos de inversión de plantas de generación, precio de los energéticos (combustibles), PIB, población y probabilidad de excedencia hidrológica.

Para identificar la situación y dar explicación a este no cumplimiento, se presenta en Figura 92 los valores promedio de las muestras de LHS, para los 10 escenarios de valores máximos y 10 de valores mínimos de emisiones en el año 2050, del modelo propuesto. Lo mismo se realiza en Figura 93 con información del modelo de referencia. Y finalmente, en Figura 94, se presenta la diferencia de los promedios entre el modelo original, menos el modelo propuesto, a fin de identificar los parámetros que presentan un mayor cambio al aplicar la metodología propuesta.

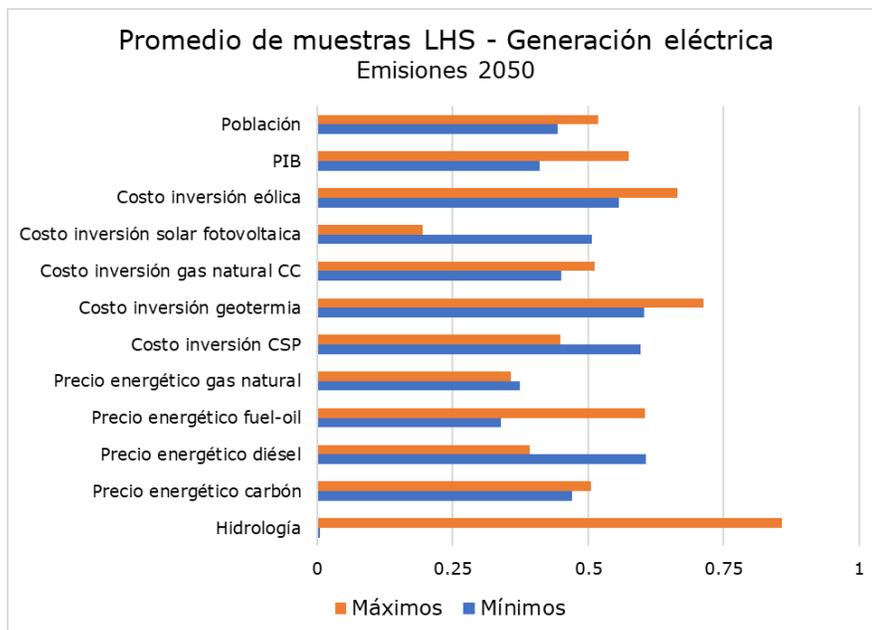


Figura 92: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector generación eléctrica, modelo propuesto.

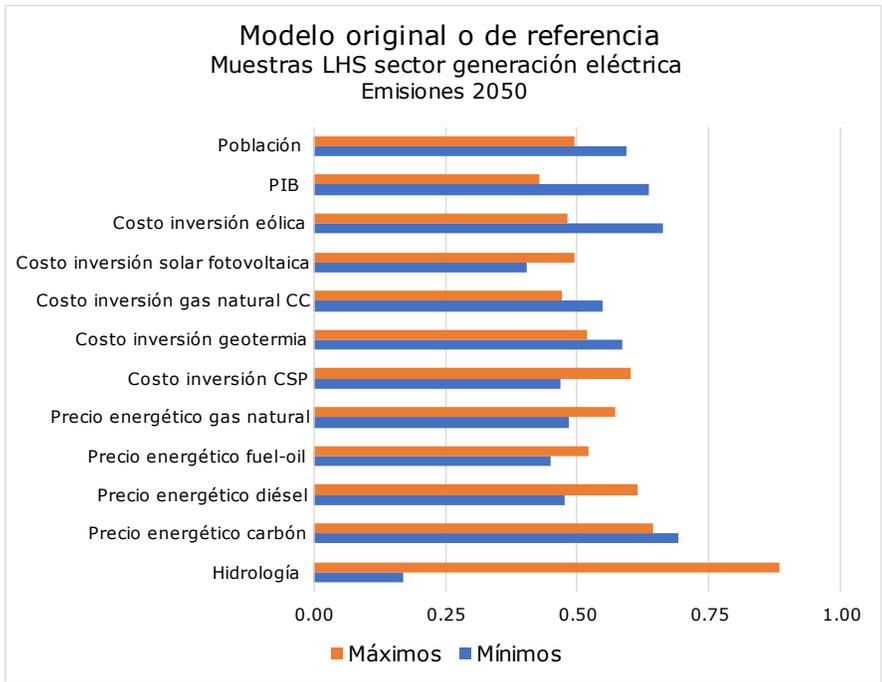


Figura 93: Promedio de muestras de LHS de 10 escenarios de valores máximos y 10 escenarios de valores mínimos de emisiones en el año 2050, para parámetros del sector generación eléctrica en modelo de referencia.

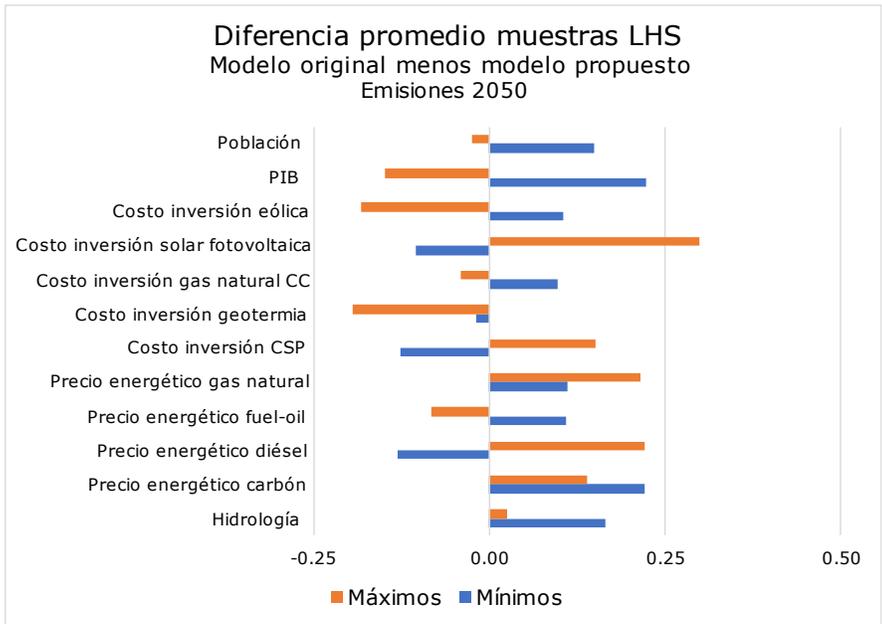


Figura 94: Diferencia entre modelo original y modelo propuesto, de los promedios de muestras de LHS para parámetros del sector generación eléctrica, en 10 escenarios de valores máximos y 10 de valores mínimos, en emisiones en el año 2050.

Recordando los pesos de cada parámetro en las componentes principales, presentes en Figura 77 y Figura 78. Se tiene que los parámetros que menos explican varianza en sus conjuntos de escenarios de valores máximos son:

- PIB.
- Hidrología.
- Población.
- Precio energético diésel.

Mientras que, para los escenarios de valores mínimos, se tienen los siguientes parámetros:

- Costo de inversión planta CSP.
- Costo de inversión planta solar fotovoltaica.
- Precio energético fuel oil.
- Precio energético gas natural.

Finalmente, si se combina la información entregada anteriormente, se tiene lo siguiente:

- En el modelo propuesto, los escenarios de valores máximos presentan un promedio mayor de las muestras LHS asociadas al PIB y un promedio menor en cuanto al precio energético diésel, mientras que la población e hidrología varían levemente. Además, en el mismo contexto, las muestras LHS de los parámetros PIB y precio energético diésel del modelo original están más cercanos a 1 que el modelo propuesto.

Por lo tanto, el modelo original presenta escenarios de valores máximos de emisiones con PIB más bajos y precio del diésel más altos que el modelo propuesto, lo cual, en la lógica de planificación del sector eléctrico, evitaría el ingreso de centrales de energía renovable ya que disminuiría la inversión (menor PIB) y se aumentaría el uso de centrales a carbón o a gas (mayor precio del diésel), presentando una mayor cantidad de emisiones de carbono equivalente.

- Por otra parte, los escenarios de valores mínimos en el modelo propuesto presentan un mayor costo de inversión en centrales CSP y fotovoltaica, y un menor precio de los energéticos gas natural y fuel-oil.

Por lo tanto, siguiendo nuevamente una lógica de planificación a mínimo costo, se priorizaría el uso de centrales térmicas y se invertiría menos en centrales de energías renovables (como la CSP y solar fotovoltaica), presentando escenarios de mayores emisiones que los escenarios de valores mínimos del modelo de referencia.

En conclusión, la metodología propuesta no logra acoplarse a una lógica de optimización de múltiples restricciones, donde los parámetros presentan una alta correlación en términos de sus resultados.

Esta correlación se debe a que el modelo de optimización considera distintas decisiones en el horizonte de evaluación: por ejemplo, el año de ingreso de una central solar fotovoltaica, dependiendo de las necesidades de aumento de capacidad, costo de oportunidad de utilizar las mismas centrales ya construidas (térmicas

principalmente), recuso hídrico que influye en las decisiones a mediano y largo plazo, entre otras situaciones.

En cambio, en los otros sectores, se presentan trayectorias fijas las cuales solo se modifican en función de la incertidumbre aplicada, y a su vez sus modelos presentan ecuaciones propias de cada aspecto del sector: por ejemplo, en transporte, lo que ocurre con los vehículos particulares no tiene mayor impacto en los vehículos de carga.

Por último, lo anterior se presenta como una oportunidad de mejora, ya que la metodología propuesta podría aplicarse de forma exclusiva a este sector, capturando cada una de las situaciones posibles e intentando proyectarlas en escenarios que expandan el rango de las métricas de desempeño.

### 5.3. Carbono neutralidad y tormenta perfecta

El modelo de referencia se caracteriza por definir muestras discretas, aleatorias y equidistantes de cada uno de los parámetros bajo incertidumbre. Esto condiciona a los parámetros a distribuir uniformemente dentro de los rangos de incertidumbre.

Dado que la esperanza de una variable que distribuye uniforme es el valor medio del rango, la combinación de parámetros que distribuyen uniformemente presentará una función de probabilidad más asociada a una distribución normal (ejemplo de esto es la distribución o histograma de las emisiones en el modelo original, presente en Figura 24).

Esta distribución normal no necesariamente sigue los escenarios de referencia, ya que los rangos de incertidumbre no son necesariamente simétricos, por lo que su esperanza se verá desplazada del valor de referencia.

Un parámetro sumamente significativo en el cumplimiento de la carbono neutralidad, es el cumplimiento de las medidas de mitigación: estas presentan un rango de incertidumbre entre 0 (mínimo 2050) y 1 (máximo 2050). Por lo tanto, las muestras de LHS de estos parámetros distribuyen uniformemente con esperanza en 0.5.

Entonces, si la metodología propuesta considera la proyección de muestras en condiciones críticas, su distribución probabilística no necesariamente es uniforme ni normal. Esto queda comprobado en las emisiones presentes en modelo propuesto iterado 2 veces, donde las emisiones máximas y mínimas presentan una mayor ocurrencia de forma considerable (ver Figura 27).

Por lo tanto, el modelo propuesto no es un buen indicador de la probabilidad de ocurrencia de estos escenarios críticos.

## 5.4. Aprendizaje no supervisado

Una de las características fundamentales del modelo propuesto es su dependencia al criterio del diseñador. Esto ya que se propone un método de aprendizaje no supervisado, como es el algoritmo de *clustering k-means*.

Luego, se puede hacer crítica y mejora de cada una de las decisiones tomadas durante el proceso, aun cuando sean fundamentadas por métricas de validación o indicadores estadísticos.

Algunas de las críticas posteriores al diseño y observación de resultados son:

- 1) Utilizar las 4 métricas de desempeño en una misma matriz: esto ya que se pudieron realizar distintas correlaciones entre CAPEX, OPEX y los niveles de emisiones. Por tanto, se estaban agregando escenarios de alto CAPEX y alto OPEX, pero que podían ser contrarios en sus niveles de emisiones, entregando señales contrarias a la representatividad de los escenarios.
- 2) Considerar sectores como comercio y público: dada la magnitud en sus métricas de desempeño, estos sectores pudieron haber sido combinados como un único sector, o incluso omitidos, a fin de disponer una mayor exploración de escenarios del resto de los sectores.

## 5.5. Uso de la metodología y oportunidades

Si bien la probabilidad de ocurrencia de los escenarios críticos representados por la metodología propuesta es baja, esto permite evaluar combinaciones que previamente no se consideraban por las características del muestreo LHS.

Por tanto, el algoritmo propuesto permite orientar la búsqueda hacia los escenarios límites en modelos donde se desconozcan los rangos máximos y mínimos.

Por ejemplo, puede ser útil en la exploración trayectorias óptimas donde la función objetivo maximice o minimice su área bajo la curva, o en modelos orientados a explorar parámetros de mayor sensibilidad o influencia.

# Capítulo 6

## Conclusiones

El trabajo desarrollado cumple el objetivo principal de expandir el rango de las métricas de desempeño consideradas, siendo estas las emisiones, CAPEX y OPEX en el periodo 2020-2050, y emisiones al año 2050.

Los principales resultados se resumen en un aumento del rango de incertidumbre en un 16% de las emisiones GEI acumuladas en el periodo 2020-2050, 30% en el costo de inversión acumulado, 11% en el costo operacional acumulado y un 43% en las emisiones GEI del año 2050.

Todo lo anterior utilizando una metodología que desarrolla nuevos escenarios a partir de la proyección de escenarios representativos de los niveles más altos y bajos en las métricas mencionadas, así como de la combinación entre los máximos y mínimos de los distintos sectores.

Se tiene además una metodología capaz de identificar variables (parámetros del modelo con rango de incertidumbre definida) de mayor incidencia en el resultado de emisiones deseado. Esto es útil para el tomador de decisiones.

Dado que la metodología propuesta utiliza el modelo de referencia como lo que se conoce por "caja negra", es decir, solo modifica sus entradas en función de sus salidas previas, entrega resultados no sesgados por el conocimiento previo del problema y su modelación.

Lo anterior permite conocer el rango de exploración con mínimo número de exploraciones, siendo entonces lo desarrollado de gran utilidad en la búsqueda de escenarios de borde o de valores máximos y mínimos.

Se tiene una crítica del proceso de diseño de la metodología propuesta, permitiendo mejoras en cuanto al algoritmo. Y se reconocen ciertas utilidades del algoritmo en otros modelos de proyección u exploración de escenarios.

Por lo tanto, se tiene como trabajo futuro el desarrollo y evaluación de la metodología propuesta en cuanto al nivel de detalle con el que trabaja, por ejemplo, considerando nuevas métricas de desempeño, o métricas anuales en vez de acumuladas. A su vez, es posible indagar respecto al comportamiento de la metodología propuesta en sectores específicos (solo en transporte, por ejemplo), y en modelos de optimización.

# Bibliografía

- [1] Gobierno de Chile, «Contribución Nacional Tentativa de Chile (INDC) para el Acuerdo Climático París 2015,» 2015.
- [2] Gobierno de Chile, «Contribución Determinada a Nivel Nacional (NDC) de Chile, Actualización 2020,» 2020.
- [3] Ministerio de Energía, «Carbono neutralidad en el sector energía - Proyección de consumo energético nacional 2020,» 2020.
- [4] Banco Interamericano de Desarrollo, «Opciones para lograr la carbono-neutralidad en Chile: una evaluación bajo incertidumbre,» 2021.
- [5] GAMS Development Corp., «GAMS: System Overview,» [En línea]. Available: <https://www.gams.com/products/gams/gams-language/>. [Último acceso: 2021].
- [6] H. Heitsch y W. Roemisch, «Scenario Tree Generation for Multi-stage Stochastic Programs,» de *Stochastic Optimization Methods in Finance and Energy*, vol. 163, Springer, 2011, pp. 313-342.
- [7] J. Amat, «Ciencia de Datos,» [En línea]. Available: <https://www.cienciadedatos.net/documentos/py19-pca-python.html>. [Último acceso: 2021].
- [8] J. Mallea, «Evaluación del uso de modelos probabilísticos en la estimación de las trayectorias de emisiones de gases de efecto invernadero en Chile,» 2016.
- [9] J. Silvente, L. G. Papageorgiou y V. Dua, «Scenario tree reduction for optimisation under uncertainty using sensitivity analysis,» de *Computers & Chemical Engineering*, vol. 125, 2019, pp. 449-459.
- [10] J. E. González, «Simulación de Monte Carlo para escenarios de alta penetración de generación solar y eólica.,» 2020.
- [11] M. Sánchez, «Dimensionamiento óptimo de sistemas de almacenamiento basado en técnicas de reducción de escenarios.,» 2018.

- [12] X. S. e. al, «Smart Sampling for Reduced and Representative Power System Scenario,» *IEEE Open Access Journal of Power and Energy* , vol. 8, nº .293-302, 2021.
- [13] H. H. e. al, «Security-Based Active Demand Response Strategy Considering Uncertainties in Power Systems,» *IEEE Access*, vol. 5, pp. 16953-16962, 2017.
- [14] Q. W. e. al., «Modified Algorithms for Fast Construction of Optimal Latin-Hypercube Design,» *IEEE Access*, vol. 8, pp. 191644-191658, 2020.
- [15] B. M.-I. a. S. T. S. Madadi, «A Data Clustering Based Probabilistic Power Flow Method for AC/VSC-MTDC,» *IEEE Systems Journal*, vol. 13, nº 4, pp. 4324-4334, 2019.

# Anexo

## Parámetros con incertidumbre

Tabla 8: Parámetros de sector comercio bajo incertidumbre.

Sector	Parámetro	Min 2050	Max 2050
<b>Comercio</b>	Elasticidad comercio	1	1.2
<b>Total</b>	PIB	0	1.7621
	Población	0	1.6354

Tabla 9: Parámetros de sector industria y minería bajo incertidumbre.

Sector	Parámetro(s)	Min 2050	Max 2050	Parámetro(s)	Min 2050	Max 2050
<b>Industria y minería</b>	Intensidad y producción del sector cemento	0.9	1.1	Intensidad del sector cobre en uso de energía	0.9	1.77
	Intensidad y producción del sector pesca	0.9	1.1	Intensidad del sector hierro	0.9	1.1
	Elasticidad en otras industrias.	0.9	1.05	Intensidad en otras industrias.	0.9	1.1
	Precio energético biomasa, carbón, coque, diésel, electricidad, fueloil, hidrógeno, kerosene, gas natural, gas y solar.	0.9	1.1	Costo de inversión calefacción con biomasa, carbón, coque, diésel, electricidad, fueloil, hidrógeno, kerosene, gas natural, gas o solar.	0.9	1.1
	Costo inversión motor diésel,	0.9	1.1	Intensidad del sector celulosa	0.9	1.1

	electricidad, hidrógeno o gas					
	Intensidad en otras minerías	0.9	1.1	Producción en otras minerías	0.3	1
	Intensidad y producción del sector salitre	0.9	1.1	Rendimiento del sector acero	0.9	1.1
	Intensidad y producción del sector azúcar	0.9	1.1	Producción del sector acero	0.9	1.05
	Medida calor solar en industria del cobre	0	1	Medida electrificación mina rajo abierto	0	1
	Medida calor solar en otras industrias	0	1	Medida calor por hidrógeno en industria del acero	0	1
	Producción del sector cobre	0	9.50	Producción del sector hierro	0	2.25
	Producción del sector celulosa	0	1			
<b>Total</b>	Población	0	1.6354	PIB	0	1.7621

Tabla 10: Parámetros de sector público bajo incertidumbre.

<b>Sector</b>	<b>Parámetro</b>	<b>Min 2050</b>	<b>Max 2050</b>
<b>Público</b>	Elasticidad del sector público	1	1.1
<b>Total</b>	PIB	0	1.7621
	Población	0	1.6354

Tabla 11: Parámetros de sector residencial bajo incertidumbre.

<b>Sector</b>	<b>Parámetro(s)</b>	<b>Min 2050</b>	<b>Max 2050</b>
<b>Residencial</b>	Rendimiento departamento en ACS, cocina, aparatos eléctricos y calefacción	0.9	1.1
	Rendimiento casa en ACS, cocina, aparatos eléctricos y calefacción	0.9	1.1
	Rendimiento del sector residencial	0.95	1.05
	Ratio de ocupación residencial	0.8	1
	Medida SST en ACS departamentos	0	1
	Medida SST en ACS casas	0	1
	Reacondicionamiento térmico hogar	0	24.54
<b>Total</b>	PIB	0	1.7621
	Población	0	1.6354

*SST: sistemas solares térmicos.*

*ACS: agua caliente sanitaria.*

Tabla 12: Parámetros de sector transporte bajo incertidumbre.

<b>Sector</b>	<b>Parámetro(s)</b>	<b>Min 2050</b>	<b>Max 2050</b>	<b>Parámetro(s)</b>	<b>Min 2050</b>	<b>Max 2050</b>
<b>Transporte</b>	Medida electromovilidad	0	1	Medida aviones hidrógeno	0	1
	Medida camiones hidrógeno	0	1	Nivel de actividad buses, vehículos particulares, taxis y camiones	0.8	1.2
	Costo de inversión avión kerosene	0.9	1	Costo de inversión avión hidrógeno	1	1.1
	Costo de inversión bus diésel	0.9	1.1	Costo de inversión bus eléctrico	0.8	1.1
	Costo de inversión vehículo eléctrico, vehículo híbrido	0.9	1.1	Rendimiento avión hidrógeno y avión kerosene	0.9	1.1
	Rendimiento tren pkm (pasajero-km)	0.95	1.05	Rendimiento bus diésel, bus eléctrico	0.95	1.05
	Rendimiento vehículo diésel, vehículo gasolina	0.95	1.05	Rendimiento vehículo eléctrico	0.7	1.3
	Rendimiento vehículo híbrido (parte gasolina)	0.8	1.2	Rendimiento vehículo híbrido (parte eléctrica)	0.7	1.3
	Rendimiento taxi eléctrico, taxi gasolina, tren eléctrico, camión diésel, camión hidrógeno	0.95	1.05	Costo inversión cambio de vehículo particular a: bicicleta, bus,	0.8	1.2

				teletrabajo, tren.		
	Reparto modal vehículo particular a bus, a bicicleta, a teletrabajo o a tren	0.95	1.05	Ratio de ocupación bus, vehículo, taxi y tren	0.95	1.05
	PKM aviación	0.9	1.1	Bus-km	0.9	1.1
	Saturación de PKM de aviación.	1	1.2	TKM tren de carga.	0.9	1.1
	Costo inversión camiones hidrógeno.	1	1.2	Distancia de viaje vehículos particulares.	0.9	1.1
	PKM trenes	0.95	1.05			
<b>Total</b>	Población	0	1.6354	PIB	0	1.7621

Tabla 13: Parámetros de sector generación eléctrica bajo incertidumbre.

<b>Sector</b>	<b>Parámetro(s)</b>	<b>Min 2050</b>	<b>Max 2050</b>
<b>Generación eléctrica</b>	Hidrología	0	2
	Precio energético carbón, diésel, fuel-oil y gas natural	0	1
	Costo de inversión centrales CSP, geotérmica, gas natural CC, solar fotovoltaica y eólica.	0	10000
<b>Total</b>	PIB	0	1.7621
	Población	0	1.6354