



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**Mejora en la estimación de consumo de
demanda base de los usuarios aplicado a los
sistemas eléctricos de potencia.**

TESIS

Que para obtener el título de

Ingeniero Eléctrico Electrónico

P R E S E N T A

Luis Joaquín Hernandez Núñez

DIRECTOR DE TESIS

Dr. Néstor González Cabrera



Ciudad Universitaria, Cd. Mx., 2022

AGRADECIMIENTO

Al finalizar esta maravillosa etapa de mi vida quiero extender un profundo agradecimiento, a quienes hicieron posible este sueño, aquellos que junto a mí caminaron en todo momento y siempre fueron mi inspiración, apoyo y fortaleza. Esta mención en especial es para mis padres, mis hermanos y mi pareja. Muchas gracias a ustedes por demostrarme que «El verdadero amor no es otra cosa que el deseo inevitable de ayudar al otro para que este se supere.»

Así mismo, me gustaría expresar mi profunda gratitud al asesor de mi tesis el Dr. Néstor González Cabrera por sus valiosas y constructivas sugerencias durante la planificación y desarrollo de este trabajo de investigación. Su disposición a dar su tiempo tan generosamente ha sido muy apreciada.

Finalmente me gustaría extender mi agradecimiento a las instituciones que me ayudaron en el desarrollo de mi carrera, la Facultad de Ingeniería junto a cada docente, quienes con su apoyo y enseñanzas constituyen la base de mi vida profesional, y a la Dirección General de Asuntos del Personal Académico (DGAPA) por otorgarme la beca de investigación bajo el proyecto PAPIIT IA105021 denominada Integración de la respuesta de la demanda a los eléctricos considerando modelos de aprendizaje.

RESUMEN

Con el incremento en el uso de las tecnologías derivado del crecimiento exponencial de la población, se ha incrementado la necesidad del uso de la energía eléctrica para la vida cotidiana, dicho de otra manera, la demanda de electricidad ha experimentado un evidente crecimiento en los últimos años, lo que conlleva, a tener que mejorar la gestión de los sistemas eléctricos de potencia para poder satisfacer la necesidad de los usuarios de estar conectados a la electricidad. En este aspecto, es necesario contar con métodos predictivos del consumo eléctrico como lo es la Carga de Referencia del Cliente o CBL por sus siglas en inglés, que calcula el patrón de consumo de los usuarios. Para ello, se hace uso de la herramienta de Aprendizaje Automático que cuenta con un conjunto de modelos estadísticos que mejora la estimación de los datos, reduciendo su error.

Específicamente, en este trabajo se utiliza el modelo de Regresión de Proceso Gaussiano (RPG) basado en un modelo de regresión lineal, el cual es comparado con el modelo de Media Móvil Integrada Auto-Regresiva (ARIMA) y así, validar la eficiencia de la estimación del CBL. De igual manera, se hace uso de indicadores de rendimiento que justifican la eficiencia del modelo propuesto, siendo estos el Error Cuadrático Medio (MSE), el Error Absoluto Medio (MAE) y el Error Cuadrado de la Media Raíz (RMSE).

La relación CBL-ARIMA presenta un MSE de $6.8288 * 10^{-8}\%$ mientras que el MSE de la relación CBL-RPG es de $5.4198 * 10^{-8}\%$ para el caso uno, justificando así la superioridad del modelo propuesto en este documento. Aplicando este enfoque a un sistema de 24 nodos para validar su eficiencia. Los resultados obtenidos de CBL-RPG permiten una predicción precisa de la generación de energía con sus respectivos costos de operación, obteniendo costos proporcionales a la demanda y evitando la generación innecesaria.

ABSTRACT

With the increase in the use of technologies derived from the exponential growth of the population, the need for the use of electrical energy for daily life has increased, in other words, the demand for electricity has experienced an evident growth in recent years, which entails, to have to improve the management of electrical power systems to be able to satisfy the need of users to be connected to electricity. In this aspect, it is necessary to have predictive methods of electricity consumption such as the Customer Baseline Load or CBL, which calculates the consumption pattern of users. To do this, the Machine Learning tool is used, which has a set of statistical models that improves the estimation of the data, reducing its error.

Specifically, this paper uses the Gaussian Process Regression (RPG) model based on a linear regression model, which is compared with the Auto-Regressive Integrated Moving Average (ARIMA) model and thus validate the efficiency of the CBL estimate. Similarly, performance indicators are used to justify the efficiency of the proposed model, these being the Mean Square Error (MSE), the Mean Absolute Error (MAE) and the Square Error of the Root Mean (RMSE).

The CBL-ARIMA ratio presents MSE of $6.8288 * 10^{-8}\%$ while the MSE of the CBL-RPG ratio is $5.4198 * 10^{-8}\%$ for case one, thus justifying the superiority of the model proposed in this document. Applying this approach to a 24-node system to validate its efficiency. The results obtained from CBL-RPG allow an accurate prediction of power generation with their respective operating costs, obtaining costs proportional to demand and avoiding unnecessary generation.

SIMBOLOGÍA Y ABREVIACIONES

OS	Operador del Sistema
OM	Operador del Mercado
REI	Red Eléctrica Inteligente
DR	Respuesta a la Demanda
OPF	Flujos de Potencia Óptimos
CBL	Carga de Referencia del Cliente
RBTS	Sistema de prueba Roy Billinton
RPG	Regresión de Proceso Gaussiano
ARIMA	Media Móvil Integrada Auto-Regresiva
AR	Polinomio Auto-Regresivo
MA	Polinomio de Media Móvil
RLG	Regresión Lineal Generalizada
TREE	Árbol de regresiones
ACR	Árboles de Clasificación y Regresión
SVM	Soporte de regresión de la máquina vectorial
PG	Proceso Gaussiano
CEM	Construcción y Evaluación de Modelos
RN	Red Neuronal
MSE	Error Cuadrático Medio
RMSE	Error cuadrado de la media raíz
R-Cuadrado	Coefficiente de determinación
MAE	Error Medio Absoluto
PYME	Pequeñas y Medianas Empresas
IEEE	Instituto de Ingenieros Eléctricos y Electrónicos

PBP	Programas basados en precios
PBI	Programas basados en incentivos
TDU	Tiempo de uso
PTR	Precios en Tiempo Real
PC	Precio Crítico
PEDR	Programa de emergencia de DR
CI	Servicios de carga interrumpibles
CDC	Control Directo de Carga
OC	Oferta de la carga
CNB	Clasificador de Naive Bayes
y_t	Respuesta de un modelo de serie temporal lineal
ε_t	Proceso estocástico
x_t	Covariables exógenas
$c, \beta, a_1, a_w, b_1, b_v$	Parámetros de la ecuación polinómica
w, v	Orden del modelo ARIMA
$\phi(L)$	AR no estacional estable de grado v
D	Grado de integración nocional
$\Phi(L)$	AR estacional multiplicativo
s	Grado del polinomio de diferenciación estacional
D_s	Grado de integración estacional
θ	MA no fuera de tiempo invertible grado w
Υ	MA estacional multiplicativo
ε	Serie de innovaciones aleatorias
f	Funciones de densidad
T	Parámetro de dispersión

$b(\varphi)$	Función de identidad
$h(Y, \tau), T(y), A(\varphi), d(\tau)$	Funciones de distribución
φ	Parámetro canónico
Var	Varianza de la distribución
η	Predictor lineal
γ	Parámetros desconocidos
X	Variables independientes
I_G	Impureza de Gini
f_i	Probabilidad de que pertenezca a una clase
i	Clases
m	Clase a la que pertenece
K	Kernel
α_1, α_2	Vectores de soporte seleccionados
σ	Anchura del Kernel
β_0, β_1	Determinados valores
o_j, r_j	Conjunto de entrenamiento
ω	Datos de entrada
o^T	Vector de entrada transpuesto
r	Funciones de respuesta base
$z(o)$	Función media
$k(o, o')$	Función de covarianza
u	Distancia euclidiana entre o_i y o_j
s_l	Escala de longitud característica
s_f	Desviación estándar de la señal
N	Paso de ajuste

$z(\lambda)$	Neurona artificial
λ	Función de activación
u_0, ψ_0, ι_0	Parámetros iniciales
A_1, A_2, A_3	Conjuntos activos
η_1, η_2, η_3	Vector de parámetros
L_1, L_2, L_3	Probabilidad de registro
t	Períodos de tiempo a analizar
$C_{i,t}^j$	Oferta de energía del i -ésimo generador en el bus i para el periodo t
$Pg_{i,t}^j$	Generación de energía del bus i para el periodo t
Ng	Número total de generadores en el sistema
Nb	Total de buses en el sistema
$Pd_{i,t}^j$	Demanda de energía activa en el bus i para el período t
$Qg_{i,t}^j$	Potencia reactiva generada en el bus i para el período t
$Qd_{i,t}^j$	Demanda de potencia reactiva en el bus i para el período t
Θ	Ángulo nodal
$V_{i,t}^j$	Magnitud del voltaje en el nodo para el período t
$V_{i,t}^{j,min}, V_{i,t}^{j,max}$	Límites de voltaje nodal mínimo y máximo
$P_{im,j}$	Flujo de energía de la línea a través de los nodos im
n	Número total independiente de elementos en todo el sistema eléctrico
sr	Varianza de ruido
x_i	Observaciones

\bar{X} Media de las observaciones

Nu Número de observaciones

ÍNDICE GENERAL

AGRADECIMIENTO	I
RESUMEN	II
ABSTRACT	III
SIMBOLOGÍA Y ABREVIACIONES	IV
ÍNDICE GENERAL	IX
ÍNDICE DE TABLAS	XI
ÍNDICE DE FIGURAS	XII
CAPÍTULO 1. INTRODUCCIÓN.	1
1.1. La nueva industria eléctrica.....	1
1.2. Revisión bibliográfica	1
1.3. Motivación y justificación.....	3
1.4. Hipótesis.....	4
1.5. Objetivos generales y específicos.....	4
1.5.1. Objetivo general	4
1.5.2. Objetivos específicos.....	4
1.6. Novedad de la propuesta de tesis.....	4
1.7. Estructura de la tesis.....	5
CAPÍTULO 2. MODELOS ESTADÍSTICOS AVANZADOS	6
2.1. Aprendizaje automático	6
2.1.1. Aprendizaje supervisado	6
2.1.2. Aprendizaje sin supervisión	7
2.2. Modelo autorregresivo integrado de media móvil.....	7
2.3. Regresión Lineal Generalizada.....	8
2.4. Árboles de regresiones	10
2.5. Soporte de regresión de la máquina vectorial.....	10
2.6. Regresión de proceso gaussiano.....	11
2.7. Construcción y evaluación de modelos	13
2.8. Red neuronal.....	13

2.9. Comparación entre las técnicas de estimación	14
2.9.1. Validaciones	14
2.9.2. Rendimiento	15
2.9.3. Comportamiento gráfico.....	15
2.9.4. Trabajo de referencia IEEE	16
<i>CAPÍTULO 3. RELACIÓN CARGA DE REFERENCIA DEL CLIENTE – REGRESIÓN DE PROCESO GAUSSIANO.</i>	19
3.1. Concepto de la Carga de Referencia del Cliente	19
3.2. Respuesta a la demanda.....	19
3.3. Regresión de Proceso Gaussiano aplicado a la Carga de Referencia del Cliente.....	21
3.4. Flujos de Potencia Óptimos.....	21
<i>CAPÍTULO 4. METODOLOGÍA PROPUESTA</i>	23
4.1. Adquisición de datos.	23
4.2. Estimación del CBL mediante los modelos estadísticos RPG y ARIMA.	23
4.3. Análisis de las estimaciones obtenidas mediante los modelos estadísticos.....	24
4.4. Evaluación de cada período del CBL – RPG en los Flujos de Potencia Óptimos.....	24
4.5. Análisis de los resultados obtenidos mediante el modelo OPF	25
4.6. Metodología propuesta para la estimación del CBL	25
4.6.1. Trabajo de referencia IEEE	26
<i>CAPÍTULO 5. CASO DE ESTUDIO</i>	29
5.1. Datos Generales del sistema eléctrico de potencia (Sistema de 24 nodos)	29
5.2. Caso 1.- CBL a nivel sistema	30
5.3. Caso 2.- CBL a nivel región (varios nodos agrupados).....	37
5.4. Caso 3.- CBL a nivel nodo	46
<i>CAPÍTULO 6. CONCLUSIONES Y TRABAJOS FUTUROS</i>	56
6.1. Conclusiones	56
6.2. Trabajos futuros.....	56
<i>BIBLIOGRAFÍA</i>	58
<i>APÉNDICE</i>	62

ÍNDICE DE TABLAS

Tabla 1. Kernels populares que se emplean con el SVM	11
Tabla 2. Algunos tipos de Kernel para RPG	12
Tabla 3. Estadísticas del modelo.	15
Tabla 4. Error Cuadrático Medio.....	18
Tabla 5. Resumen de las iteraciones para el CBL basado en RPG	21
Tabla 6. Indicadores de rendimiento a nivel sistema.....	31
Tabla 7. Indicadores de rendimiento por región.....	40
Tabla 8. Indicadores de rendimiento por nodo	49
Tabla 9. Datos de los nodos.....	62
Tabla 10. Datos de los generadores	63
Tabla 11. Datos de las líneas	64
Tabla 12. Datos de demanda total del sistema	65
Tabla 13. Datos de demanda en la región 1.....	66
Tabla 14. Datos de demanda en la región 2.....	67
Tabla 15. Datos de demanda en el nodo 1.....	68
Tabla 16. Datos de demanda en el nodo 2.....	69
Tabla 17. Datos de demanda en el nodo 7.....	70

ÍNDICE DE FIGURAS

Figura 1. Clasificación de las técnicas del aprendizaje automático.....	6
Figura 2. Representación de una neurona artificial.	14
Figura 3. Serie Temporal de la demanda para 96 horas previas.....	16
Figura 4. Resultados de los pronósticos para 24 horas.....	17
Figura 5. Acercamiento de la hora 13:00 a 17:00.....	17
Figura 6. Clasificación del programa de respuesta a la demanda.....	20
Figura 7. Metodología propuesta de esta tesis.....	26
Figura 8. RBTS de seis nodos.	27
Figura 9. Comparación entre estimaciones de los modelos CNB, ARIMA y los datos reales.	27
Figura 10. Sistema IEEE de 24 buses RTS.	29
Figura 11. Estimación de la demanda a nivel sistema.	30
Figura 12. Coeficiente de determinación de la estimación del modelo RPG a nivel sistema	31
Figura 13. Potencia activa estimada con los datos del modelo RPG a nivel sistema.....	32
Figura 14. Potencia activa estimada con los datos reales a nivel sistemas.....	32
Figura 15. Relación energía-generadores (RPG) a nivel sistema.....	33
Figura 16. Relación energía-generadores (datos reales) a nivel sistema.....	33
Figura 17. Costos operativos estimados a nivel sistema.....	34
Figura 18. Costos marginales estimado a nivel sistema con RPG.....	34
Figura 19. Costos marginales estimado a nivel sistema con los datos reales.....	35
Figura 20. Relación costos-nodos (RPG) a nivel sistema.....	35
Figura 21. Relación costos-nodos (datos reales) a nivel sistema.....	36
Figura 22. RTS de 24 nodos dividido por regiones.....	37
Figura 23. Estimación de la demanda de la región 1.....	38
Figura 24. Estimación de la demanda de la región 1 entre la hora 7:00 y la hora 13:00.....	38
Figura 25. Estimación de la demanda de la región 2.....	39
Figura 26. Estimación de la demanda de la región 2 entre la hora 13:00 y la hora 21:00.....	39
Figura 27. Coeficiente de determinación de la estimación del modelo RPG para la región 1.....	40
Figura 28. Coeficiente de determinación de la estimación del modelo RPG para la región 2.....	41
Figura 29. Potencia activa estimada con los datos del modelo RPG a nivel regional.....	41
Figura 30. Potencia activa estimada con los datos reales a nivel regional.....	42
Figura 31. Relación energía-generadores (RPG) a nivel regional.....	42
Figura 32. Relación energía-generadores (datos reales) a nivel regional.....	43
Figura 33. Costos operativos estimados a nivel región.....	43
Figura 34. Costos marginales estimado a nivel región con RPG.....	44
Figura 35. Costos marginales estimado a nivel región con los datos reales.....	44
Figura 36. Relación costos-nodos (RPG) a nivel región.....	45
Figura 37. Relación costos-nodos (datos reales) a nivel región.....	45
Figura 38. Estimación de la demanda del nodo 1.....	46

Figura 39. Estimación de la demanda del nodo 1 entre la hora 1:00 y la hora 12:00.....	47
Figura 40. Estimación de la demanda del nodo 2.....	47
Figura 41. Estimación de la demanda del nodo 2 entre la hora 1:00 y la hora 12:00.....	48
Figura 42. Estimación de la demanda del nodo 7.....	48
Figura 43. Estimación de la demanda del nodo 7 entre la hora 1:00 y la hora 15:00.....	49
Figura 44. Coeficiente de determinación de la estimación del modelo RPG para el nodo 1.	50
Figura 45. Coeficiente de determinación de la estimación del modelo RPG para el nodo 2	50
Figura 46. Coeficiente de determinación de la estimación del modelo RPG para el nodo 7	51
Figura 47. Potencia activa estimada con los datos del modelo RPG a nivel nodal	51
Figura 48. Potencia activa estimada con los datos reales a nivel nodal	52
Figura 49. Relación energía-generadores (RPG) a nivel nodal	52
Figura 50. Relación energía-generadores (datos reales) a nivel nodal	53
Figura 51. Costos operativos estimados a nivel nodal	53
Figura 52. Costos marginales estimado a nivel nodal con RPG.....	54
Figura 53. Costos marginales estimado a nivel nodal con los datos reales	54
Figura 54. Relación costos-nodos (RPG) a nivel nodal.....	55
Figura 55. Relación costos-nodos (datos reales) a nivel nodal.....	55

CAPÍTULO 1. INTRODUCCIÓN.

1.1. La nueva industria eléctrica

Desde inicios del siglo pasado hasta hace pocas décadas los sistemas eléctricos eran operados de forma verticalmente integrada, esto quiere decir, que una sola empresa era la encargada de generar, transmitir y distribuir la energía eléctrica. Esta a su vez era controlada en distintos niveles jerárquicos para poder desarrollar una operación controlada, continua y con un nivel de confiabilidad mínimo. Lo que a su vez presentaba apagones continuos debido a la falta de inversión para mejorar la infraestructura eléctrica.

La nueva industria eléctrica opera bajo el paradigma del programa de Redes Eléctricas Inteligentes (REI), debido a que la finalidad de este programa es mejorar el sistema eléctrico, siendo este eficiente, seguro, flexible, resiliente, de calidad, confiable y sustentable. La arquitectura de una REI debe estar integrada por los clientes y un proveedor de servicios, quien supervisa los productos ofrecidos por terceros.

Otros sistemas que son parte de las REI son la operación (que gestiona el flujo de electricidad de los distintos dominios de la red) y el mercado (que coordina a los que participan en el comercio de servicios energéticos dentro de la REI).

Dentro de las operaciones que realizan las REI se encuentran la integración de vehículos eléctricos, almacenamientos de energía, energías renovables, la mejora de la eficiencia energética y que los consumidores tengan opciones de quienes puedan suministrar energía.

1.2. Revisión bibliográfica

La gestión tradicional por el lado de la demanda ha sido superada por la evolución y mejora de la gestión en la adquisición de patrones de consumo, lo que representa un cambio radical en la calidad de vida de las personas, brindándoles la oportunidad de acceder fácilmente a la información y controlar su consumo eléctrico.

La demanda de electricidad aumenta constantemente cada año debido a las continuas mejoras en la calidad de vida y las condiciones climáticas extremas de frío y calor [1]. Requerir una estimación precisa del consumo de electricidad por usuario a nivel de distribución y potencia. Los programas de Respuesta a la Demanda (DR por sus siglas en inglés) [2] se consideran un modelo prometedor para mejorar la confiabilidad de los sistemas eléctricos, porque incentivan a los consumidores a cambiar o reducir su consumo de electricidad durante ciertos períodos en respuesta a las señales técnico-económicas emitidas por el proveedor

del servicio eléctrico [3]. Por otro lado, los estudios han demostrado que la exclusión de la demanda de energía por variaciones de precios en el mercado de la electricidad es inaceptable desde un punto de vista económico, ya que ha plagado al mercado con muchas ineficiencias [4].

Uno de los subprogramas de DR es la Carga de Referencia del Cliente (CBL por sus siglas en inglés), que puede ser particularmente desafiante para los usuarios de DR, ya que la carga eléctrica diaria puede variar de forma aleatoria y significativa [5]. Uno de los principales aspectos en la estimación del CBL es su implementación flexible en los usuarios residenciales y su aplicación en los programas de recuperación energética ante fallas repentinas en la red eléctrica, porque involucra los intereses de múltiples partes, incluidas las empresas de servicios públicos y los usuarios finales, que se revisa en [6-7]. Además, el uso de la DR basada en precios está respaldado por el amplio despliegue de pruebas y el despliegue de contadores inteligentes que permiten la cuantificación de la capacidad de respuesta del consumidor a las señales de precios a través de una estimación de referencia, como se puede evaluar en [8].

En [9] los autores presentan un modelo CBL para medir y evaluar los criterios de desempeño de los programas de gestión de carga, aunque la solución se presenta para redes inteligentes. Las series de tiempo de los patrones de demanda de electricidad son la fuente clave de información sobre el comportamiento del consumo de los clientes, como se presenta en [10], dicha información se puede usar en programas de DR para la predicción del consumo de los usuarios, lo cual es relevante tanto para la planificación como para el funcionamiento del sistema energético [11].

El equilibrio de la demanda y la generación de energía utilizando las últimas tecnologías de gestión de carga se considera un requisito inmediato para los programas de DR, así como para mejorar el rendimiento de las redes de distribución de electricidad [12], por lo que se han realizado muchos intentos en todo el mundo para diseñar programas eficaces de recuperación ante desastres, que favorezcan a clientes reales en la industria comercial y residencial [13]. En este contexto, los programas que utilizan métodos estadísticos de aprendizaje automático pueden realizar una mejor gestión de datos, lo que representa una mayor precisión en la estimación y el modelado del consumo [14].

Las principales técnicas para estimar el consumo de electricidad en los sectores doméstico e industrial se basan en el método de mínimos cuadrados [15], por ejemplo, la Media Móvil Integrada Auto-regresiva (ARIMA por sus siglas en inglés) metódicamente.

Obtener una estimación precisa de los patrones de consumo de electricidad de los usuarios permitirá reducir el consumo de energía desde el punto de vista del usuario final, así como la certeza operativa y una

reducción de la generación para el operador de la red. Incluso permitirá al usuario participar en programas de gestión de la eficiencia energética [16].

En combinación con lo anterior, los Flujos de Potencia Óptimos (OPF por sus siglas en inglés) son una herramienta para evaluar las contribuciones de los usuarios en los programas de DR, facilitando la adición de variables definidas por el usuario, costos y restricciones al problema OPF estándar. El objetivo principal de OPF es minimizar el costo de generación para una demanda de carga dada [17].

Finalmente, de acuerdo con la necesidad de una estimación precisa del CBL, este trabajo propone la estimación del consumo de demanda del usuario de la CBL a nivel nodal de una manera más precisa a través de la Regresión de Proceso Gaussiano (RPG), con el fin de obtener una óptima respuesta para un horizonte de 24 horas, además, esta misma estimación brindará señales de consumo de los usuarios al área del operador del sistema para planificar un escenario óptimo en la toma de decisiones para la generación de energía. Para analizar la efectividad del modelo CBL-RPG se consideran dos escenarios: primero, se compara con el modelo ARIMA utilizado en [18] y segundo, la comparación con un día de la semana. Además, se considera que el OPF aprecia los precios nodales y la generación como señales utilizando el sistema de energía RBTS de 6 nodos como referencia.

1.3. Motivación y justificación

Con el continuo crecimiento de la población, se incrementa la necesidad de consumir energía para satisfacer la calidad de vida de los usuarios. En los últimos años se han estado produciendo cambios irregulares en los hábitos de consumo por parte de los clientes. Por lo consiguiente, los operadores de las plantas generadoras se han visto en la necesidad de realizar estimaciones más precisas para cumplir con la demanda de energía de la manera más económica y confiable posible.

El conjunto de modelos estadísticos de la herramienta Aprendizaje automático (Machine Learning) permiten mejorar el proceso de estimación, al aplicar algoritmos capaces de aprender a partir de la información de patrones de datos, mejorando la eficiencia de los procesos a través de la automatización.

Considerando lo anterior, en este trabajo se profundiza sobre la estimación del consumo de energía eléctrica, para reducir la incertidumbre que pudiera causar el comportamiento de la carga, y estudiar las consecuencias que causa en los sistemas eléctricos de potencia para prevenir y eliminar estos efectos en la planeación de la generación de corto plazo.

1.4. Hipótesis

Al realizar mejores estimaciones de la demanda eléctrica, se pueden tener beneficios económicos para los usuarios finales y brindar certeza en la estimación de la demanda a los operadores del sistema. A su vez, se reduce la generación excedente o insuficiente durante algunos periodos de tiempo.

1.5. Objetivos generales y específicos

1.5.1. Objetivo general

Hacer uso del aprendizaje automático para mejorar la estimación de datos de demanda eléctrica, a través del modelo de Regresión de Proceso Gaussiano (RPG), asegurando que se obtendrán estimaciones de costos de generación, generación de potencia activa y costos marginales más precisas en comparación de un modelo tradicional.

1.5.2. Objetivos específicos

- Realizar estimaciones de demanda eléctrica empleando la herramienta *Aprendizaje Automático*, garantizando que el modelo de Regresión de Proceso Gaussiano (modelo propuesto) es el mejor para realizar estimaciones.
- Obtener indicadores de rendimiento que demuestren que el modelo propuesto es el mejor modelo para estimar datos de demanda eléctrica. Haciendo énfasis en que los indicadores deben presentar un valor de error inferior al 3% e inferior al valor del modelo tradicional.
- Analizar el comportamiento de las estimaciones por medio de su ajuste en la curva del quinto día, de tal manera, que se observe que la estimación hecha con el modelo propuesto es más preciso y, por consiguiente, sea el mejor modelo para estimar datos de demanda.
- Implementar las estimaciones obtenidas con el modelo propuesto en un sistema eléctrico de potencia de 24 nodos y analizar su comportamiento en tres niveles (sistema, regional y nodal).
- Comparar la estimación técnico-económica hecha mediante el modelo propuesto y los datos del quinto día, demostrando que las estimaciones de la generación de potencia activa, costos operativos y costos marginales de potencia activa se aproximan significativamente a los datos del quinto día.

1.6. Novedad de la propuesta de tesis

Al realizar estimaciones con la herramienta Aprendizaje Automático (Statistics and Machine Learning Toolbox) de Matlab, se obtienen patrones de datos más precisos debido a la capacidad del software por mantenerse en constante aprendizaje. Esta herramienta al tener modelos estadísticos más exactos que los

modelos tradicionales, se puede tener una mayor certeza y confiabilidad en la respuesta estimada. Los algoritmos y propuestas desarrolladas en esta tesis fueron presentadas en:

1.- Hernández-Núñez J., Campos-Romero J., González-Cabrera N. (2022) Improving the Customer Baseline Technique Based on a Learning Machine Applied to a Power System. In: Flores Rodríguez K.L., Ramos Alvarado R., Barati M., Segovia Tagle V., Velázquez González R.S. (eds) Recent Trends in Sustainable Engineering. ICASAT 2021. Lecture Notes in Networks and Systems, vol 297. Springer, Cham. https://doi.org/10.1007/978-3-030-82064-0_4

2.- J. Campos, L. Hernández and N. González, “Técnicas estadísticas para el pronóstico de la demanda mediante Matlab”, 2021 IEEE Reunión Internacional de Verano de Potencia, Acapulco, Guerrero, 2021.ta sea semejante a la realidad.

1.7. Estructura de la tesis

Este proyecto está constituido por seis capítulos, siendo el capítulo uno la introducción donde se mencionan los antecedentes, la descripción de la propuesta y la composición del trabajo. En el segundo capítulo se describe lo que es el Aprendizaje Automático, los modelos estadísticos que componen esta herramienta y los indicadores que se utilizan para comparar su funcionamiento. En el tercer capítulo se presenta la aplicación del OPF con base en la relación entre el modelo estadístico propuesto RPG y el consumo base de los clientes, a través del DR. En el cuarto capítulo se hace mención de la metodología planteada para realizar el proyecto, desde la obtención de los datos, pasando por el método de estimación hasta la evaluación del sistema una vez que fue aplicado el OPF. En el quinto capítulo se aplica el método propuesto a tres casos de estudio, uno de ellos a nivel sistema (24 nodos), a nivel regional (12 nodos) y a nivel nodal. Finalmente, en el sexto capítulo se hace mención de las conclusiones con base en los resultados adquiridos tanto del modelo propuesto como del modelo tradicional en comparación con los datos esperados y se da una propuesta de cómo puede ser mejorado este proyecto, en caso de ser retomado a futuro.

CAPÍTULO 2. MODELOS ESTADÍSTICOS AVANZADOS

2.1. Aprendizaje automático

El aprendizaje automático enseña a las computadoras a hacer lo que es natural para los humanos y los animales: aprender de la experiencia. Los algoritmos de aprendizaje automático utilizan métodos computacionales para "aprender" información directamente de los datos sin depender de una ecuación predeterminada como modelo. Los algoritmos mejoran su rendimiento de forma adaptativa a medida que aumenta el número de muestras disponibles para el aprendizaje.

Los algoritmos de aprendizaje automático encuentran patrones naturales en los datos que generan información y los ayudan a tomar mejores decisiones y predicciones. Se utilizan todos los días para tomar decisiones críticas en diagnóstico médico, negociación de acciones, pronóstico de carga de energía y más. Los sitios de medios se basan en el aprendizaje automático para examinar millones de opciones y brindarle recomendaciones de canciones o películas. Las empresas y/o compañías de distribución, lo utilizan para conocer mejor el comportamiento de compra de sus clientes.

El aprendizaje automático como se observa en la Figura 1, utiliza dos tipos de técnicas [19]: aprendizaje supervisado, que entrena un modelo con datos de entrada y salida conocidos para que pueda predecir resultados futuros, y aprendizaje no supervisado, que encuentra patrones ocultos o estructuras intrínsecas en los datos de entrada.

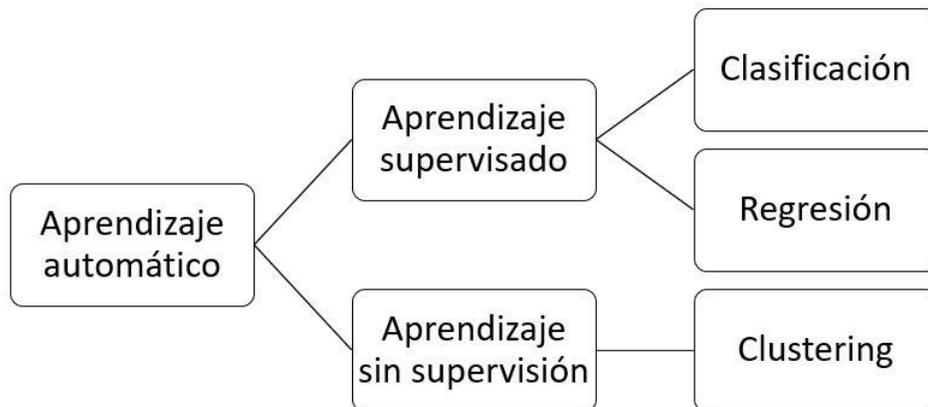


Figura 1. Clasificación de las técnicas del aprendizaje automático

2.1.1. Aprendizaje supervisado

El objetivo del aprendizaje automático supervisado es construir un modelo que haga predicciones basadas en evidencia en presencia de incertidumbre. Un algoritmo de aprendizaje supervisado toma un

conjunto conocido de datos de entrada y respuestas conocidas a los datos (salida) y entrena un modelo para generar predicciones razonables para la respuesta a nuevos datos.

El aprendizaje supervisado utiliza técnicas de clasificación y regresión para desarrollar modelos predictivos.

- **Las técnicas de clasificación** predicen respuestas discretas, por ejemplo, si un correo electrónico es genuino o spam, o si un tumor es canceroso o benigno. Los modelos de clasificación clasifican los datos de entrada en categorías. Las aplicaciones típicas incluyen imágenes médicas, reconocimiento de voz y calificación crediticia.
- **Las técnicas de regresión** predicen respuestas continuas, por ejemplo, cambios de temperatura o fluctuaciones en la demanda de energía. Las aplicaciones típicas incluyen pronóstico de carga de electricidad y comercio algorítmico.

2.1.2. Aprendizaje sin supervisión

El aprendizaje no supervisado encuentra patrones ocultos o estructuras intrínsecas en los datos. Se utiliza para extraer inferencias de conjuntos de datos que constan de datos de entrada sin respuestas etiquetadas.

La agrupación en clústeres es la técnica de aprendizaje no supervisada más común. Se utiliza para el análisis exploratorio de datos para encontrar patrones o agrupaciones ocultos en los datos.

2.2. Modelo autorregresivo integrado de media móvil

Un modelo de serie temporal lineal para el proceso de respuesta y_t e innovaciones aleatorias ε_t es un proceso estocástico en el que la respuesta actual es una función lineal de las respuestas anteriores, las innovaciones actuales y anteriores, y las covariables exógenas x_t . En notación de ecuaciones de diferencia, la forma general de un modelo de serie temporal lineal es [20]:

$$y_t = c + x_t\beta + a_1y_{t-1} + \dots + a_wy_{t-w} + \varepsilon_t + b_1\varepsilon_{t-1} + \dots + b_v\varepsilon_{t-v} \quad (2.1)$$

Donde $c, \beta, a_1, a_w, b_1, b_v$ son parámetros de la ecuación polinómica

Dados w y v cuyas variables representan el orden del modelo ARIMA, todos los coeficientes son estimables. Expresado en notación de operador de retraso, el formulario de modelo general es:

$$a(L)y_t = c + x_t\beta + b(L) \quad (2.2)$$

Los polinomios del operador de retraso en el modelo a menudo se expresan como productos de polinomios para efectos estacionales no estacionales, multiplicativos e integración:

$$\phi(L)(1-L)^D\Phi(L)(1-L^s)^{D_s}y_t = c + x_t\beta + \theta(L)Y(L)\varepsilon_t \quad (2.3)$$

donde $\phi(L)$ es un polinomio Auto-Regresivo (AR por sus siglas en inglés) no estacional estable de grado v , D es el grado de integración no estacional, $\Phi(L)$ es un polinomio AR estacional multiplicativo, s es el grado del polinomio de diferenciación estacional, D_s es el grado de integración estacional, c es la constante del modelo, β es el coeficiente de regresión de covariables exógenas, θ es un polinomio de Media Móvil (MA por sus siglas en inglés) no fuera de tiempo invertible grado w , Y es un polinomio MA estacional multiplicativo y ε es una serie de innovaciones aleatorias.

2.3. Regresión Lineal Generalizada

La Regresión Lineal Generalizada o RLG pertenece a una familia exponencial de distribuciones sobredispersas, la cuales son una generalización de la familia exponencial y el modelo de dispersión exponencial de distribuciones e incluye esas distribuciones de probabilidad, parametrizadas por φ y τ , cuyas funciones de densidad f (o función de masa de probabilidad, para el caso de una distribución discreta) se pueden expresar en la forma [21]:

$$f_Y = h(Y|\varphi, \tau) = \exp\left(\frac{b(\varphi)^T T(y) - A(\varphi)}{d(\tau)}\right) \quad (2.4)$$

El parámetro de dispersión T , generalmente es un valor conocido y generalmente está relacionado con la varianza de la distribución. Las funciones $h(Y, \tau)$, $b(\varphi)$, $T(y)$, $A(\varphi)$ y $d(\tau)$ son conocidas. Muchas distribuciones comunes están en esta familia, incluyendo binomial, multinomial y binomial normal, exponencial, gamma, Poisson, Bernoulli y para un número fijo de ensayos.

Para el escalar Y y φ , esto se reduce a:

$$f_Y = h(Y|\varphi, \tau) = \exp\left(\frac{b(\varphi)T(Y) - A(\varphi)}{d(\tau)}\right) \quad (2.5)$$

φ está relacionado con la media de la distribución. Si $b(\varphi)$ es la función de identidad, se dice que la distribución está en forma canónica (o forma natural). Tenga en cuenta que cualquier distribución se puede convertir a una forma canónica reescribiendo φ como φ' y luego aplicar la transformación $\varphi = b(\varphi')$. Siempre es posible convertir $A(\varphi)$ en términos de la nueva parametrización, incluso si $b(\varphi')$ no es una

función de uno a uno. Si, además, $T(Y)$ es la identidad y τ se conoce, entonces φ se llama el parámetro canónico (o parámetro natural) y está relacionado con la media a través de

$$\mu = E(Y) = \Delta A(\varphi) \quad (2.6)$$

Para Y y φ esto se reduce a

$$\mu = E(Y) = A'(\varphi) \quad (2.7)$$

Bajo este escenario, la varianza de la distribución puede mostrarse como

$$Var(Y) = \Delta^2 A(\varphi) d(T) \quad (2.8)$$

Para Y y φ esto se reduce a

$$Var(Y) = A''(\varphi) d(T) \quad (2.9)$$

El predictor lineal es la cantidad que incorpora la información sobre las variables independientes en el modelo. El símbolo η denota un predictor lineal. Está relacionado con el valor esperado de los datos a través de la función de enlace.

El parámetro η se expresa como combinaciones lineales de parámetros desconocidos γ . Los coeficientes de la combinación lineal se representan como la matriz de las variables independientes X . Por lo tanto, η puede expresarse como

$$\eta = X\gamma \quad (2.10)$$

La función de enlace proporciona la relación entre el predictor lineal y la media de la función de distribución. Existen muchas funciones de enlace de uso común, y su elección se basa en varias consideraciones. Siempre hay una función de enlace canónico bien definida que se deriva del exponencial de la función de densidad de la respuesta. Sin embargo, en algunos casos tiene sentido tratar de hacer coincidir el dominio de la función de enlace con el rango de la media de la función de distribución, o usar una función de enlace no canónica con fines algorítmicos.

$$E(Y) = \mu = g^{-1}(\eta) \quad (2.11)$$

2.4. Árboles de regresiones

Se utiliza un árbol de decisión como un modelo predictivo que mapea observaciones sobre un artículo a conclusiones sobre el valor objetivo del artículo [22]. Es uno de los enfoques de modelado predictivo utilizados en estadísticas, minería de datos y aprendizaje automático. Los modelos de árbol, donde la variable de destino puede tomar un conjunto finito de valores se denominan árboles de clasificación. En estas estructuras de árbol, las hojas representan etiquetas de clase y las ramas representan las conjunciones de características que conducen a esas etiquetas de clase. Los árboles de decisión, donde la variable de destino puede tomar valores continuos (por lo general números reales) se llaman árboles de regresión (TREE).

Los algoritmos para la construcción de árboles de decisión suelen trabajar de manera top-down, escogiendo en cada paso la variable que mejor divide el conjunto de elementos. Estos algoritmos miden generalmente la homogeneidad de la variable de destino dentro de los subconjuntos. Algunos ejemplos se dan a continuación. Estas métricas se aplican a cada subconjunto candidato, y los valores resultantes se combinan (por ejemplo, un promedio) para proporcionar una medida de la calidad de la división.

Mediante el algoritmo de Árboles de Clasificación y Regresión (ACR), la impureza de Gini es una medida de cuán a menudo un elemento elegido aleatoriamente del conjunto sería etiquetado incorrectamente si fue etiquetado de manera aleatoria de acuerdo a la distribución de las etiquetas en el subconjunto. La impureza de Gini se puede calcular sumando la probabilidad de cada elemento elegido multiplicado por la probabilidad de un error en la categorización de ese elemento. Alcanza su mínimo (cero) cuando todos los casos del nodo corresponden a una sola categoría de destino.

Para calcular la impureza de Gini representada como I_G de un conjunto de elementos

$$I_G(f) = 1 - \sum_{i=1}^m f_i^2 \quad (2.12)$$

Donde f_i es la probabilidad de que un ejemplo sea de la clase i y m la clase a la que pertenece.

2.5. Soporte de regresión de la máquina vectorial

Una máquina de vectores de soporte (SVM por sus siglas en inglés) es un algoritmo de aprendizaje supervisado que se puede emplear para clasificación binaria o regresión [23]. Las SVM's son muy populares en aplicaciones como el procesamiento del lenguaje natural, el habla, el reconocimiento de imágenes y la visión artificial.

Una SVM construye un hiperplano óptimo en forma de superficie de decisión, de modo que el margen de separación entre las dos clases en los datos se amplía al máximo. Los vectores de soporte hacen referencia a un pequeño subconjunto de las observaciones de entrenamiento que se utilizan como soporte para la ubicación óptima de la superficie de decisión.

El entrenamiento de una SVM consta de dos fases:

1. Transformar los predictores (datos de entrada) en un espacio de características altamente dimensionales. En esta fase es suficiente con especificar el kernel (Tabla 1); los datos nunca se transforman explícitamente en el espacio de características. Este proceso se conoce comúnmente como el truco del kernel.
2. Resolver un problema de optimización cuadrática que se ajuste a un hiperplano óptimo para clasificar las características transformadas en dos clases. El número de características transformadas está determinado por el número de vectores de soporte.

Para construir la superficie de decisión solo se requieren los vectores de soporte seleccionados α_1 y α_2 de los datos de entrenamiento. Una vez hecho el entrenamiento de la SVM, el resto de los datos de entrenamiento son irrelevantes.

Tabla 1. Kernels populares que se emplean con el SVM

Tipo de SVM	Kernel de Mercer	Descripción
Función gaussiana	$K(\alpha_1, \alpha_2) = \exp\left(-\frac{\ \alpha_1 - \alpha_2\ ^2}{2\sigma^2}\right)$	Aprendizaje de una clase. σ representa la anchura del kernel.
Lineal	$K(\alpha_1, \alpha_2) = \alpha_1^T \alpha_2$	Aprendizaje de dos clases.
Polinómica	$K(\alpha_1, \alpha_2) = (\alpha_1^T \alpha_2 + 1)^\rho$	ρ representa el orden del polinomio.
Sigmoide	$K(\alpha_1, \alpha_2) = \tanh(\beta_0 \alpha_1^T \alpha_2 + \beta_1)$	Representa un kernel de Mercer solo para determinados valores β_0 y β_1 .

2.6. Regresión de proceso gaussiano

Los Procesos Gaussianos (PG) se definen como una distribución de probabilidad sobre funciones aleatorias. Son sobre colecciones infinitas de variables (funciones), tal que cualquier subconjunto de variables aleatorias finitas tiene una distribución gaussiana multivariable [24]. Estos modelos probabilísticos son basados en kernel, esto quiere decir que en el aprendizaje se espera que los puntos con valores predictivos similares, naturalmente tengan valores de respuesta cercana (objetivo). En los procesos gaussianos la función de covarianza (kernel) expresa esta similitud. Considere un conjunto de entrenamiento: $(o_j, r_j); j = 1, 2, \dots, n$, donde $o_j \in R^d$, $r_j \in R$ procedente de una distribución desconocida. Un modelo de Regresión de Proceso

Gaussiano (RPG por sus siglas en inglés) aborda la cuestión de predecir el valor de una variable de respuesta y dado el nuevo vector de entrada o y los datos de entrenamiento. Un modelo de regresión lineal tiene la siguiente forma:

$$r = o^T \omega + \delta \quad (2.13)$$

donde ω son los datos de entrada, o^T vector de entrada transpuesto.

Un modelo RPG explica la respuesta introduciendo variables latentes $f(o_j), j = 2, 3, \dots, n$ de un PG y funciones de bases explícitas h . La función de covarianza de las variables latentes captura la suavidad de las funciones de respuesta base r y proyectan entradas o en un espacio de características dimensional p .

Un PG es un conjunto de variables aleatorias, de modo que cualquier número finito de ellas tiene una distribución gaussiana conjunta. Si $f(o), o \in R^d$ es un PG, se le dan observaciones o_1, o_2, \dots, o_n la distribución conjunta de las variables aleatorias $f(o_1), f(o_2), \dots, f(o_n)$ es gaussiano, además un PG se define por su función media $z(o)$ y su función de covarianza $k(o, o')$. Es decir, si $f(o), o \in R^d$, entonces

$$e(f(o)) = z(o) \quad (2.14)$$

$$Cov[f(o), f'(o)] = e[f(o) - z(o), f(o') - z(o')] = k(o, o') \quad (2.15)$$

Las funciones integradas del kernel (covarianza) son mostradas en la Tabla 2 con la misma escala de longitud para cada predictor son:

Tabla 2. Algunos tipos de Kernel para RPG

Kernel	Función
Exponencial al cuadrado	$k(o, o' \theta) = s_f^2 \exp\left(-\frac{1}{2}\right) \left(\frac{(o_i - o_j)^T (o_i - o_j)}{s_l^2}\right)$
Exponencial	$k(o, o' \theta) = s_f^2 \exp\left(\frac{u}{s_l^2}\right)$
Mater 3/2	$k(o, o' \theta) = s_f^2 \left(1 + \frac{\sqrt{3}u}{s_l}\right) \exp\left(-\frac{\sqrt{3}u}{s_l}\right)$
Mater 5/2	$k(o, o' \theta) = s_f^2 \left(1 + \frac{\sqrt{5}u}{s_l} + \frac{5u^2}{3s_l^2}\right) \exp\left(-\frac{\sqrt{5}u}{s_l}\right)$
Núcleo cuadrático racional	$k(o, o' \theta) = s_f^2 \left(1 + \frac{u^2}{2\alpha s_l^2}\right)^{-\alpha}$

Donde u es la distancia euclidiana entre o_i y o_j , s_l es la escala de longitud característica, y s_f es la desviación estándar de la señal.

2.7. Construcción y evaluación de modelos

Para la construcción y evaluación de modelos (CEM) se debe determinar cuáles son las entradas y salidas esperadas, para ello, habría que responder a las siguientes cuestiones [25]:

- Objetivo principal, ¿Qué vamos a tratar de predecir?
- Características objetivo
- Mejora esperada
- Estado actual de la variable objetivo
- Medición de la variable objetivo

Entonces para realizar la construcción del modelo es necesario definir adecuadamente el problema. A continuación, realizar la recopilación de los datos con los que vamos a tratar. Es importante establecer el protocolo de evaluación y cuáles son los que están disponibles para establecer en el estudio. Una vez recopilados los datos deben tratarse correctamente y dividirse. No todos los problemas pueden resolverse, sólo podemos hacer cierta hipótesis hasta que se tenga un modelo funcional. Al crear un modelo de clasificación predictiva de alta calidad, es importante seleccionar las entidades (o predictores) correctas y ajustar los hiper parámetros (parámetros de modelo que no se estiman).

2.8. Red neuronal

Las redes neuronales (RN) son una implementación muy sencilla de un comportamiento local observado en nuestros cerebros [26]. El cerebro está compuesto de neuronas, las cuales son elementos individuales de procesamiento. La información viaja entre las neuronas, y basado en la estructura y ganancia de los conectores neuronales, la red se comporta de forma diferente.

La Figura 2 representa una neurona artificial, donde cada neurona está conectada con otra neurona por medio de un peso de ajuste representado por la letra N , el primer subíndice indica la neurona destino, mientras que el segundo subíndice indica la neurona de origen.

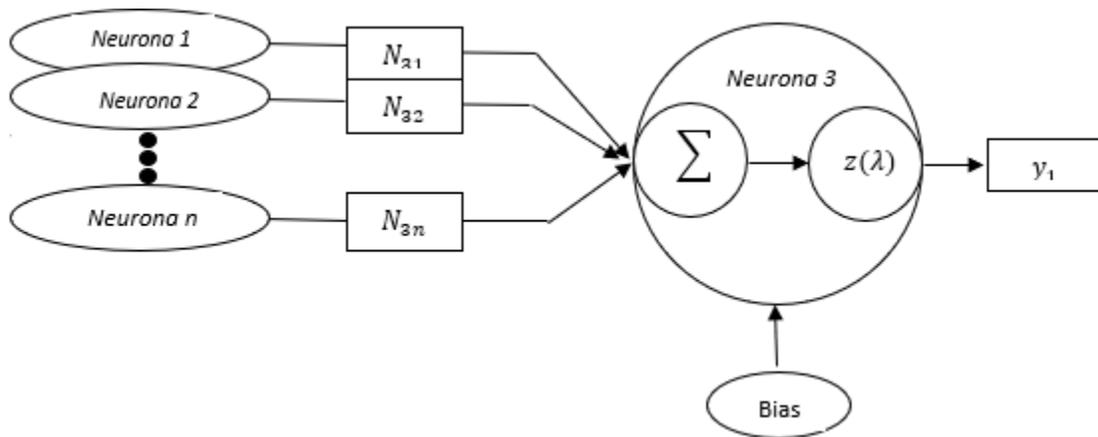


Figura 2. Representación de una neurona artificial.

Una neurona artificial está formada por un sumador y una función de activación, representada por $z(\lambda)$, donde la función de activación λ debe ser un sigmoide, continua, real, de rango acotado y tener una derivada positiva. Una de las funciones más populares en las RN es la de logística:

$$z(\lambda) = \frac{1}{1 + e^{(-\lambda)}} \quad (2.16)$$

Los resultados corresponden a las puntuaciones de clasificación predichas (o probabilidades posteriores).

2.9. Comparación entre las técnicas de estimación

2.9.1. Validaciones

La validación estima el rendimiento del modelo en los nuevos datos y le ayuda a elegir el mejor modelo. La validación protege contra el sobreajuste. Un modelo que es demasiado flexible y sufre de sobreajuste tiene una peor precisión de validación.

- **Validación cruzada:** Seleccione el número de pliegues (o divisiones) para particionar el conjunto de datos. Si elige k pliegues, entonces la aplicación:
 - Particiona los datos en k conjuntos o pliegues disjuntos
 - Para cada pliegue de validación:
 - Entrena un modelo utilizando las observaciones de plegado de entrenamiento (observaciones que no están en el pliegue de validación)
 - Evalúa el rendimiento del modelo mediante datos de validación
 - Calcula el error de validación medio en todos los pliegues

Este método proporciona una buena estimación de la precisión predictiva del modelo final entrenado utilizando el conjunto de datos completo. El método requiere múltiples ajustes, pero hace un uso eficiente de todos los datos, por lo que funciona bien para conjuntos de datos pequeños.

- **Validación de holdout:** Seleccione un porcentaje de los datos que desea utilizar como conjunto de validación. La aplicación entrena un modelo en el conjunto de entrenamiento y evalúa su rendimiento con el conjunto de validación. El modelo utilizado para la validación se basa en solo una parte de los datos, por lo que la validación de holdout es apropiada solo para grandes conjuntos de datos. El modelo final se entrena utilizando el conjunto de datos completo.
- **Validación de resustitución:** No hay protección contra el sobreajuste. La aplicación utiliza todos los datos para el entrenamiento y calcula la tasa de error en los mismos datos. Sin ningún dato de validación separado se obtiene una estimación poco realista del rendimiento del modelo en los nuevos datos, es decir, es probable que la precisión de la muestra de entrenamiento sea irrealistamente alta, y es probable que la precisión predictiva sea menor.

2.9.2. Rendimiento

Los parámetros estadísticos del modelo se utilizan para evaluar y comparar modelos. Los parámetros de los resultados de la capacitación se calculan en el conjunto con la estimación misma. En la Tabla 3 se muestran los principales parámetros de referencia para evaluar la precisión de estimación del CBL.

Tabla 3. Estadísticas del modelo.

Estadísticas	Descripción
Error Cuadrado de la Media Raíz (RMSE)	Siempre es positivo y sus unidades coinciden con las unidades de su respuesta.
Error Cuadrático Medio (MSE)	El MSE es el cuadrado del RMSE
Error Medio Absoluto (MAE)	Siempre positivo y similar al RMSE, pero menos sensible a los valores atípicos.

2.9.3. Comportamiento gráfico

La gráfica predicha frente a Real se utiliza para comprobar el rendimiento del modelo, en esta gráfica se comprende qué tan bien el modelo de regresión hace predicciones para diferentes valores de respuesta. Un modelo de regresión perfecto tiene una respuesta predicha igual a la respuesta verdadera, por lo que todos los puntos se encuentran en una línea diagonal. La distancia vertical desde la línea hasta cualquier punto es el

error de la predicción para ese punto. Un buen modelo tiene pequeños errores, lo que significa que las predicciones están dispersas cerca de la línea.

El gráfico de residuos muestra la diferencia entre las respuestas predichas y las verdaderas. Por lo general, un buen modelo tiene residuos dispersos simétricamente alrededor de 0. Si puede ver patrones claros en los residuos, es probable que pueda mejorar su modelo. Busque estos patrones:

- Los residuos no se distribuyen simétricamente alrededor de 0.
- Los residuos cambian significativamente de tamaño de izquierda a derecha en la parcela.
- Se producen valores atípicos, es decir, residuos que son mucho más grandes que el resto de los residuos.
- Un patrón claro y no lineal aparece en los residuos.

2.9.4. Trabajo de referencia IEEE

En [27] se realiza una comparación entre los métodos de regresión de la herramienta de Aprendizaje Automático para realizar la estimación de datos de consumo eléctrico. En [27] se puede observar que se utilizan datos de cuatro días previos (Figura 3) para obtener una predicción del día siguiente a los datos tomados en un sistema de seis nodos.

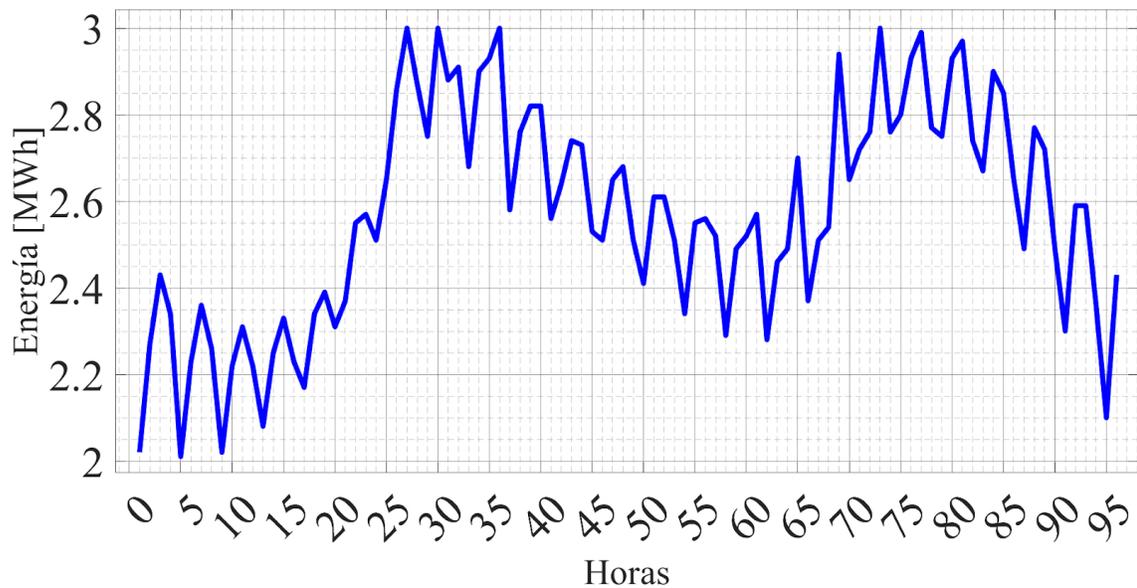


Figura 3. Serie Temporal de la demanda para 96 horas previas.

En la Figura 4 tomada de la referencia mencionada, se puede observar una comparación entre las respuestas de los métodos, en donde se puede observar en primera instancia que los métodos del Aprendizaje Automático realizan una mejor estimación que el modelo ARIMA.

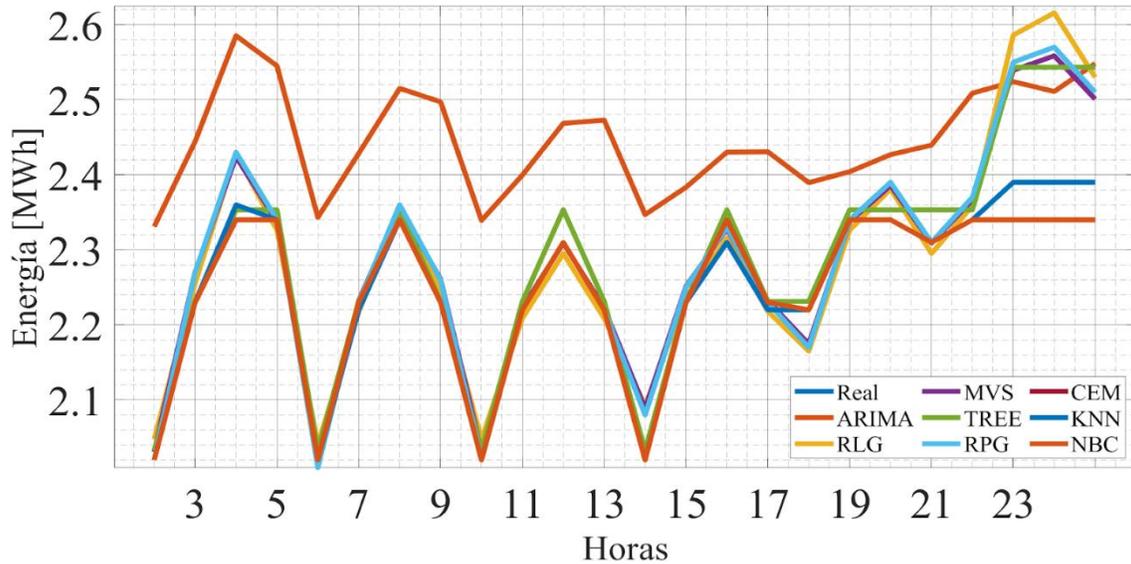


Figura 4. Resultados de los pronósticos para 24 horas.

Para visibilizar mejor la comparación entre los métodos se realiza un acercamiento en el periodo comprendido entre la hora 13:00 y la hora 17:00, mostrado en la Figura 5.

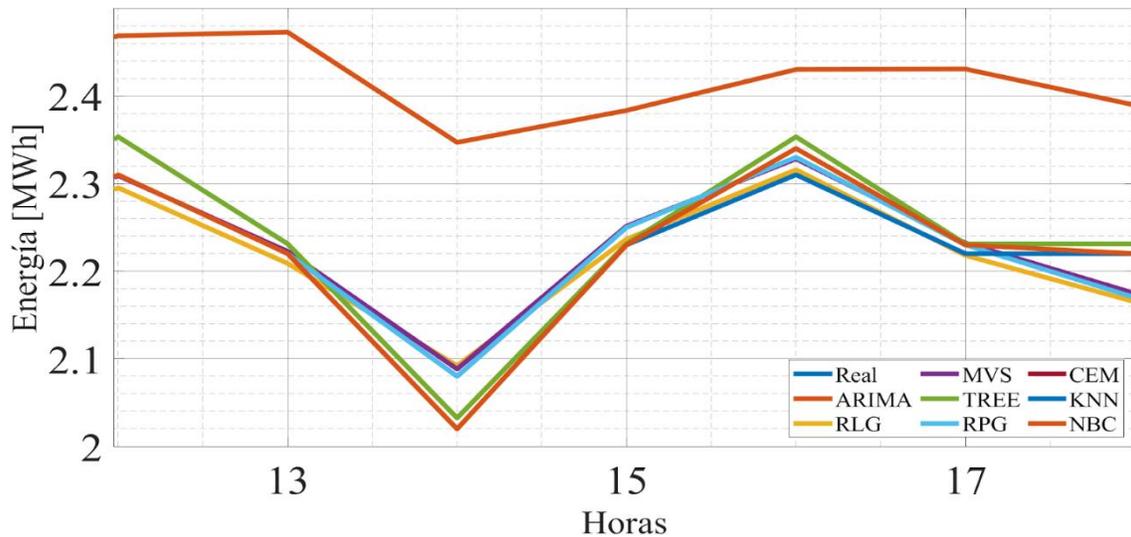


Figura 5. Acercamiento de la hora 13:00 a 17:00

Se logra observar que RPG tiene un comportamiento muy similar a los datos reales; sin embargo, no es el único que aparenta tener ese mismo compartimento, por lo cual, se calculó el MSE de cada uno de los métodos para finalmente determinar cuál fue mejor.

Tabla 4. Error Cuadrático Medio.

Técnica estadística	MSE
ARIMA	1.5×10^{-3}
Regresión lineal generalizada	0.37×10^{-3}
Clasificación de máquinas vectoriales de soporte	0.037×10^{-3}
Árboles de regresión	0.099×10^{-3}
Regresión de proceso gaussiano	5.2×10^{-9}
Construcción y evaluación de modelos	6.1×10^{-3}
Clasificador vecino más cercano	0.66
Clasificador de Naive Bayes	0.58

Citando a los autores del artículo [27], “Debido a las herramientas en las que se apoya la ciencia de los datos (Data Science), ha permitida que ésta sea una alternativa para la gestión y predicción de datos, ya que al analizar las respuestas que se obtienen por los mismos, podemos afirmar que son superiores a la técnica ARIMA”.

Al observar la Figura 4, notamos que el comportamiento de la predicción obtenida mediante el modelo RPG es muy similar al comportamiento de los datos reales (datos del quinto día). Sin embargo, hay otros modelos como el SVM ó RLG que siguen casi de cerca el comportamiento de los datos reales, para ello, se muestra el valor del MSE con la finalidad de comprobar qué modelo obtiene el menor error.

El modelo RPG obtuvo un valor de MSE más bajo que el resto de los modelos como se puede observar en la Tabla 4, donde se muestra en un subrayado amarillo los valores de MSE del modelo ARIMA, RPG y CNB, siendo el menor de estos RPG obteniendo un valor de 5.2×10^{-9} .

CAPÍTULO 3. RELACIÓN CARGA DE REFERENCIA DEL CLIENTE – REGRESIÓN DE PROCESO GAUSSIANO.

3.1. Concepto de la Carga de Referencia del Cliente

El valor ideal de la Carga de Referencia del Cliente (CBL por sus siglas en inglés) es la cantidad de carga de electricidad que el cliente usaría si no hubiera un evento de Respuesta de la Demanda (DR por sus siglas en inglés). Sin embargo, para un evento de DR real, los consumidores se comportan de manera diferente y no hay forma de obtener el valor CBL ideal. Por lo tanto, el objetivo del DR es estimar el valor CBL ideal con la mayor precisión posible. La función utilizada para la estimación de CBL se denomina función de línea de base y las funciones de línea de base más populares se describen en esta sección.

Algunos de los errores de CBL se deben a diferentes condiciones entre cada día del evento y los días sin eventos relevantes. Por ejemplo, cuando la actividad del cliente, el clima o la condición de estacionalidad son diferentes para los días sin eventos cuyos datos se utilizan para calcular el CBL, la diferencia aumenta el error del CBL estimado.

Los modelos CBL tradicionales se pueden dividir en modelos CBL simples y modelos CBL ajustados. Un modelo CBL simple se basa en una función de línea de base con los datos de carga históricos u otros datos externos antes del día del evento como entrada, mientras que un modelo CBL ajustado utiliza adicionalmente datos del día del evento, como la carga horaria previa al evento o los datos meteorológicos del evento para aplicar un ajuste al modelo CBL simple [28].

3.2. Respuesta a la demanda

La DR se considera ampliamente como un medio prometedor para mantener la confiabilidad y mejorar la flexibilidad de los sistemas de energía. Generalmente, el DR se puede clasificar en dos categorías: basada en el precio y basados en incentivos como se aprecia en la figura 6. El primero permite a las empresas de servicios públicos del sector eléctrico gestionar las cargas industriales directamente y, generalmente, es adoptado por grandes clientes comerciales e industriales con patrones de carga altamente predecibles. De igual forma, estas empresas obtienen la capacidad de DR de los clientes a través de diferentes programas y luego comercian en el mercado de la electricidad.

El segundo está impulsado por la voluntad de los clientes y es más aplicable a las pequeñas y medianas empresas (PYME). Por ejemplo, una empresa de servicios eléctricos de DR basado en incentivos logra

reducciones de carga al proporcionar incentivos financieros a los participantes de DR para reducir su consumo de electricidad durante el período de tiempo específico después de que se activa la DR.

La DR puede reducir los consumos cambiando la carga de electricidad de los clientes finales cuando el precio del mercado mayorista es temporalmente alto debido a la alta demanda; por lo tanto, puede proporcionar beneficios económicos tanto a los proveedores de respuesta a la demanda como a los clientes. Además, puede reducir las inversiones generales de la planta y los costos de capital a largo plazo al reducir la demanda máxima y mejorar la confiabilidad del sistema eléctrico. Los clientes de la DR se pueden dividir en tres sectores principales: clientes comerciales, industriales y residenciales. Si bien el sector residencial es un segmento importante en su conjunto, tiene una carga reducible por cliente considerablemente menor en comparación con los grandes clientes comerciales e industriales [29].

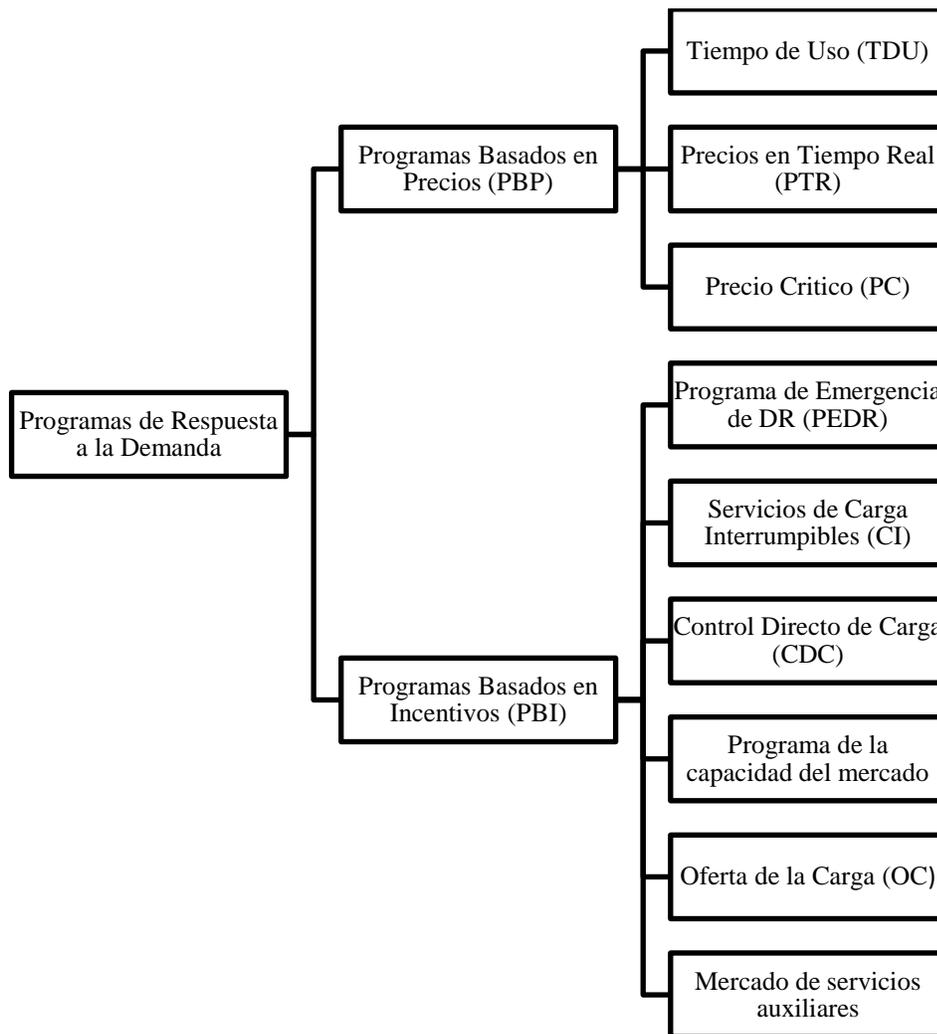


Figura 6. Clasificación del programa de respuesta a la demanda

3.3. Regresión de Proceso Gaussiano aplicado a la Carga de Referencia del Cliente

Para un subconjunto de datos registrados de un CBL por hora del consumo de carga de los días hábiles de una semana, se estima el comportamiento del patrón de demanda del día siguiente. Se selecciona el conjunto activo (datos de interés) y calcula las estimaciones de parámetros en una serie de iteraciones.

En la primera iteración, el software utiliza los valores de los parámetros iniciales (ν_0, ψ_0, ι_0) y los agrupa una nueva variable inicial $\eta_0 = [\nu_0, \rho_0, \theta_0]$ para seleccionar un conjunto activo A_1 . Maximiza la probabilidad logarítmica marginal RPG o su aproximación utilizando η_0 como los valores iniciales y A_1 para calcular las nuevas estimaciones de parámetros η_1 . A continuación, calcula la nueva probabilidad de registro L_1 utilizando η_1 y A_1 .

En la segunda iteración, el software selecciona el conjunto activo A_2 uso de los valores de parámetro en η_1 . Luego, usando η_1 como los valores iniciales y A_2 , maximiza la probabilidad logarítmica marginal RPG o su aproximación y estima los nuevos valores de parámetros η_2 . Luego usando η_2 y A_2 , calcula el nuevo valor de probabilidad de registro L_2 .

En la Tabla 5 se resumen las iteraciones y el conjunto de parámetros obtenidos de loa estimación del CBL.

Tabla 5. Resumen de las iteraciones para el CBL basado en RPG

Número de iteración	Conjunto activo	Vector de parámetros	Probabilidad de registro
1	A_1	η_1	L_1
2	A_2	η_2	L_2
3	A_3	η_3	L_3
...

El método de predicción predeterminado es para $n \leq 10000$, donde n es el número de observaciones en los datos de entrenamiento.

3.4. Flujos de Potencia Óptimos

El modelo de Flujos de Potencia Óptimos (OPF por sus siglas en inglés) busca encontrar la mejor solución de las ecuaciones que describen la potencia neta inyectada en los nodos del sistema eléctrico de potencia bajo un determinado criterio, por lo que, a través de este modelo se pueden obtener los costos de todos los generadores conectados al sistema eléctrico, para todos los períodos de tiempo t a analizar [16-18]. El modelo matemático se expresa como:

$$Min = \sum_{i=1}^{Ng} C_{i,t}(Pg_{i,t}) \quad (3.1)$$

Sujeto a

$$\sum_{i=1}^n (Pg_{i,t} - Pd_{i,t}) = \sum_{i=1}^n P_{i,t}(V, \theta) \quad (3.2)$$

$$\sum_{i=1}^n (Qg_{i,t} - Qd_{i,t}) = \sum_{i=1}^n Q_{i,t}(V, \theta) \quad (3.3)$$

$$Pg_{i,t}^{min} \leq Pg_{i,t} \leq Pg_{i,t}^{max} \quad (3.4)$$

$$Qg_{i,t}^{min} \leq Qg_{i,t} \leq Qg_{i,t}^{max} \quad (3.5)$$

$$V_{i,t}^{min} \leq V_{i,t} \leq V_{i,t}^{max} \quad (3.6)$$

$$P_{im,t} \leq P_{im,t}^{max} \quad (3.7)$$

donde $C_{i,t}^j$ es la oferta de energía del i -ésimo generador en el bus i , en el período t , $Pg_{i,t}^j$ es la potencia de generación del bus i en el período t , Ng es el número total de generadores en el sistema, Nb es el total de buses en el sistema, t es el período de tiempo a analizar, $Pd_{i,t}^j$ es la demanda de energía activa en el bus i para el período t , donde cada nodo i tiene un patrón de consumo diferente asociado con el período de tiempo t , $Qg_{i,t}^j$ es la potencia reactiva generada en el bus i y el período t , $Qd_{i,t}^j$ es la demanda de potencia reactiva en el bus i y el período t , θ es el ángulo nodal, $V_{i,t}^j$ es la magnitud del voltaje en el nodo i y el período t , $V_{i,t}^{min}$, $V_{i,t}^{max}$ son los límites de voltaje nodal mínimo y máximo, $P_{im,t}$ es el flujo de energía de la línea a través de los nodos im , n es el número total independiente de elementos en todo el sistema eléctrico (como generadores, transformadores, líneas de transmisión). Donde cada período t se refiere al comportamiento de la demanda estimada de CBL total.

La ecuación (3.1) representa los costos operativos (función objetivos), las restricciones (3.2) y (3.3) representan el balance nodal de potencia activa y reactiva respectivamente. Las restricciones (3.4) y (3.5) son las potencias máximas y mínimas, reales y reactivas respectivamente de cada generador. La restricción (3.6) es el límite de magnitud de voltaje del bus i . La restricción (3.7) es el límite físico de las líneas de transmisión para el flujo de energía.

CAPÍTULO 4. METODOLOGÍA PROPUESTA

4.1. Adquisición de datos.

Mediante la base de datos de [30] se obtiene un archivo con datos de consumo eléctrico para distintas regiones. Esta referencia es un portal web de la organización Interconexión Pensilvania-Nueva Jersey-Maryland (PJM por sus siglas en inglés), la cual, almacena datos de demanda de los estados antes mencionados.

Para cada uno de los nodos del sistema propuesto, se le adaptaron cuatro días de consumo previo, los cuales sirven para entrenar a los modelos estadísticos Regresión de Proceso Gaussiano (RPG) y Media Móvil Integrada Auto-Regresiva (ARIMA), mientras que el quinto día sirve como referencia (datos esperados).

4.2. Estimación del CBL mediante los modelos estadísticos RPG y ARIMA.

Con los datos seleccionados y previamente almacenados en una base de datos, se ajusta el modelo RPG utilizando una función de base lineal (ecuación 2.13) y el método de ajuste exacto para estimar los parámetros.

La adaptación de un modelo RPG implica estimar los siguientes parámetros del modelo a partir de los datos:

- Función de covarianza de la ecuación 2.15 parametrizado en términos del kernel (Tabla 2)
- Varianza de ruido sr^2 , como la sumatoria del cuadrado de cada una de las n observaciones (x_i) menos la media de las observaciones (\bar{X}), entre el número de observaciones Nu menos uno.

$$sr^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{Nu} \quad (4.1)$$

- Vector de coeficiente de funciones de base fija especificado como un vector p por 1, donde p es el número de columnas en la matriz base.

Una vez ajustado el modelo RPG se aplica la secuencia algorítmica mencionada en la sección 3.3, que consiste en realizar una estimación de nuevos datos, a partir de los valores iniciales estimados con el ajuste del modelo y un conjunto de datos activos (datos predictivos).

Por otro lado, para la estimación hecha con el modelo ARIMA, primero se ajusta el modelo con base en los datos de entrada, para ello, se utiliza el modelo de serie temporal explicado en la sección 2.2. Este ajuste especifica la forma funcional para poder hacer uso de una de las variaciones del modelo.

Posteriormente se infieren varianzas condicionales de la serie de respuesta subyacente, y se usa la máxima verosimilitud restringida para nuevamente ajustar el modelo y obtener los datos de respuesta. Debido a que se pueden especificar entradas de datos de muestra previa de diferentes longitudes, se supone que todos los conjuntos especificados tienen estas características:

- El último conjunto de resultados (fila) en cada conjunto ocurre simultáneamente.
- El primer resultado en la muestra de estimación sigue inmediatamente al último resultado en la premuestra, con respecto a la frecuencia de muestreo.

4.3. Análisis de las estimaciones obtenidas mediante los modelos estadísticos

Parte de las actividades que se realizan para llevar a cabo el análisis de las estimaciones se mencionan en la sección 2.9. Partiendo del rendimiento de los modelos estadísticos, se pueden obtener distintos porcentajes de error al realizar las estimaciones como se menciona en la sección 2.9.2, siendo el modelo con los porcentajes de error más bajos, el mejor modelo predictor.

Seguido de esto se obtiene una gráfica para comparar las estimaciones hechas con los modelos mencionados en la sección 4.2 y los datos esperados del quinto día, como se menciona en la sección 2.9.3, de tal manera, que la estimación con el comportamiento gráfico con más similitud en su trayectoria a los datos reales y los porcentajes de error más bajo es nombrada el mejor método predictor.

En el caso de que el comportamiento de ambas estimaciones sea muy similar a los datos reales, se prosigue a comparar dato por dato en las tablas los valores de las estimaciones y los datos reales, para verificar con exactitud las diferencias mínimas y máximas entre los tres conjuntos de datos.

Conociendo que modelo estadístico es mejor para realizar la estimación, se obtiene la gráfica de residuos dispersor que es otra forma de observar la diferencia entre los resultados estimados con los datos verdaderos.

4.4. Evaluación de cada período del CBL – RPG en los Flujos de Potencia Óptimos.

Debido a que la estimación hecha con el modelo RPG presenta un error menor en comparación con el modelo ARIMA, además de adaptarse a un comportamiento similar a los datos reales, se precisa que el RPG es el mejor modelo para obtener una estimación del CBL de los usuarios. Por lo que, con la respuesta estimada del modelo RPG, se prosigue a aplicar el método de Flujos de Potencia Óptimos (OPF por sus siglas en inglés) para obtener la información del sistema eléctrico de potencia. Para ello, se almacenan los datos de la estimación hecha con el modelo RPG en una matriz, la cual es introducida al método OPF como los datos de potencia activa y reactiva en los nodos del sistema, es decir, los datos estimados con el modelo RPG

multiplican los valores nominales de potencia activa y reactiva del sistema para generar nuevos valores de potencias en los nodos basados en la demanda obtenida.

Con los nuevos valores de potencia en el sistema, se prosigue a aplicando las ecuaciones 3.1 a 3.7 de la sección 3.4 en los patrones de demanda obtenidos previamente, se obtienen los parámetros del sistema por nodo, como los costos operativos, generación de potencia activa, generación de potencia reactiva, voltajes, los precios marginales de localización de la potencia activa, los precios marginales de localización de la potencia reactiva, así como, los flujos de potencia activa en el sistema y sus pérdidas.

4.5. Análisis de los resultados obtenidos mediante el modelo OPF

Los parámetros obtenidos del sistema eléctrico de potencia mediante el modelo OPF, se analizan casi de la misma forma que se analizaron las estimaciones de los modelos estadísticos. En aquellos parámetros técnico-económicos donde solo se obtenga un dato por hora, se obtienen gráficas bidimensionales para comparar el comportamiento gráfico de los resultados obtenidos de la estimación de RPG con el comportamiento gráfico de los resultados obtenidos de los datos reales.

Para dar una mejor visualización de la diferencia entre los resultados, se realiza un acercamiento en un periodo de tiempo en donde se aprecien mejor los resultados y en las tablas de resultados obtenidas, se verifican con exactitud las diferencias mínimas y máximas entre los conjuntos de datos.

Por otro lado, para aquellos parámetros técnico-económicos donde se obtiene más de un dato por hora como es el caso de la potencia activa generada, se obtiene una gráfica en tres dimensiones, debido a que en este tipo de gráfica se tienen tres variables que analizar a la vez.

De igual manera que con las gráficas bidimensionales, se realizan acercamientos y cortes en periodos de tiempo específicos para poder apreciar con mayor facilidad la comparación entre los dos conjuntos de datos; sin embargo, los cortes serán a dos dimensiones para una mejor observación en el comportamiento de los resultados. Además, se verificarán los valores obtenidos en las tablas de resultados para observar con exactitud las diferencias mínimas y máximas entre los conjuntos de datos.

En caso de que se obtengan resultados fuera de lo común como periodos de tiempo sin generación, se modificara la capacidad de los nodos para que la demanda sea cubierta.

4.6. Metodología propuesta para la estimación del CBL

En esta subsección se muestra la metodología propuesta en esta tesis.

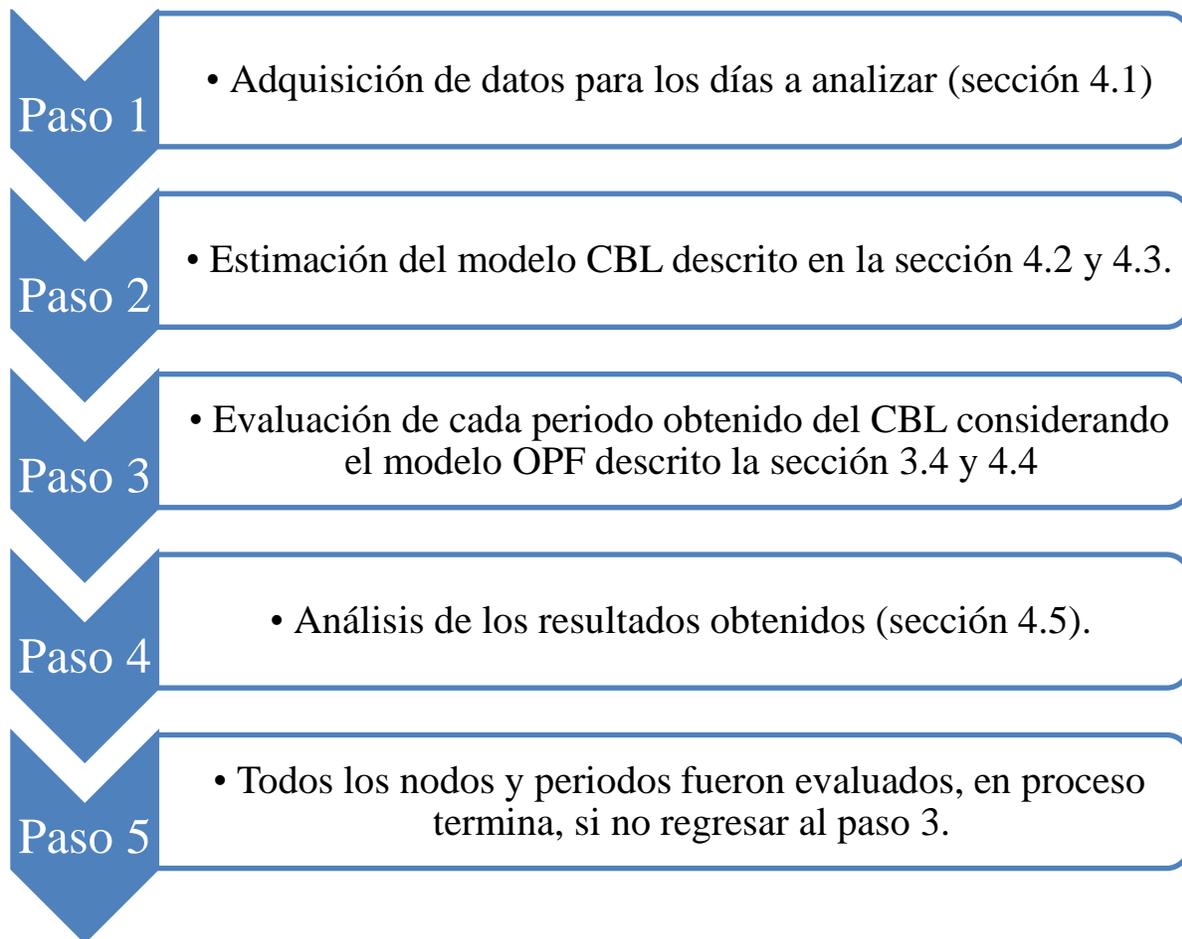


Figura 7. Metodología propuesta de esta tesis

Una vez que los resultados son obtenidos, se terminan de concluir las ventajas que tiene el método propuesto ante otros, debido a que los resultados adquiridos se asemejan más a los resultados esperados, es decir, en este último paso se observa como el comportamiento de los resultados del modelo propuesto presenta ventajas en cada uno de los nodos a lo largo del horizonte de 24 horas y se plantea la conclusión de trabajo.

4.6.1. Trabajo de referencia IEEE

En [31] se implementa la misma metodología que se utilizó en este trabajo; sin embargo, se utiliza para un sistema de seis nodos que se muestra en la Figura 8 y se utiliza como modelo de estimación el Clasificador de Naive Bayes (NBC).

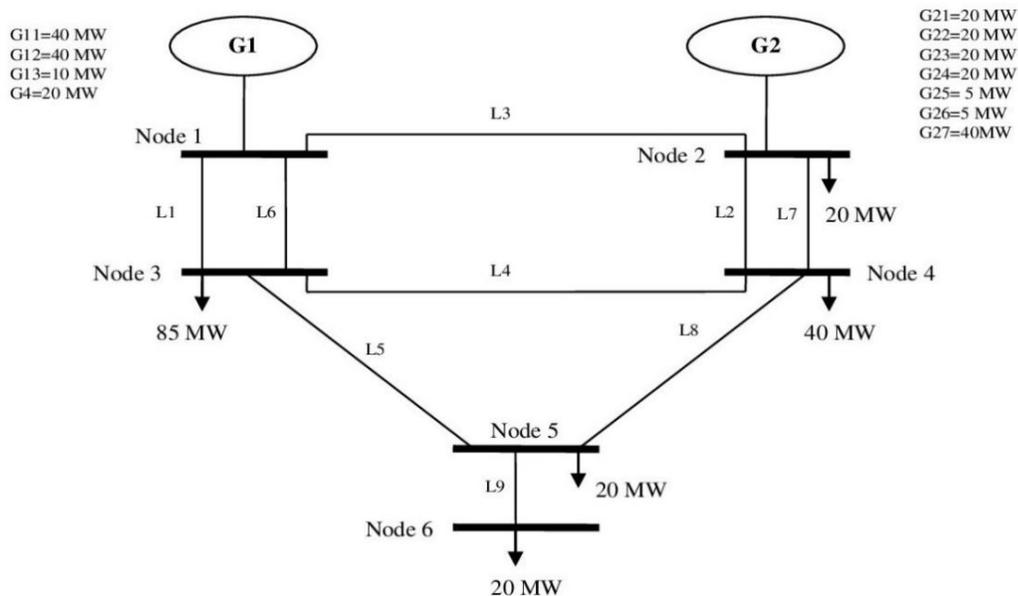


Figura 8. RBTS de seis nodos.

En la referencia antes mencionada se puede observar que los parámetros calculados mediante el modelo NBC tienen un comportamiento similar a los parámetros reales ahí establecidos, como se puede observar en la Figura 9, donde se comparan las estimaciones obtenidas del modelo NBC, ARIMA y los datos reales.

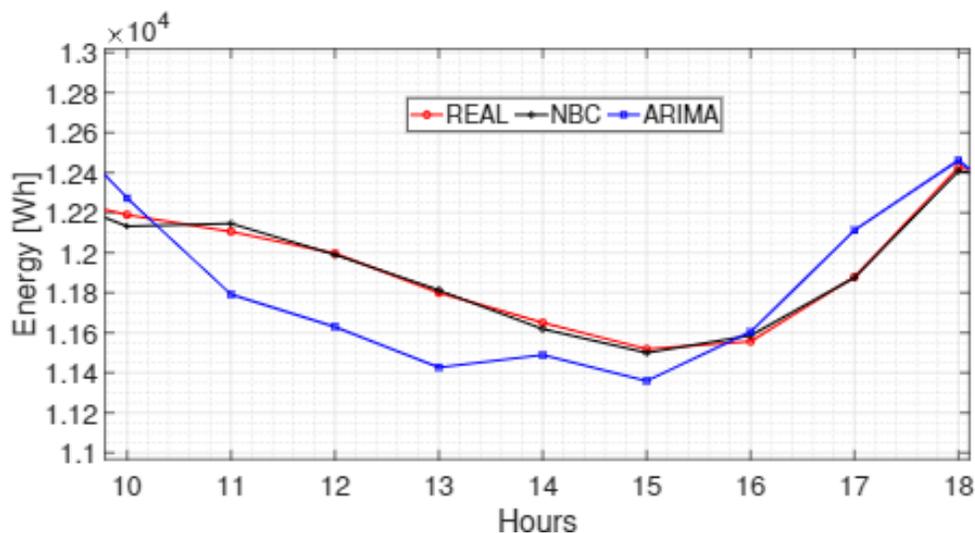


Figura 9. Comparación entre estimaciones de los modelos CNB, ARIMA y los datos reales.

No obstante, como se comprobó en la sección 2.9.4 el modelo RPG realiza una mejor estimación de datos que el modelo NBC, considerando que en la Tabla 4 se puede observar que el modelo RPG presenta un Error Cuadrático Medio (MSE por sus siglas en inglés) de $5.2 \times 10^{-9}\%$ mientras que el modelo NBC presenta

un MSE de 0.58%, siendo más de cien veces más pequeño el error obtenido con el modelo RPG. De igual forma, se puede observar en la Figura 9 que el comportamiento de los resultados obtenidos mediante el modelo RPG sigue con más fidelidad el comportamiento de los datos reales en comparación con el comportamiento de los datos del modelo NBC. En base a esto, se puede concluir que el modelo RPG realiza mejores estimaciones que el modelo NBC, por lo consiguiente realizará una mejor estimación de los parámetros del sistema.

CAPÍTULO 5. CASO DE ESTUDIO

5.1. Datos Generales del sistema eléctrico de potencia (Sistema de 24 nodos)

Se considera el sistema de potencia IEEE-RTS de 24 nodos como sistema de potencia prueba, mostrado en la Figura 10. Los parámetros eléctricos de los componentes que conforman el sistema IEEE-RTS de 24 nodos fueron tomados de la referencia [32] y se puede observar de la tabla 9 a la 11 en la sección Apéndice. Este sistema cuenta con dos niveles de tensión de operación, uno a 230 kV y otro a 138 kV, tiene 38 líneas de transmisión, 5 transformadores, 32 unidades de generación, 17 cargas, un reactor y un condensador síncrono.

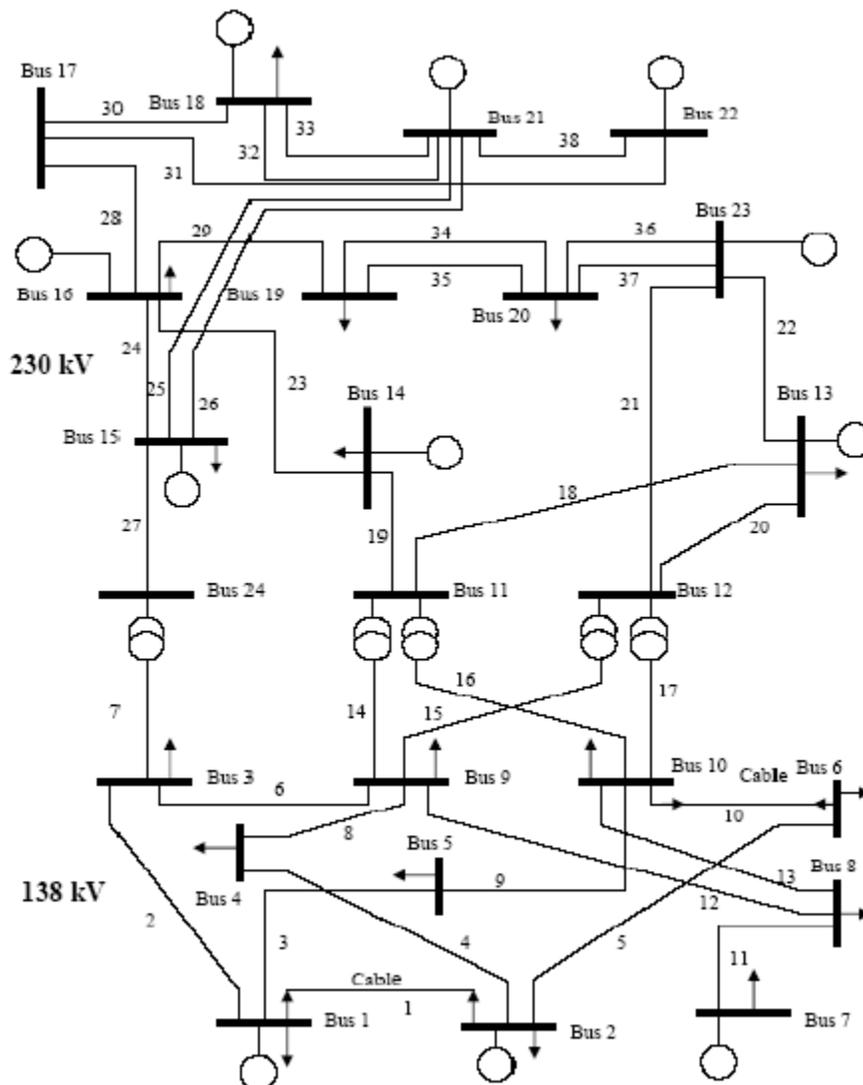


Figura 10. Sistema IEEE de 24 buses RTS.

Se plantean tres casos de estudios utilizando el sistema IEEE RTS de 24 buses (Figura 10).

Caso 1.- Este caso consiste en analizar la Carga de Referencia del Cliente (CBL por sus siglas en inglés) a nivel sistema, es decir, se obtuvo la suma de todas demandas de todos los nodos del sistema para realizar una estimación y obtener los parámetros del sistema.

Caso 2.- En el segundo caso de estudio se analizó el CBL a nivel región, por lo cual se dividió el sistema en dos regiones. La primera región contiene todos los nodos que sean alimentados con un voltaje de 138kV, mientras que la segunda región contiene todos los nodos que sean alimentados con un voltaje de 230kV. Así como en el caso anterior se realizó, para este caso se obtuvo la estimación de la demanda y los parámetros del sistema, pero basados en la suma de las demandas de los nodos de cada región.

Caso 3.- En el último caso de estudio se analizó el CBL a nivel nodal, por lo que, se realizaron diecisiete estimaciones con los datos de demanda en aquellos nodos donde se tiene carga y se obtuvieron los valores de los parámetros del sistema.

5.2. Caso 1.- CBL a nivel sistema

Los datos de demanda eléctrica del sistema se obtuvieron como resultado de la suma de los datos de demanda de cada uno de los nodos, obteniendo así 120 registros de los cinco días de la semana, los cuales, se pueden observar en la Tabla 12 de la sección Apéndice.

El patrón de demanda obtenido con los dos modelos y los datos esperados se puede observar en la Figura 11, en la cual se puede observar que el patrón estimado con el modelo ARIMA está muy por debajo en comparación con los datos reales. El CBL obtenido con el RPG tiene un comportamiento similar a los datos reales, con la diferencia máxima de 1306.534936 [Wh] en la hora 12:00, lo que representa un error de 0.42%.

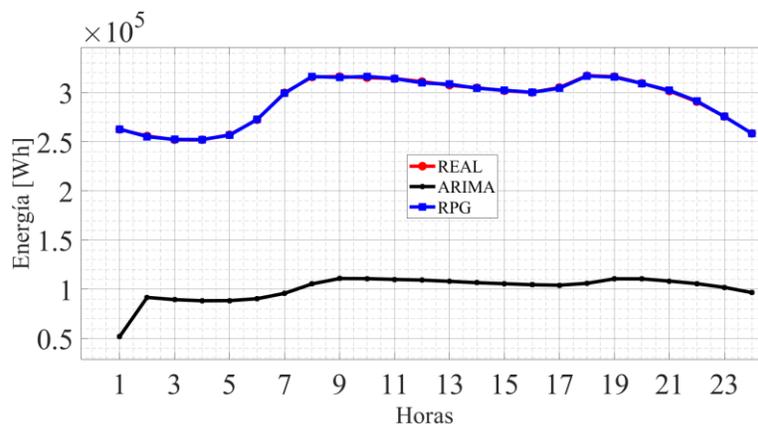


Figura 11. Estimación de la demanda a nivel sistema.

En la Tabla 6 se presentan los indicadores de rendimiento para este caso. Se aprecia que el modelo RPG tiene valores menores de RMSE, MSE y MAE en comparación con el modelo ARIMA. Por lo que, a primera instancia, se puede decir que las estimaciones hechas con el modelo propuesto son mucho más exactas que las estimaciones hechas con el modelo ARIMA.

Tabla 6. Indicadores de rendimiento a nivel sistema

RMSE		MSE		MAE	
ARIMA	RPG	ARIMA	RPG	ARIMA	RPG
$2.6132 * 10^{-4}$	$2.3280 * 10^{-4}$	$6.8288 * 10^{-8}$	$5.4198 * 10^{-8}$	$2.3326 * 10^{-4}$	$2.0729 * 10^{-4}$

La Figura 12 representa el comportamiento gráfico del coeficiente de determinación, donde se puede observar que la dispersión de los datos estimados con el modelo RPG es muy poca, ya que se encuentran cercanos a la línea diagonal (predicción perfecta).

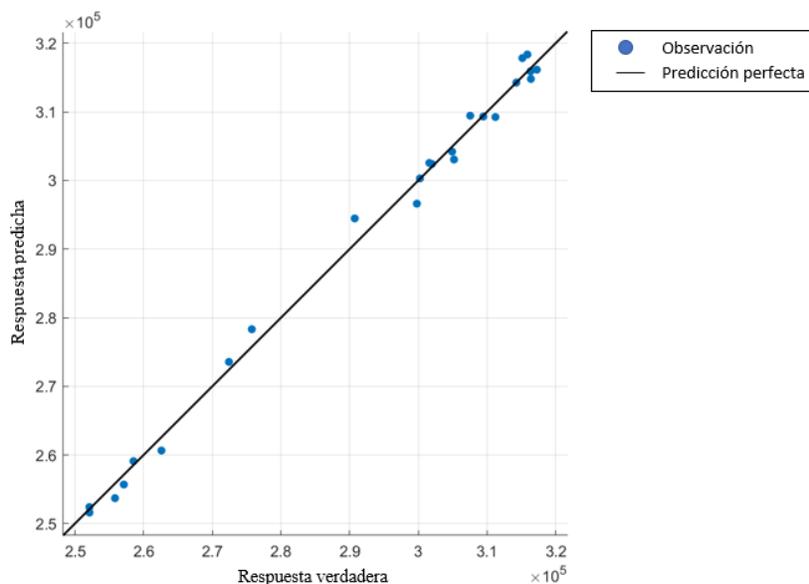


Figura 12. Coeficiente de determinación de la estimación del modelo RPG a nivel sistema

Con base en los resultados mostrados de las estimaciones de los modelos y los indicadores de rendimiento, se comprueba que el modelo RPG obtiene una mejor estimación en comparación con el modelo ARIMA. Por consiguiente, se prosiguió aplicar la herramienta OPF a la estimación hecha con el modelo RPG y a los datos reales, con la finalidad de estimar los parámetros técnico-económicos del sistema y observar la disparidad entre los datos de ambas estimaciones.

Dado el patrón de demanda estimado, se obtuvo la potencia generada por cada uno de los generadores a lo largo del periodo de 24 horas. Las estimaciones de potencia generada muestran que existe una generación de energía continua a lo largo de todo el horizonte de tiempo, como se muestra en las figuras 13 y 14.

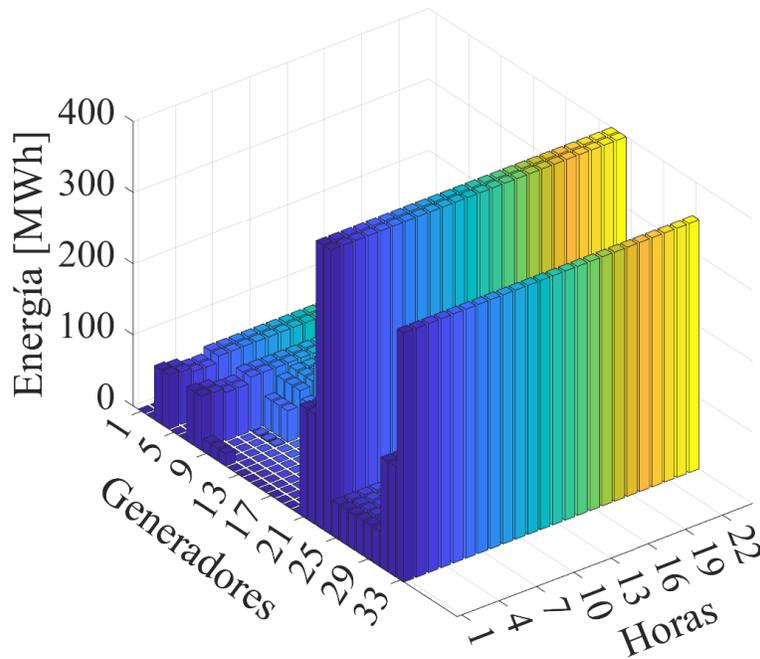


Figura 13. Potencia activa estimada con los datos del modelo RPG a nivel sistema

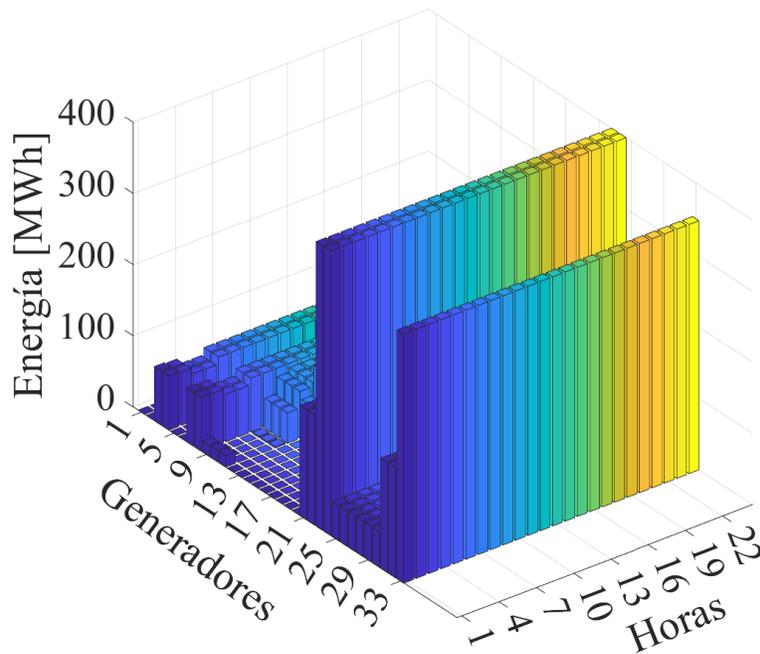


Figura 14. Potencia activa estimada con los datos reales a nivel sistemas

En las figuras 15 y 16, se puede observar un corte enfocado en la relación de energía-generadores de las dos figuras anteriores. Debido a que el comportamiento de la estimación de los dos conjuntos de datos es muy semejante, la generación de ambas estimaciones difiere muy poco, siendo la diferencia maximas entre las dos

estimaciones de 4.13 [MWh] durante la hora 10:00 en los generadores 12, 13 y 14, lo que representa un error de 4.31%.

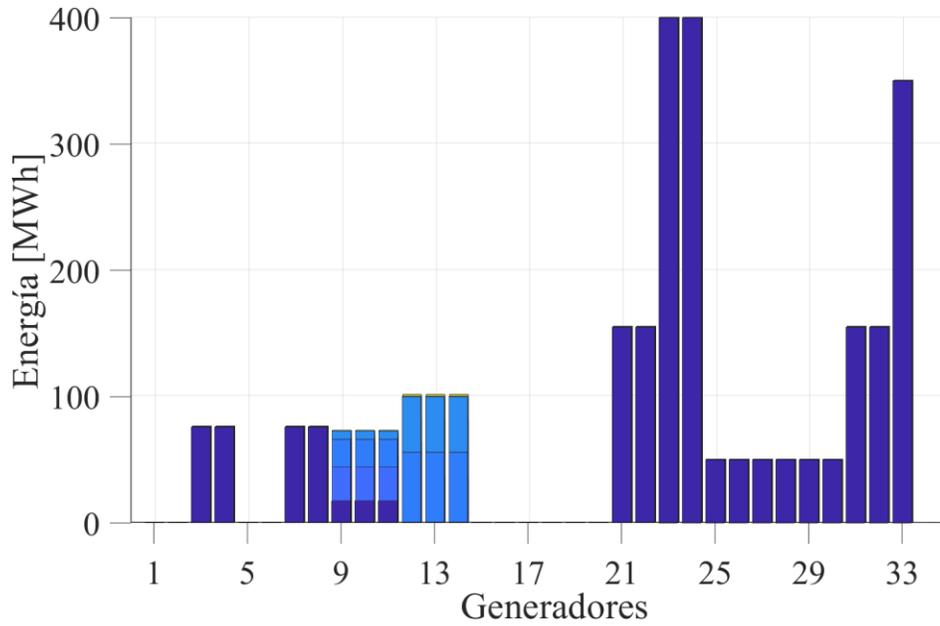


Figura 15. Relación energía-generadores (RPG) a nivel sistema

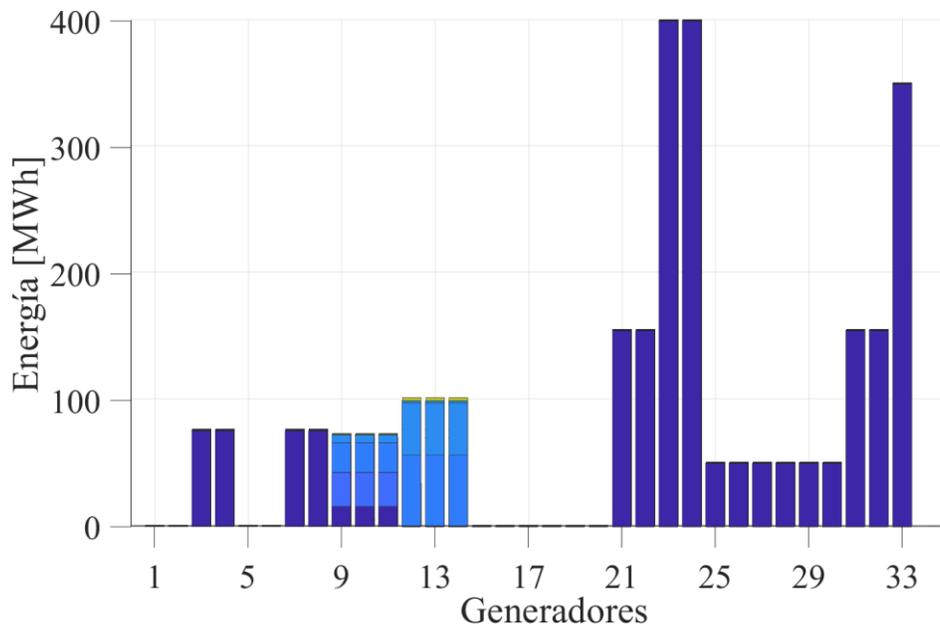


Figura 16. Relación energía-generadores (datos reales) a nivel sistema

Los costos operativos a nivel sistema se muestran en la Figura 17. El patrón estimado de costos operativos sigue fielmente al patrón estimado con los datos reales, teniendo una diferencia máxima de

714.9760404 [\$MWh] en la hora 10:00 y un error de estimación menor a 1.3% a lo largo del horizonte de 24 horas.

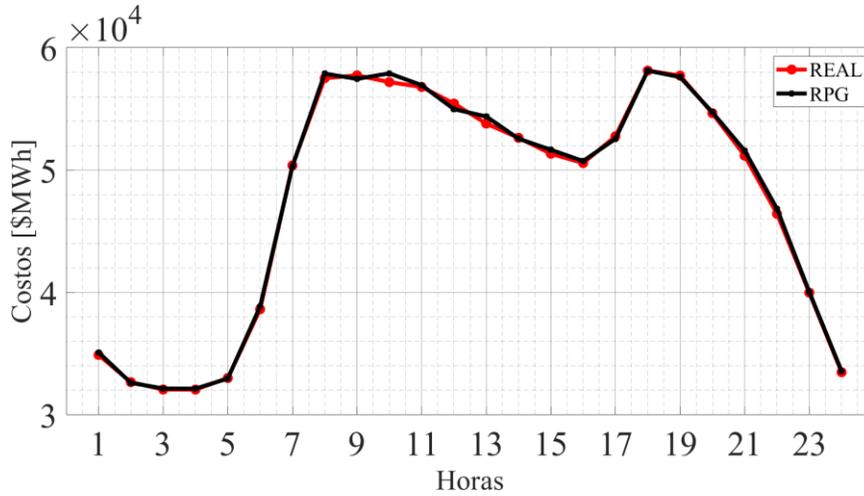


Figura 17. Costos operativos estimados a nivel sistema

Las estimaciones de los costos marginales mostrados en las figuras 18 y 19, siguen las mismas características que los dos parámetros técnico-económicos anteriores, es decir, para cada uno de los nodos existe un valor de costo marginal variante a lo largo del intervalo de tiempo, y dicho valor es muy similar para ambas estimaciones.

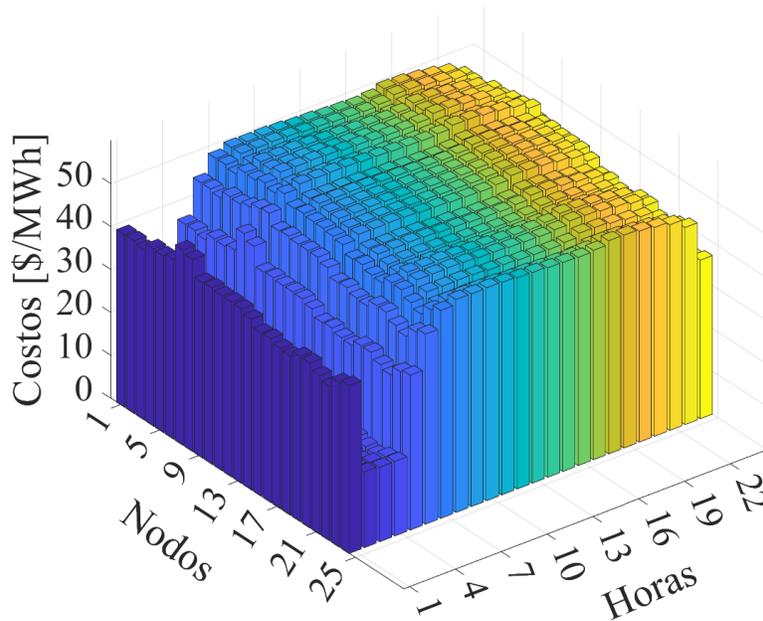


Figura 18. Costos marginales estimado a nivel sistema con RPG

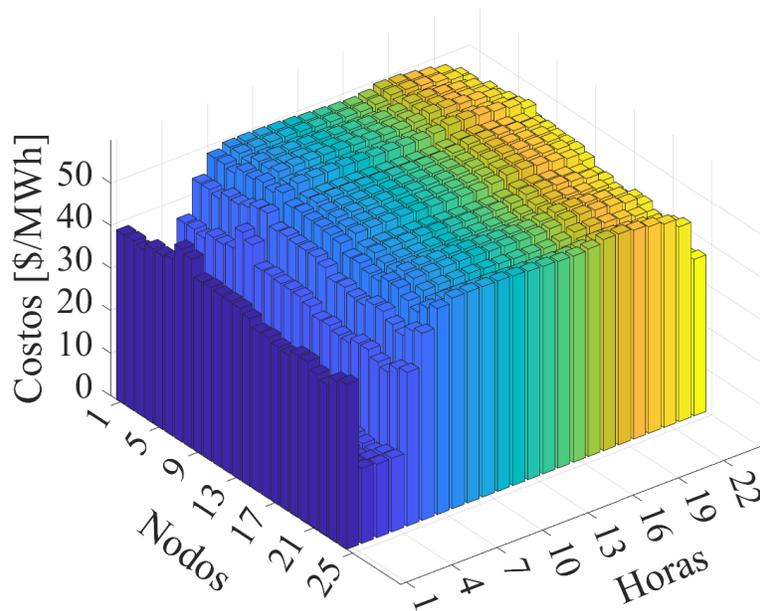


Figura 19. Costos marginales estimado a nivel sistema con los datos reales

En las figuras 20 y 21, se puede observar un corte enfocado en la relación de costos-nodos de las dos figuras anteriores. Considerando que ambas estimaciones tienen comportamientos similares, la diferencias entre los datos va a ser poca, siendo la diferencia máxima de 0.312191016 [\$/MWh] durante la hora 6:00 en el nodo seis, lo que representa un error de 0.66%.

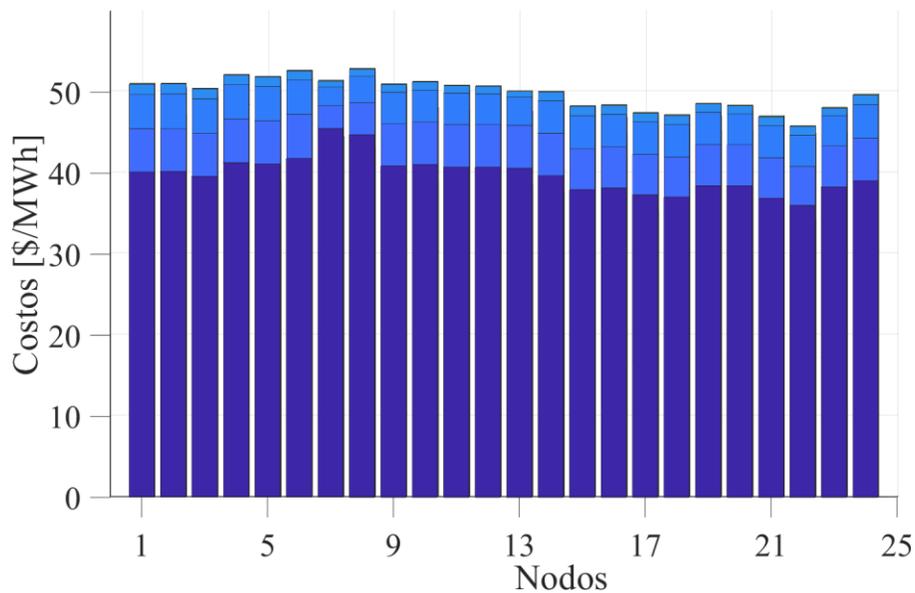


Figura 20. Relación costos-nodos (RPG) a nivel sistema

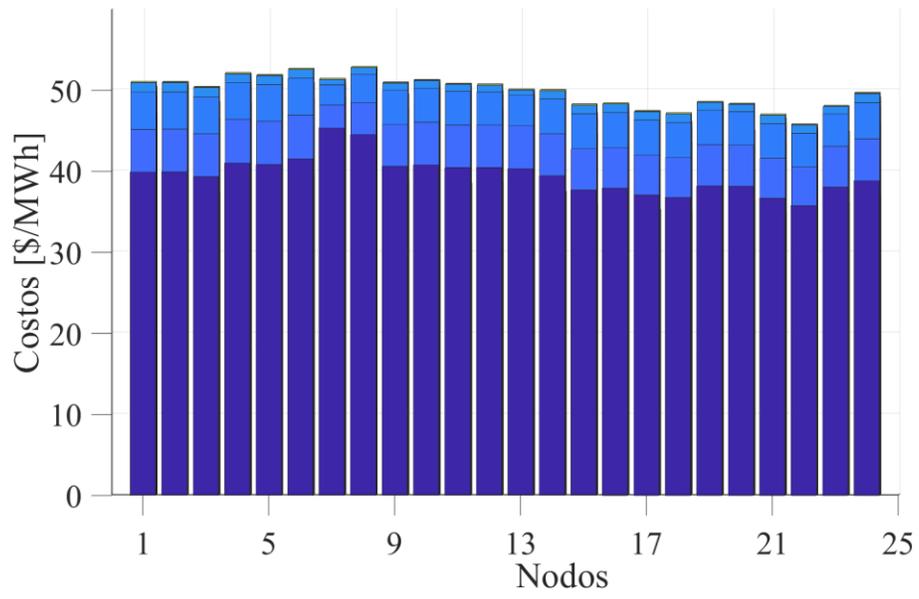


Figura 21. Relación costos-nodos (datos reales) a nivel sistema

5.3. Caso 2.- CBL a nivel región (varios nodos agrupados)

En este nivel se dividió el sistema en dos secciones de acuerdo a su nivel de voltaje como se muestra en la Figura 22, se tienen diez nodos en la región de 138kV (cuadrado azul) y catorce nodos en la región de 230kV (cuadrado verde). De la primera región (138kV) todos los nodos tienen carga, mientras que en la segunda región (230kV), tiene siete cargas ubicadas entre los nodos trece a dieciséis y los nodos dieciocho a veinte.

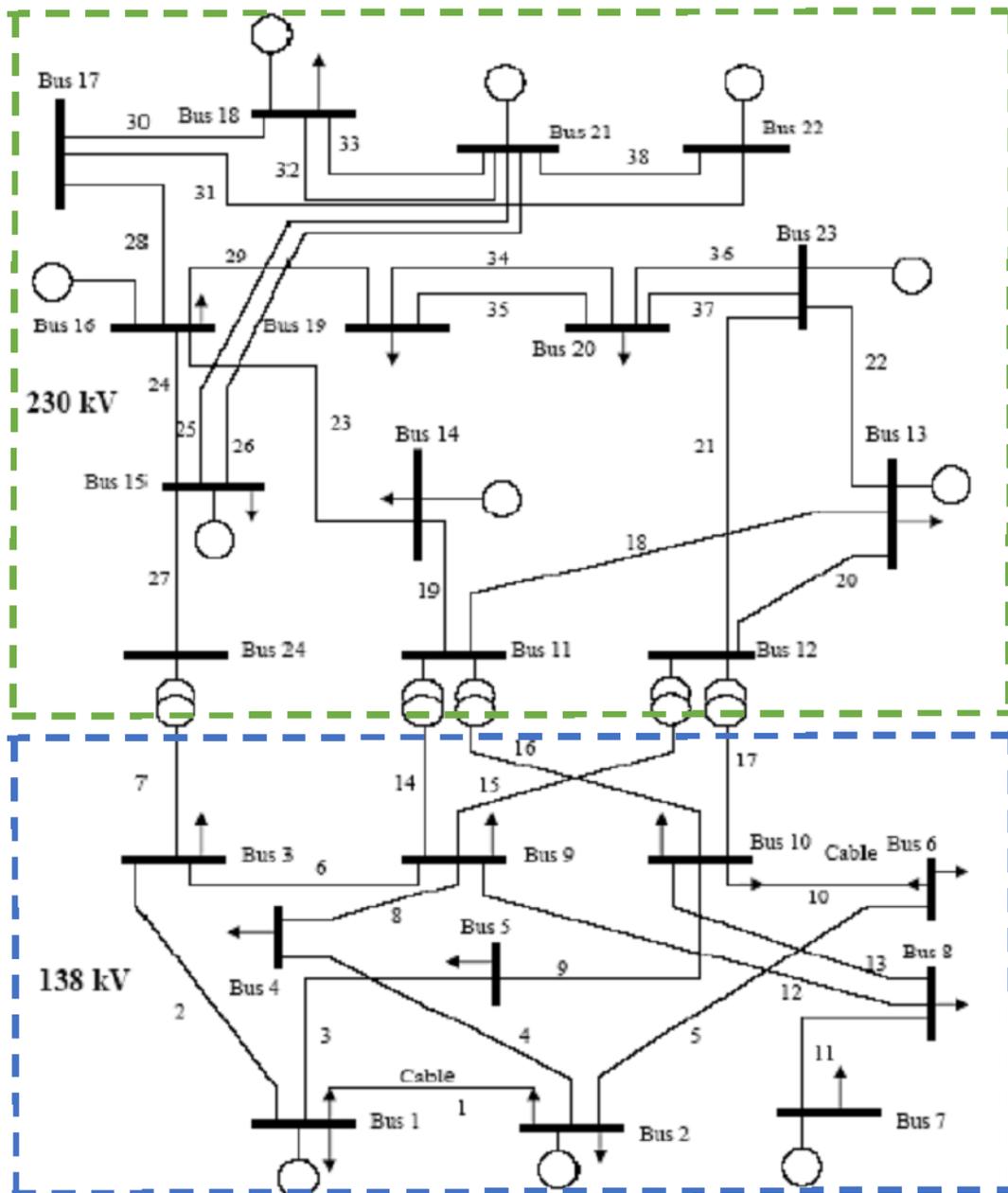


Figura 22. RTS de 24 nodos dividido por regiones.

Para obtener la demanda de cada región, se sumaron las demandas por hora de cada una de los nodos que pertenecen a la región. En las tablas 13 y 14 de la sección Apéndice, se tienen los datos de demanda eléctrica por hora de la región con un nivel de voltaje de 138kV y los datos de demanda eléctrica de la región con un nivel de voltaje de 230kV, respectivamente.

Con los datos de las tablas antes mencionadas, se obtuvieron los patrones de demanda por región. En la Figura 23 se pueden observar el patrón de demanda de la primera región. El patrón obtenido con ambos modelos sigue casi en su totalidad la trayectoria de los datos esperados, por lo que en la Figura 24, se prosiguió a realizar un acercamiento entre las horas 7:00 y 13:00, donde se logra visualizar que la estimación con el modelo RPG sigue más de cerca la trayectoria de los datos reales que la otra estimación.

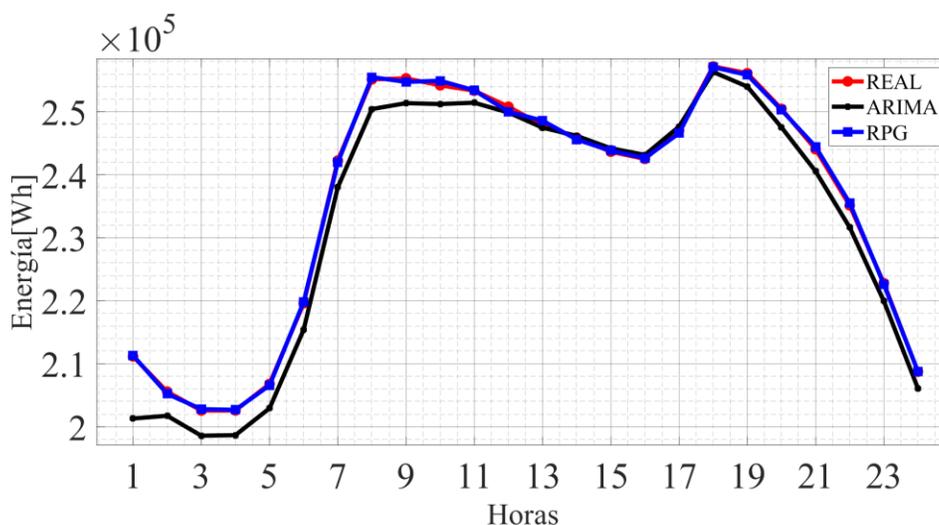


Figura 23. Estimación de la demanda de la región 1.

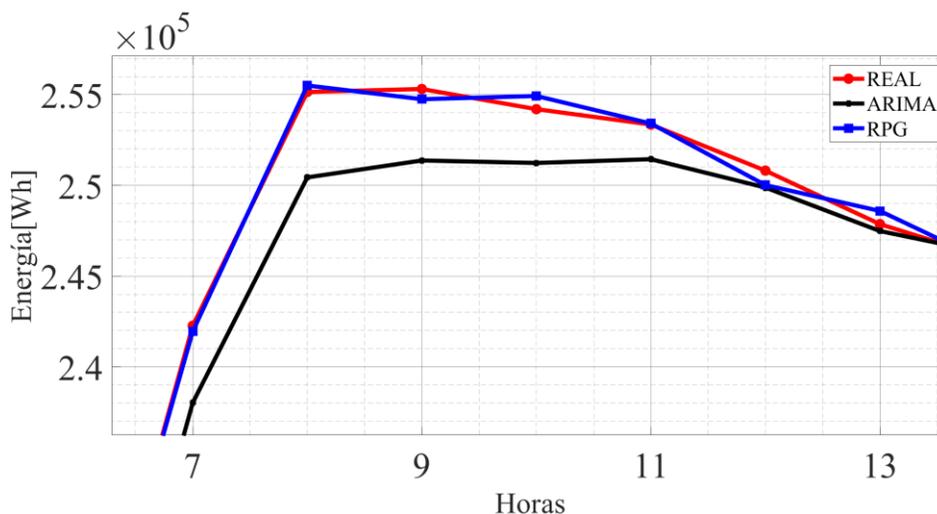


Figura 24. Estimación de la demanda de la región 1 entre la hora 7:00 y la hora 13:00

Por otro lado, en la Figura 25 se tiene el patrón de demanda de la segunda región. Al igual que en las estimaciones hechas para la región uno, para la región dos el comportamiento de ambas estimaciones sigue casi en su totalidad a la respuesta real; sin embargo, la trayectoria de la estimación del modelo RPG tiene menores diferencias entre estos datos y los datos reales, siendo la diferencia máxima de 285.0429972 [Wh] en la hora 19:00.

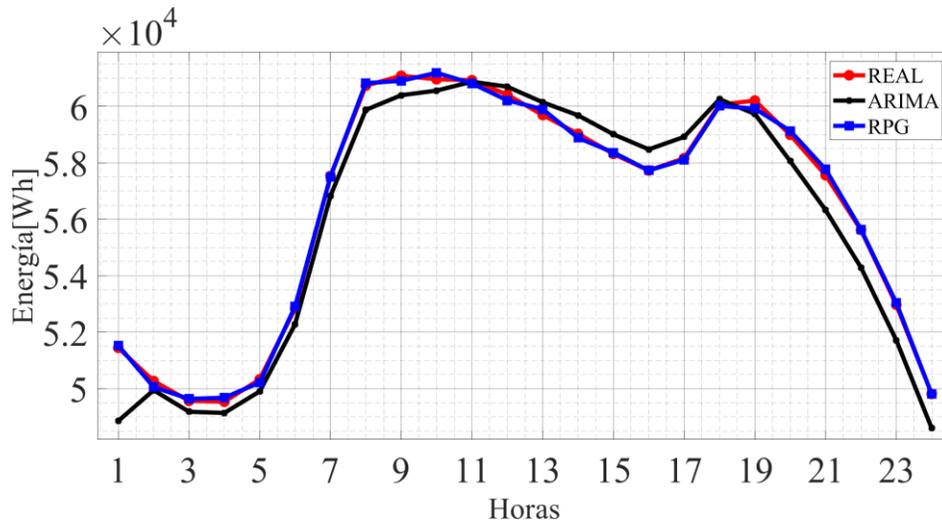


Figura 25. Estimación de la demanda de la región 2

Con la finalidad de apreciar de mejor la diferencia entre las estimaciones de los modelos y los datos reales, se tomó el periodo de tiempo comprendido entre las horas 13:00 a 21:00, y se realizó un acercamiento como se observa en la Figura 26.

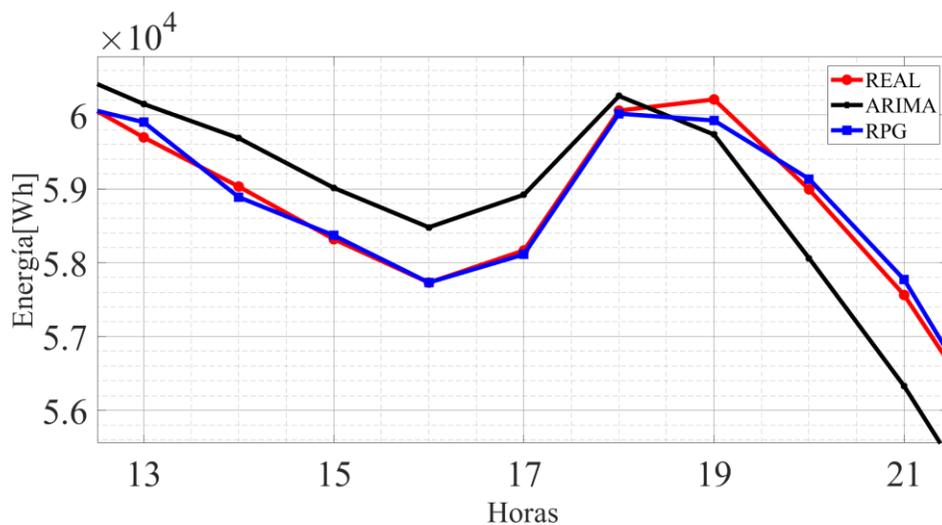


Figura 26. Estimación de la demanda de la región 2 entre la hora 13:00 y la hora 21:00

Se obtuvieron los indicadores de rendimiento para este par de estimaciones, los cuales se pueden observar en la Tabla 7, notando que el modelo propuesto es mucho mejor que el modelo tradicional, ya que sus índices de rendimiento son menores.

Región	RMSE		MSE		MAE	
	ARIMA	RPG	ARIMA	RPG	ARIMA	RPG
1	$2.0805 * 10^{-4}$	$1.9148 * 10^{-4}$	$4.3284 * 10^{-8}$	$3.6664 * 10^{-8}$	$1.8626 * 10^{-4}$	$1.7007 * 10^{-4}$
2	$4.5504 * 10^{-5}$	$4.1616 * 10^{-5}$	$2.0706 * 10^{-9}$	$1.7319 * 10^{-9}$	$4.0531 * 10^{-5}$	$3.7219 * 10^{-5}$

Tabla 7. Indicadores de rendimiento por región

Los coeficientes de determinación de las estimaciones hechas con el modelo RPG se muestran en las figuras 27 y 28. Como se puede observar en la Figura 27, la dispersión de los datos es poca, por lo consiguiente los datos estimados se encuentran muy cerca de la predicción perfecta (línea diagonal).

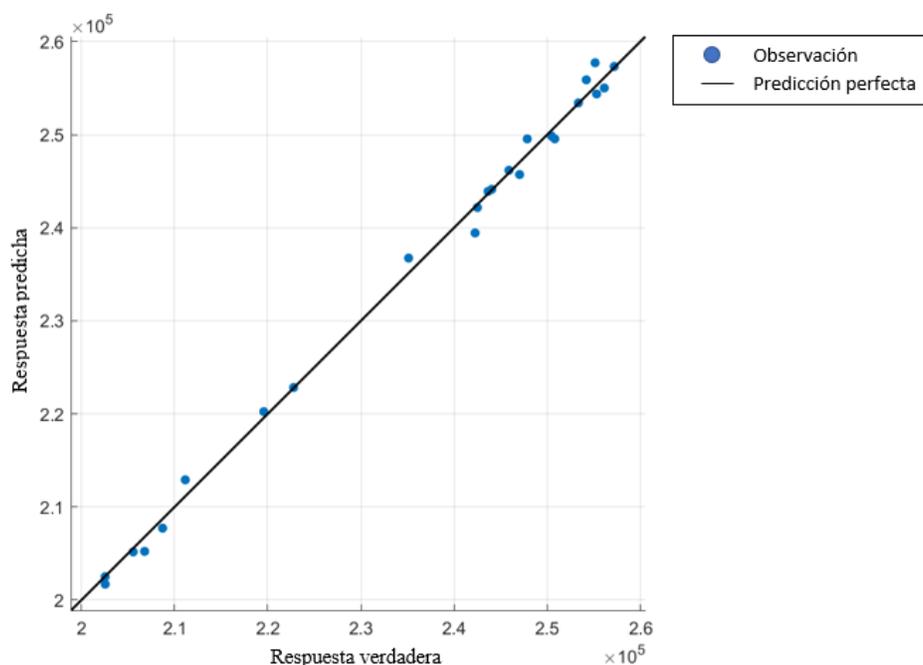


Figura 27. Coeficiente de determinación de la estimación del modelo RPG para la región 1

En el caso del coeficiente de determinación de la región dos, ilustrado en la Figura 28, se pueden observar seis datos que se encuentran un poco separadas; sin embargo, el resto de datos se encuentran muy cerca de la predicción perfecta, por lo tanto, se puede considerar que la estimación hecha con el modelo RPG

es buena, debido a que la relación de los datos con la línea de la predicción perfecta es menor a 0.2 en ambos ejes.

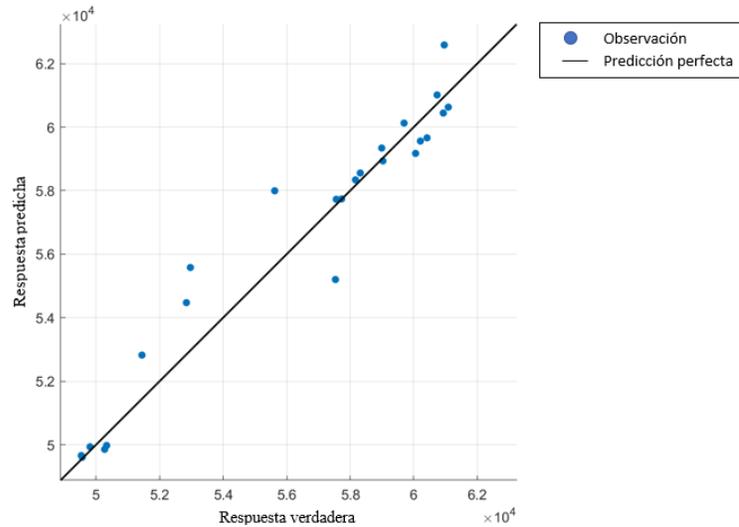


Figura 28. Coeficiente de determinación de la estimación del modelo RPG para la región 2

Como en el caso anterior, con base en el comportamiento de las estimaciones y los indicadores de rendimiento, se demostró que el modelo RPG realiza mejores estimaciones que el modelo ARIMA. Por lo mismo, se consideró aplicar la herramienta OPF solo en los datos reales y en los datos estimados el uso del modelo RPG, con el fin de obtener estimaciones de los parámetros del sistema y así, analizar las diferencias entre ambos resultados.

Las estimaciones de potencia generada muestran que existe una generación de energía continua a lo largo de todo el horizonte de tiempo, como se muestra en las figuras 29 y 30.

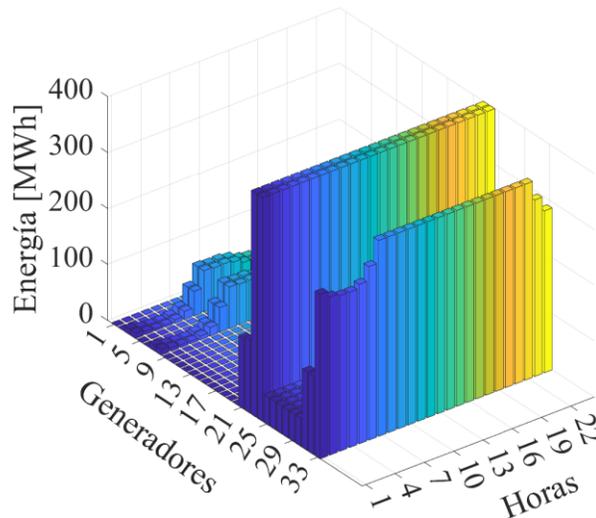


Figura 29. Potencia activa estimada con los datos del modelo RPG a nivel regional

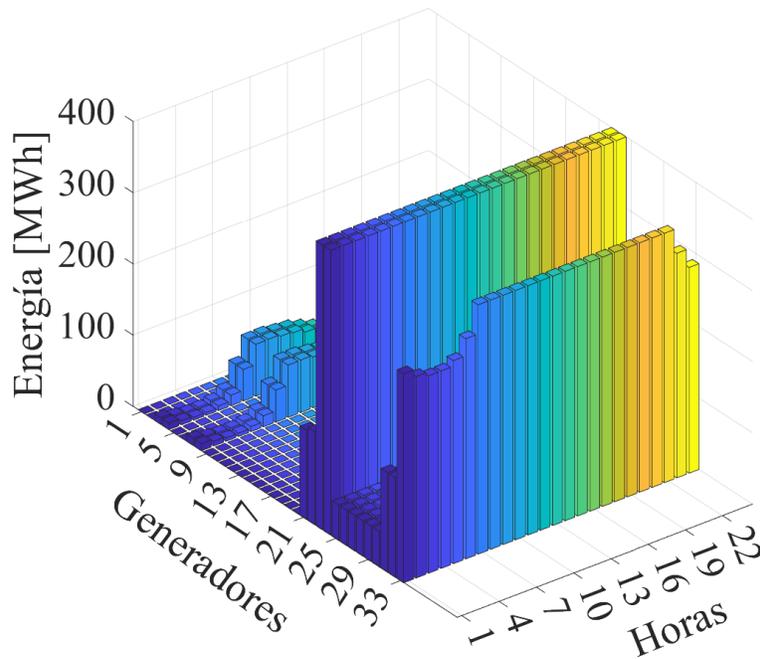


Figura 30. Potencia activa estimada con los datos reales a nivel regional

En las figuras 31 y 32, se puede observar un corte enfocado en la relación de energía-generadores de las dos figuras anteriores. Debido a que el comportamiento de la estimación de los dos conjuntos de datos es muy semejante, la generación de ambas estimaciones difiere muy poco, siendo la diferencia máximas entre las dos estimaciones de 2.154283897 [MWh] durante la hora 10:00 en los generadores 3 y 4, lo que respresenta un error de 3.04%.

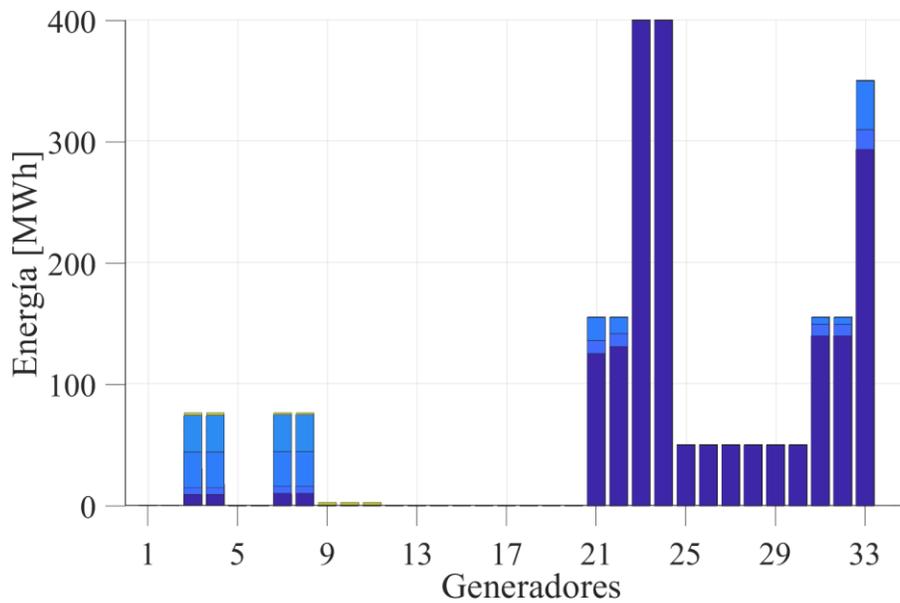


Figura 31. Relación energía-generadores (RPG) a nivel regional

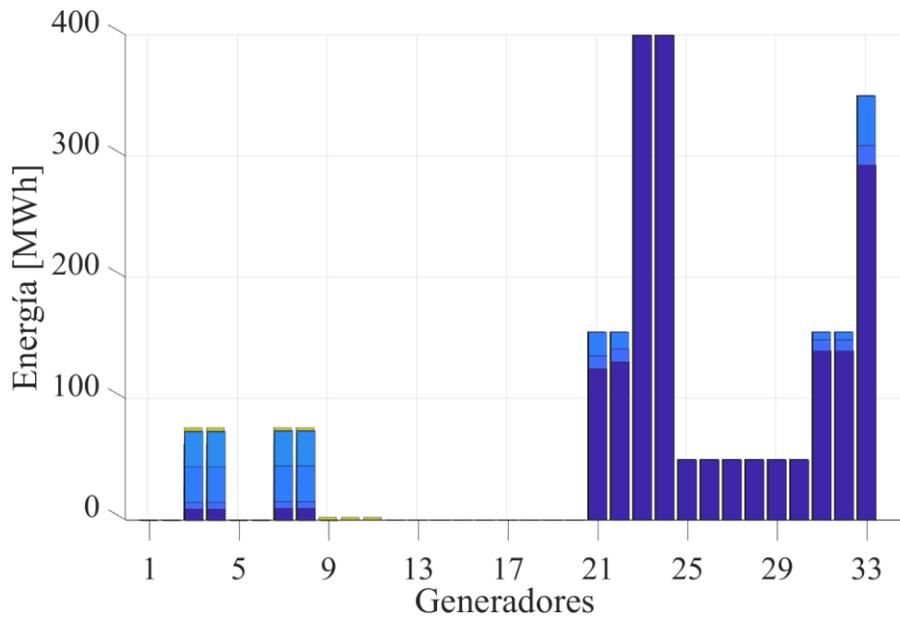


Figura 32. Relación energía-generadores (datos reales) a nivel regional

Los costos operativos a nivel región se muestran en la Figura 33. El patrón estimado de costos operativos sigue fielmente al patrón estimado con los datos reales, teniendo una diferencia máxima de 155.8788175 [\$MWh] en la hora 10:00 y un error de estimación menor al 0.5%.

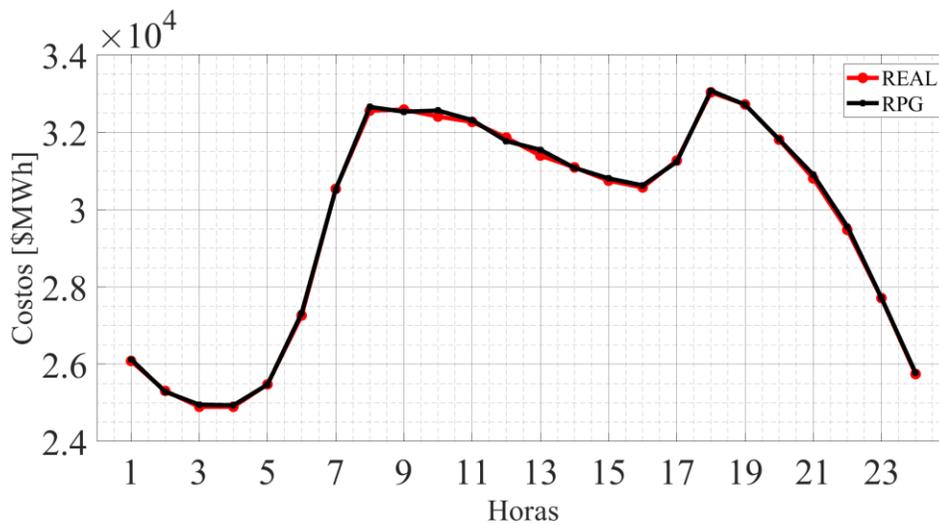


Figura 33. Costos operativos estimados a nivel región

Al igual que el caso anterior, las estimaciones de los costos marginales mostrados en las figuras 34 y 35, siguen las mismas características que los dos parámetros técnico-económicos anteriores, por lo consiguientes, las estimaciones de ambos costos son muy similares y son continuos a lo largo del horizonte de tiempo.

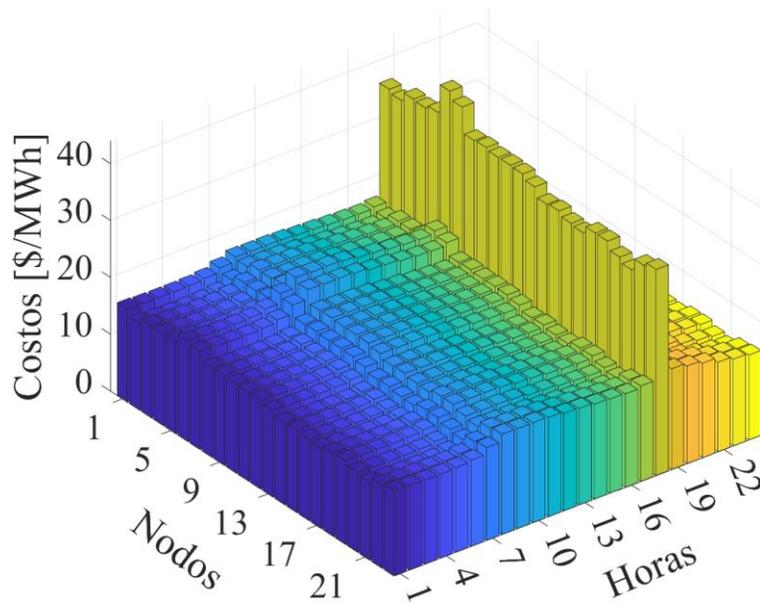


Figura 34. Costos marginales estimado a nivel región con RPG

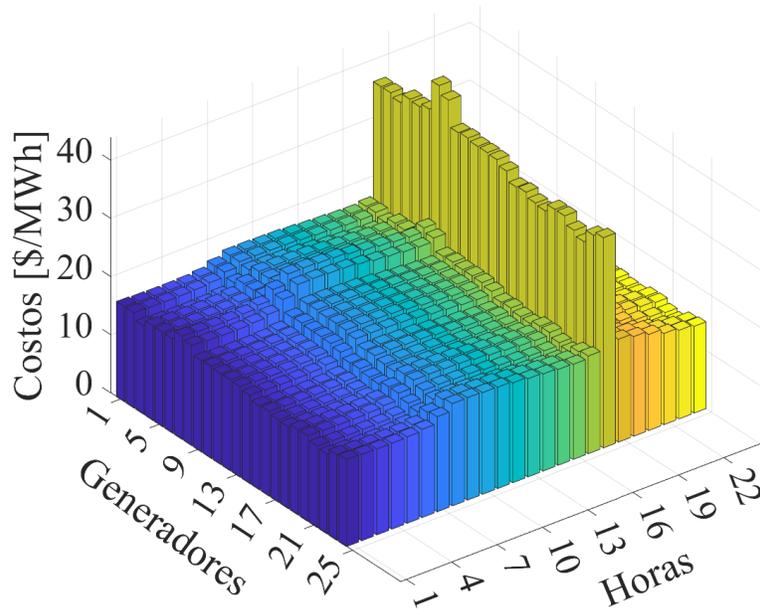


Figura 35. Costos marginales estimado a nivel región con los datos reales

En las figuras 36 y 37, se puede observar un corte enfocado en la relación de costos-nodos de las dos figuras anteriores. Considerando que ambas estimaciones tienen comportamientos similares, la diferencias entre los datos va ser poca, siendo la diferencia máxima de 0.127529464 [\$/MWh] durante la hora 7:00 en el nodo diez, lo que representa un error de 0.6%.

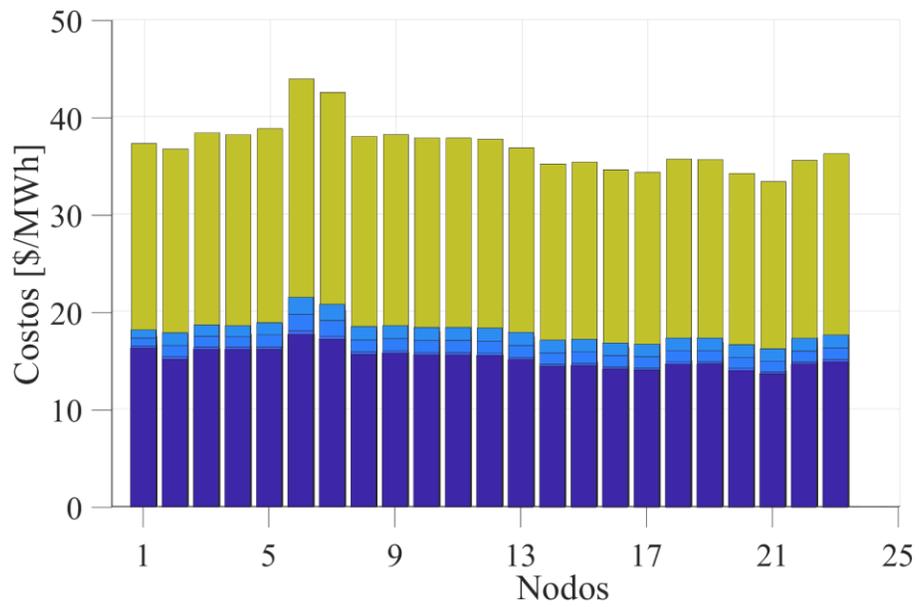


Figura 36. Relación costos-nodos (RPG) a nivel región

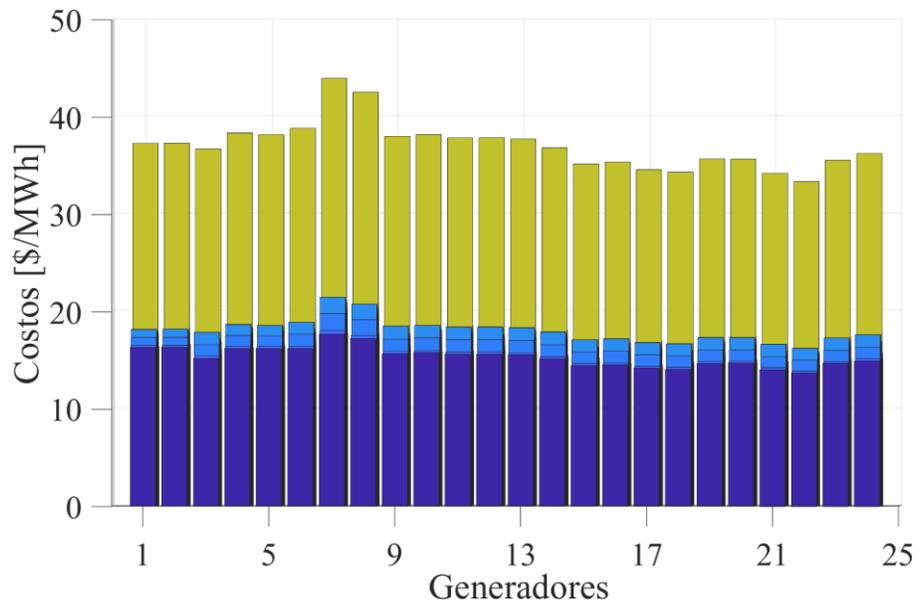


Figura 37. Relación costos-nodos (datos reales) a nivel región

5.4. Caso 3.- CBL a nivel nodo

Finalmente, en este nivel se consideró la demanda individual de cada uno de los nodos del sistema; sin embargo, para resumir la predicción se seleccionaron tres nodos para evidenciar el comportamiento del CBL. Los nodos seleccionados son el nodo uno, el nodo dos y el nodo siete, cuyas tablas de datos de demanda se muestran en la sección Apéndice referenciadas con los números 15, 16 y 17, respectivamente.

Similar al caso anterior, las estimaciones hechas por ambos modelos tienen un comportamiento similar a los datos esperados. Las estimaciones del nodo uno se muestra en la Figura 38, en la cual se puede apreciar que el modelo RPG tiene un comportamiento más fiel al patrón de demanda estimado con el modelo ARIMA.

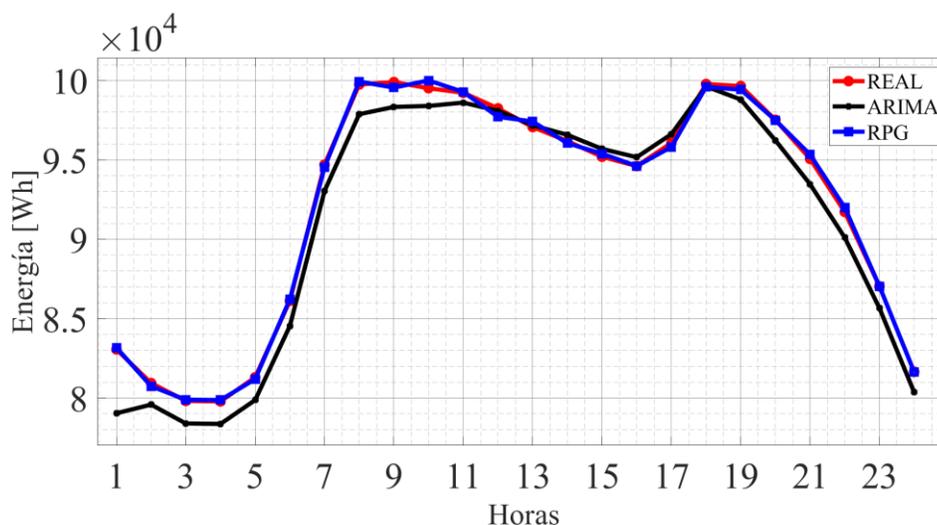


Figura 38. Estimación de la demanda del nodo 1

Realizando un acercamiento a la gráfica de la Figura 39 entre las horas 1:00 y 12:00, se puede observar con más facilidad como la gráfica obtenida mediante el modelo RPG tiene un comportamiento más cercano a los datos reales, teniendo la diferencia máxima entre los datos reales y los datos obtenidos con ARIMA de 4020.497903 [Wh] en la hora 1:00. Mientras que la diferencia máxima entre los datos del modelo RPG y los datos reales, es solo de 532.5462653[Wh] en la hora 12:00.

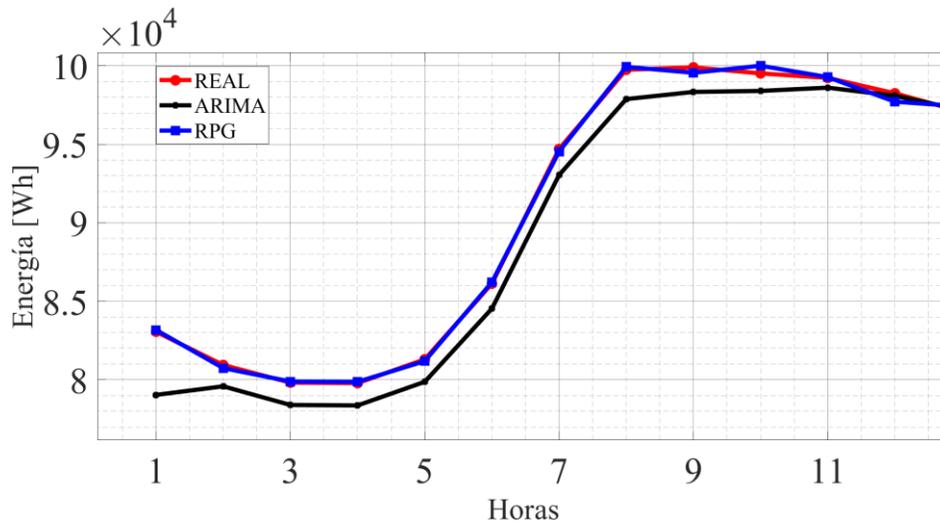


Figura 39. Estimación de la demanda del nodo 1 entre la hora 1:00 y la hora 12:00

Las estimaciones del nodo dos se pueden observar en la Figura 40, donde al igual que en la figura anterior es notorio que la estimación con el modelo RPG sigue con más exactitud el comportamiento de los datos reales. Esto se aprecia mejor en la Figura 41, en la cual se realiza un acercamiento entre las horas 1:00 y 12:00.

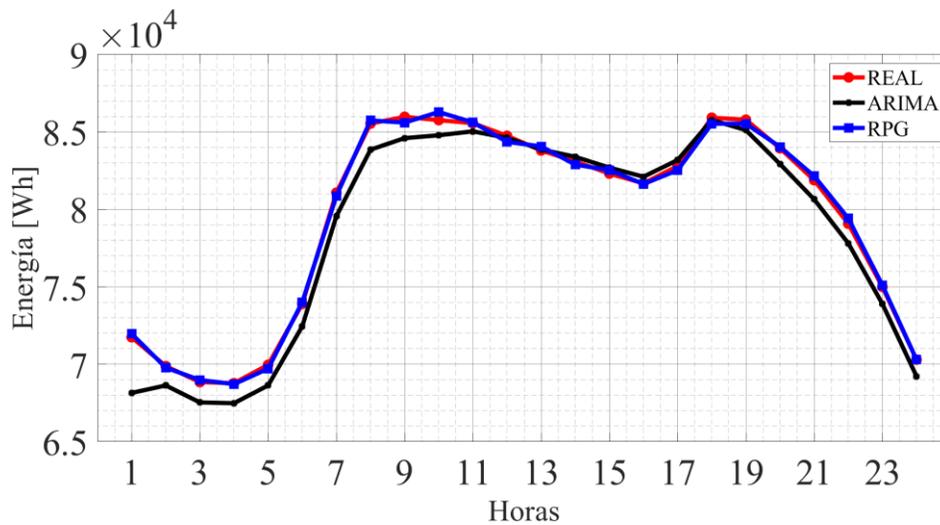


Figura 40. Estimación de la demanda del nodo 2

La diferencia máxima entre los datos estimados con el modelo ARIMA y los datos reales es de 3591.0518[Wh] en la hora 1:00, mientras que la diferencia máxima entre los datos estimados con el modelo RPG y los datos reales es de 524.2480159[Wh] en la hora 10:00.

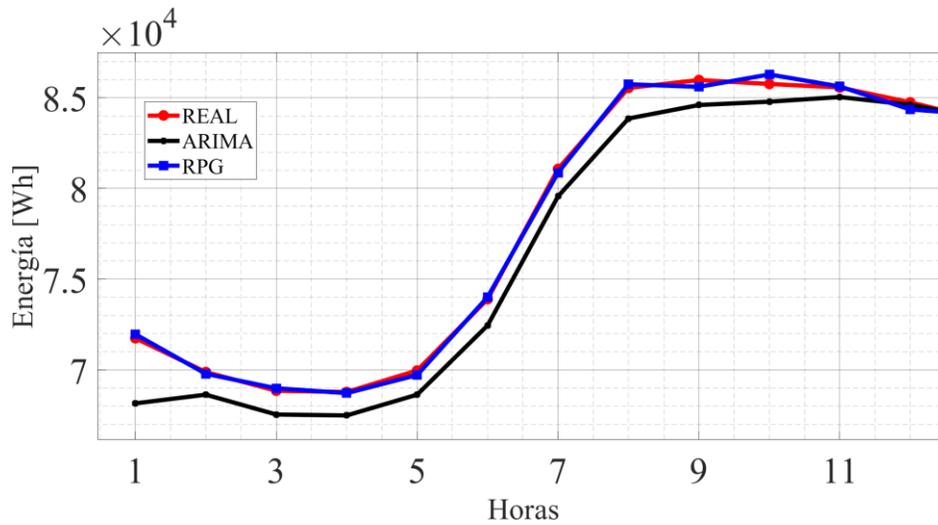


Figura 41. Estimación de la demanda del nodo 2 entre la hora 1:00 y la hora 12:00

Las estimaciones del nodo siete tuvieron un comportamiento similar a las estimaciones de los nodos anteriores, la estimación hecha con el modelo RPG es más fiel a los datos reales que la estimación hecha con el modelo ARIMA como se pueden observar en la Figura 42. Esto se aprecia mejor en la Figura 43, en la cual se realiza un acercamiento entre las horas 1:00 y 15:00.

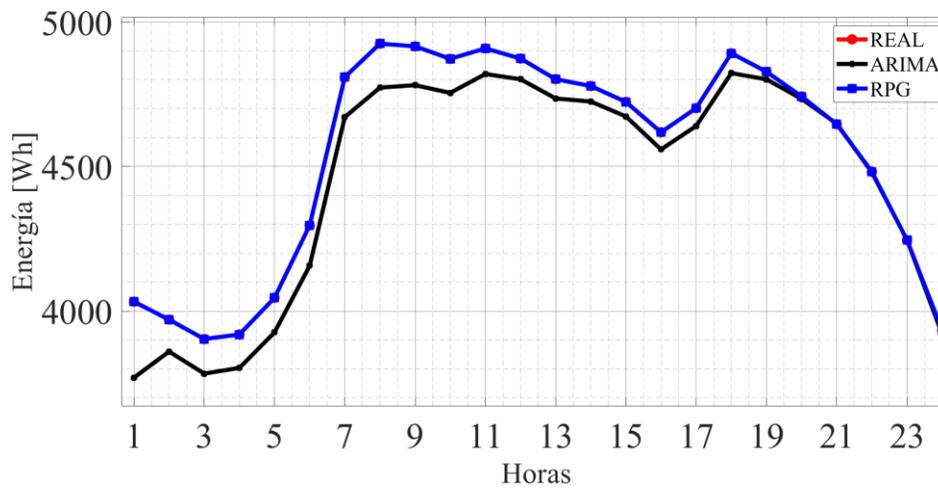


Figura 42. Estimación de la demanda del nodo 7

La diferencia máxima entre los datos estimados con el modelo ARIMA y los datos reales es de 262.5587818 [Wh] en la hora uno, mientras que la diferencia máxima entre los datos estimados con el modelo RPG y los datos reales es de 0.230961114 [Wh] en la hora 15:00.

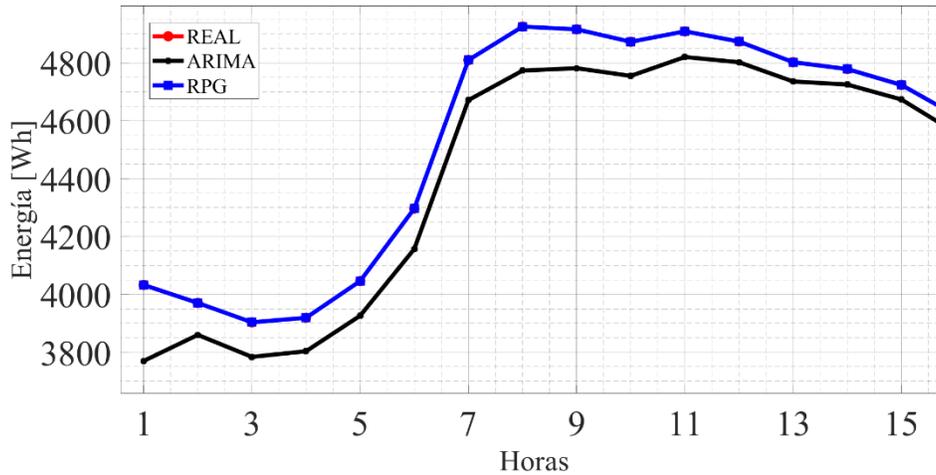


Figura 43. Estimación de la demanda del nodo 7 entre la hora 1:00 y la hora 15:00

Con la finalidad de asegurar que la metodología propuesta es mejor que la metodología tradicional, se obtuvieron un conjunto de indicadores de error como se menciona en la sección 2.9. En la Tabla 8 se pueden observar los tres indicadores usados para medir el rendimiento de los modelos.

Tabla 8. Indicadores de rendimiento por nodo

Nodos	RMSE		MSE		MAE	
	ARIMA	RPG	ARIMA	RPG	ARIMA	RPG
1	$7.7939 * 10^{-5}$	$7.3366 * 10^{-5}$	$6.0746 * 10^{-9}$	$2.2427 * 10^{-9}$	$6.9914 * 10^{-5}$	$6.5640 * 10^{-5}$
2	$6.3598 * 10^{-5}$	$6.3051 * 10^{-5}$	$4.0447 * 10^{-9}$	$1.6564 * 10^{-9}$	$5.7072 * 10^{-5}$	$5.6485 * 10^{-5}$
7	$3.9619 * 10^{-6}$	$3.7103 * 10^{-6}$	$1.5697 * 10^{-10}$	$5.7360 * 10^{-11}$	$3.5213 * 10^{-6}$	$3.3302 * 10^{-6}$

Los tres indicadores de rendimiento obtenidos con base en la estimación del modelo RPG son menores en comparación con los indicadores del modelo ARIMA.

El coeficiente de determinación de los nodos uno, dos y siete se pueden visualizar en las figuras 44, 45 y 46, respectivamente. El comportamiento de las tres gráficas es similar en los tres nodos, existe muy poca disparidad entre los datos reales y los estimados con el modelo RPG, por lo que las estimaciones se encuentran cerca de la línea diagonal (predicción perfecta).

Del coeficiente de determinación del nodo siete, el cual se puede observar en la Figura 46, resaltan dos puntos que se encuentran alejado de la predicción perfecta; sin embargo, la relación de los ejes para ambas observaciones tiene un valor aproximado de 1.06, con lo que aún se considera que la estimación es buena.

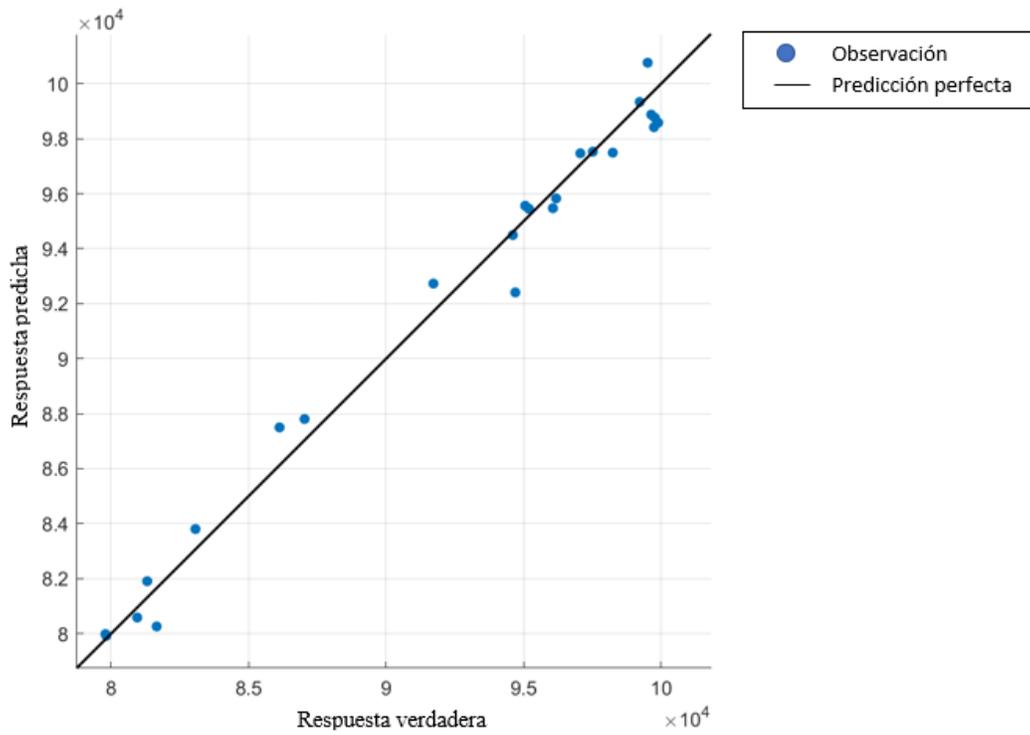


Figura 44. Coeficiente de determinación de la estimación del modelo RPG para el nodo 1.

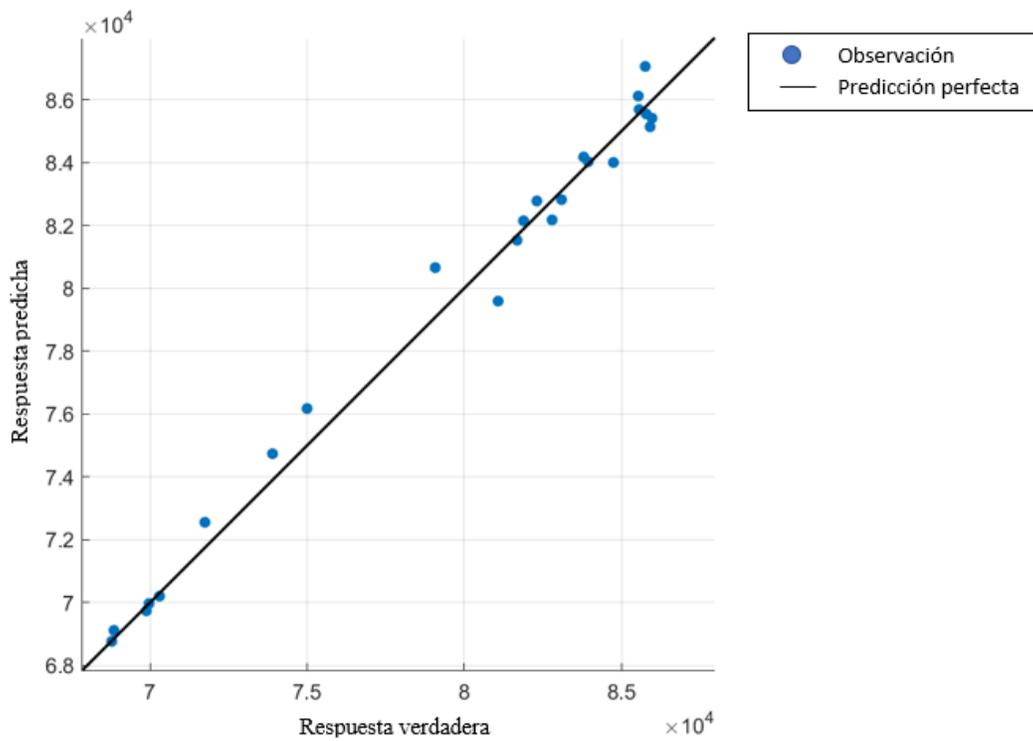


Figura 45. Coeficiente de determinación de la estimación del modelo RPG para el nodo 2

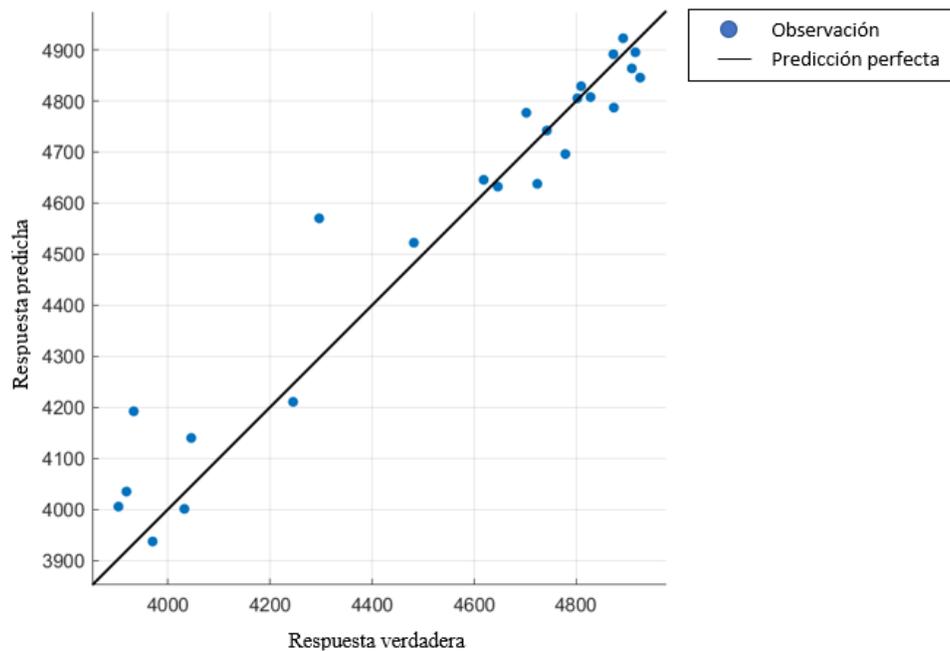


Figura 46. Coeficiente de determinación de la estimación del modelo RPG para el nodo 7

En este último caso al igual que en los dos anteriores, se encontró que el modelo RPG realiza mejores estimaciones que las realizadas con el modelo ARIMA. Esta conclusión fue tomada con base en la comparación hecha entre las respuestas gráficas de los modelos, los indicadores de rendimiento y los coeficientes de determinación obtenido para cada uno de los nodos seleccionados.

Las estimaciones de potencia generadas muestran que existe una generación de energía continua a lo largo de todo el horizonte de tiempo, como se muestra en las figuras 47 y 48.

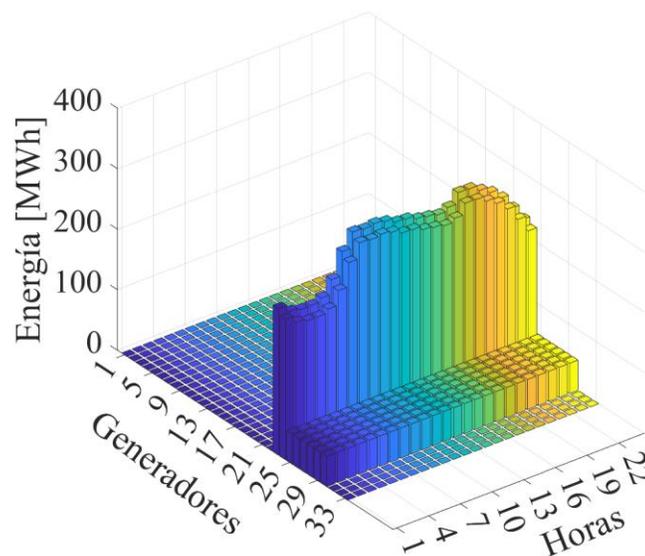


Figura 47. Potencia activa estimada con los datos del modelo RPG a nivel nodal

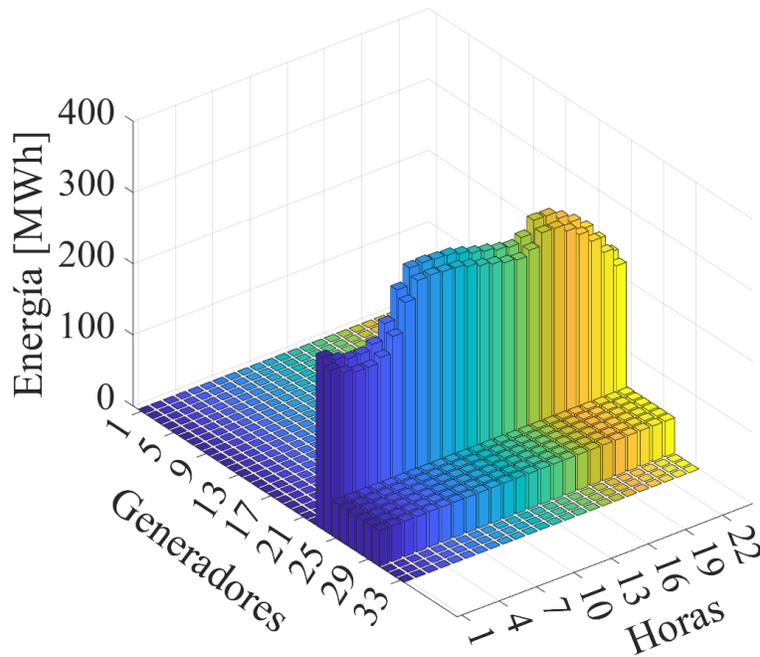


Figura 48. Potencia activa estimada con los datos reales a nivel nodal

En las figuras 49 y 50, se puede observar un corte enfocado en la relación de energía-generadores de las dos figuras anteriores. Debido a que el comportamiento de la estimación de los dos conjuntos de datos es muy semejante, la generación de ambas estimaciones difieren muy poco, siendo la diferencia máximas entre las dos estimaciones de 2.782568318 [MWh] durante la hora 10:00 en los generadores 24, lo que respresenta un error de 0.2%.

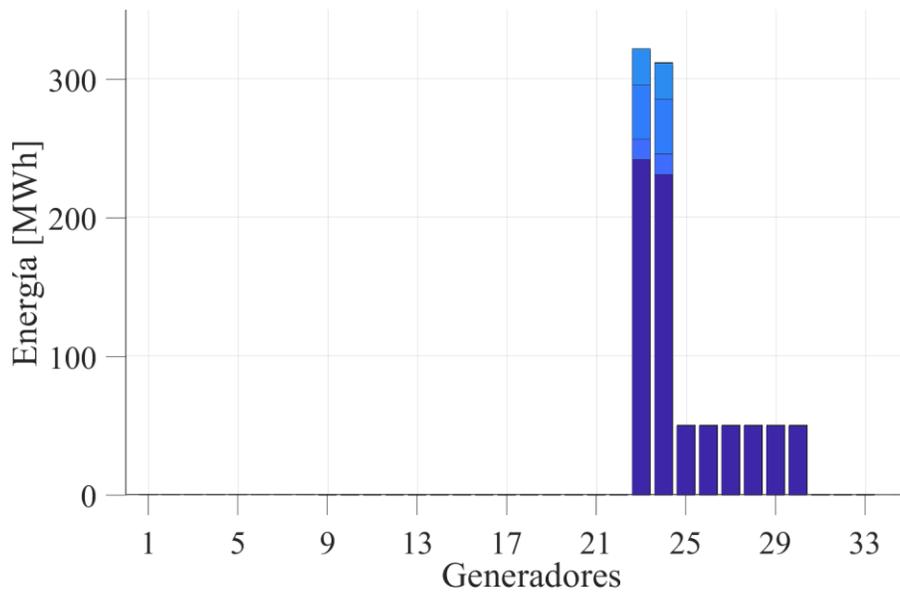


Figura 49. Relación energía-generadores (RPG) a nivel nodal

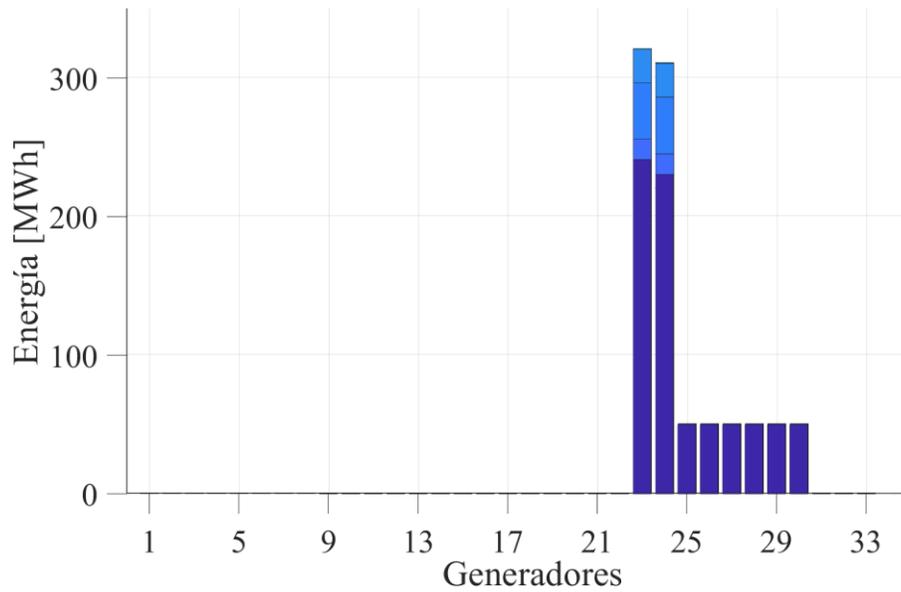


Figura 50. Relación energía-generadores (datos reales) a nivel nodal

Los costos operativos a nivel nodal se muestran en la Figura 51. El patrón estimado de costos operativos sigue fielmente al patrón estimado con los datos reales, teniendo una diferencia máxima de 25.24365822 [\$MWh] en la hora 10:00 y un error de estimación menor al 0.2%.

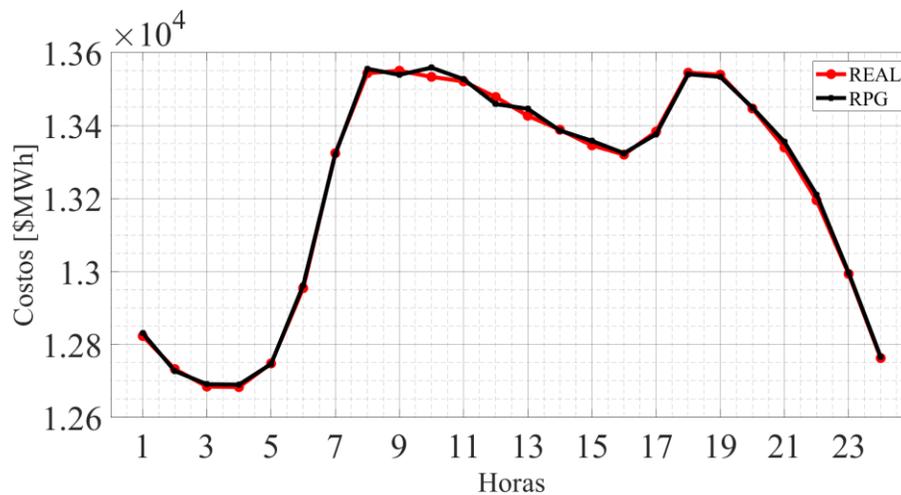


Figura 51. Costos operativos estimados a nivel nodal

Finalmente, las estimaciones de los costos marginales mostrados en las figuras 52 y 53, siguen las mismas características que los dos parámetros técnico-económicos anteriores, por lo consiguientes, las estimaciones de ambos costos son muy similares y son continuos a lo largo del horizonte de tiempo.

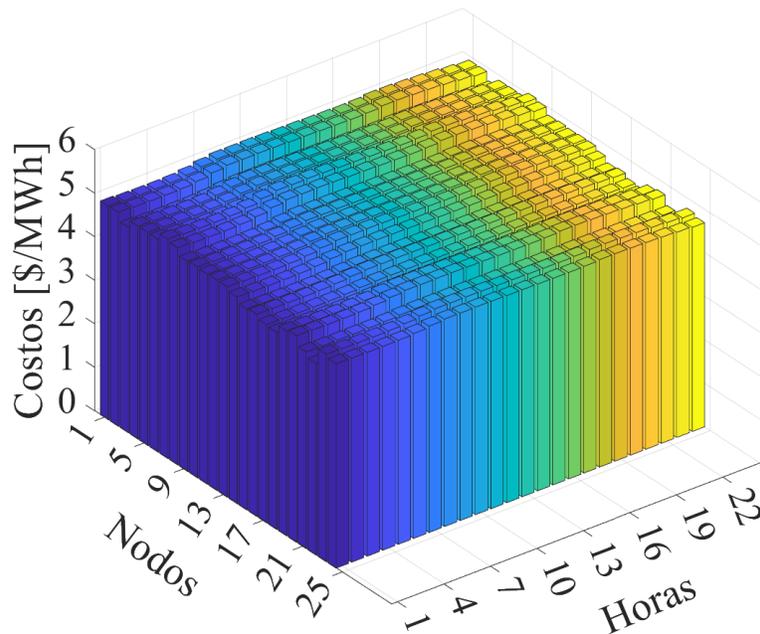


Figura 52. Costos marginales estimado a nivel nodal con RPG

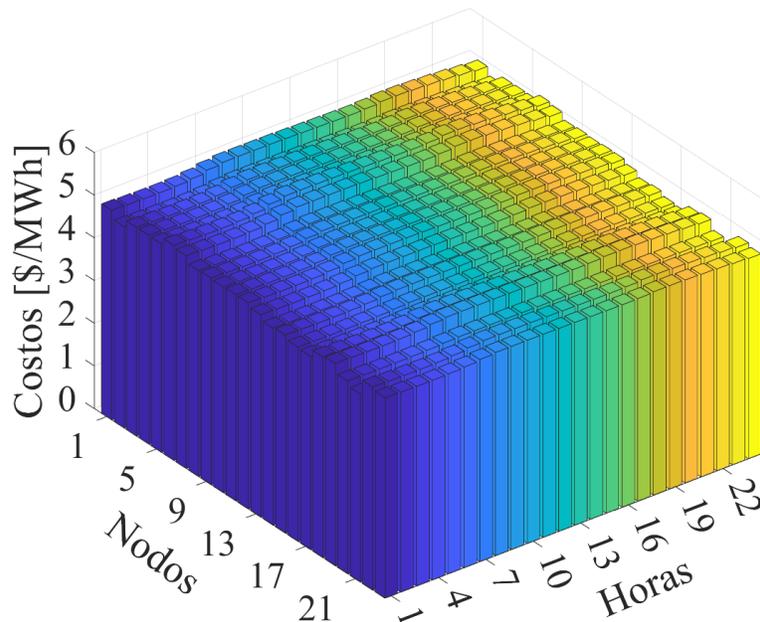


Figura 53. Costos marginales estimado a nivel nodal con los datos reales

En las figuras 54 y 55, se puede observar un corte enfocado en la relación de Costos-nodos de las dos figuras anteriores. Considerando que ambas estimaciones tienen comportamientos similares, la diferencia entre los datos va a ser poca, siendo la diferencia máxima de 0.005823544 [\$/MWh] durante la hora 7:00 en el nodo diez, lo que representa un error de 0.112%.

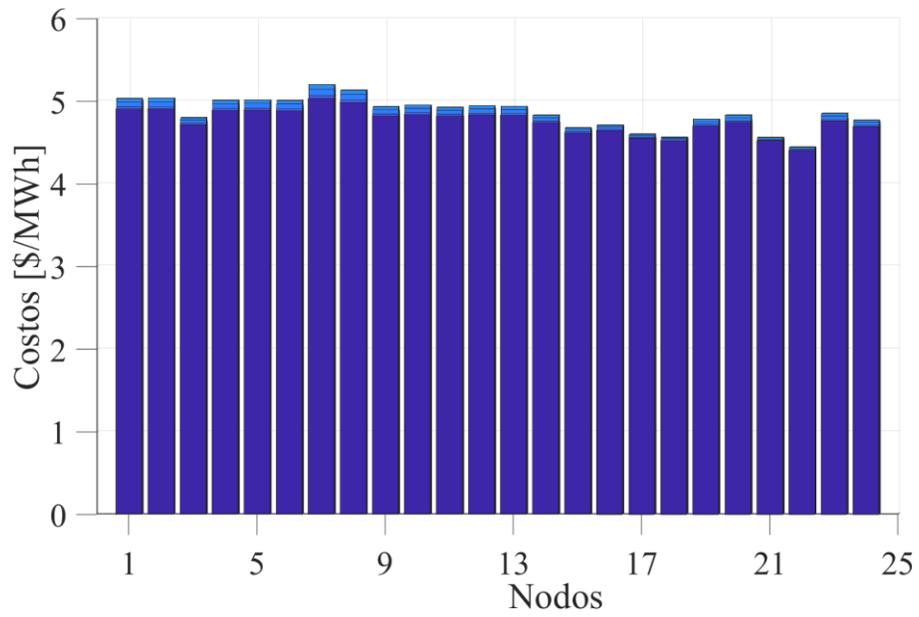


Figura 54. Relación costos-nodos (RPG) a nivel nodal

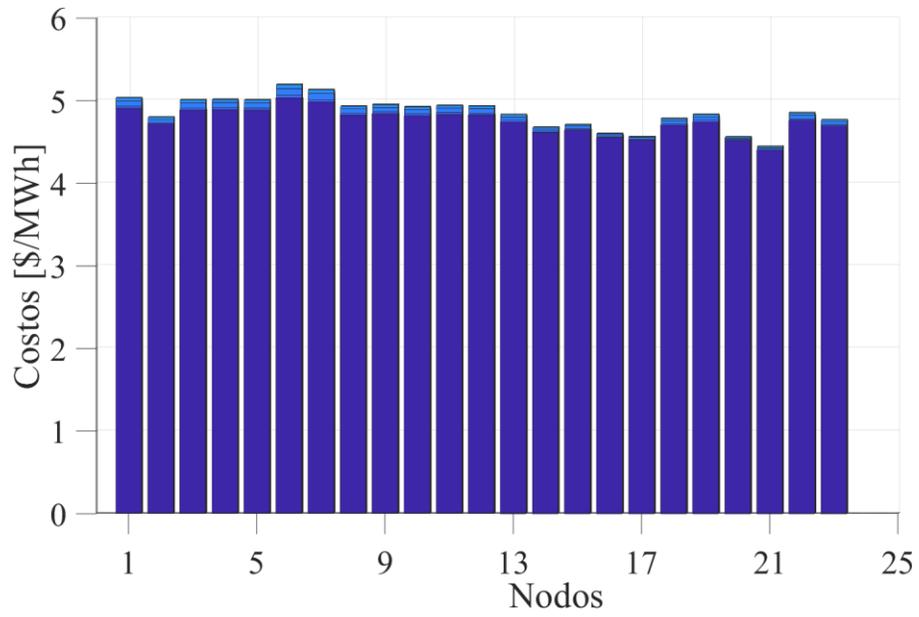


Figura 55. Relación costos-nodos (datos reales) a nivel nodal

CAPÍTULO 6. CONCLUSIONES Y TRABAJOS FUTUROS

6.1. Conclusiones

Tras el análisis previo hecho sobre el uso del Aprendizaje Automático (Machine Learning por su nombre en inglés), se demostró que el mejor modelo predictivo para la estimación de datos es el modelo de Regresión de Proceso Gaussiano (RPG) [27], por lo cual, se recurrió a hacer uso de este modelo para realizar la estimación de acuerdo a lo presentado en el capítulo 2.

Mediante la comparación de la estimación hecha con el modelo propuesto, la estimación hecha con el modelo de Media Móvil Integrada Auto-Regresiva (ARIMA) y los datos de referencia del quinto día, se pudo inferir que las estimaciones hechas con el modelo RPG son más precisas que las estimaciones hechas con el modelo ARIMA, debido a que los patrones de demanda obtenidos con RPG son más fieles a los datos de referencia, además de que sus indicadores de rendimiento obtenidos muestran menores porcentajes de error en comparación con los indicadores obtenidos con ARIMA, siendo ambos mucho menores al 3%.

Con base en lo antes mencionado, se sobreentiende que al usar los datos de la estimación de RPG para obtener los valores de los parámetros técnico-económicos del sistema, los datos de los parámetros estimados tendrán diferencias mínimas con los datos de referencia, lo que se comprueba con los valores de error obtenidos en cada uno de los casos de estudio.

Aislando los tres casos de estudio en parámetros técnico-económicos y analizando el valor de sus errores, encontramos que para el parámetro ‘Generación de potencia’ se presenta un error menor al 5%, en el parámetro ‘Costo operativo’ se presenta un error menor al 2% y en el parámetro ‘Costo marginal de potencia activa’ se presenta un error menor al 1%.

Como consecuencia de lo expuesto, se puede concluir que el modelo RPG es el mejor modelo predictivo del conjunto de modelos de Machine Learning que podemos ocupar para la estimación de la demanda de sistemas eléctricos de potencia, debido a que presenta errores por debajo del 3% lo que garantiza una estimación óptima, y en consecuencia un mejor ajuste económico en la operación del sistema.

6.2. Trabajos futuros

El uso del aprendizaje automático para la estimación de datos tiene una gran diversificación en el área del mercado de un día en adelante. Partiendo de la metodología propuesta, se pueden plantear algunos posibles trabajos futuros que se pueden desarrollar para la mejora de la misma, como por ejemplo, hacer más

robusta la predicción de la demanda ante contingencias, de tal manera que, si ocurre una imperfección en alguno de los elementos del sistema, se pueda prever esta alteración y existan menos pérdidas.

También, se puede hacer uso de los resultados obtenidos para implementarlos en modelos de comportamiento de generadores más exactos como lo son los métodos de asignación de unidades.

Finalmente, se considera que la metodología propuesta en este trabajo puede ser enfocada a energías renovables para la estimación del viento o las emisiones del sol, así como, sus aplicaciones a los sistemas eléctricos de potencia y los sistemas de distribución.

BIBLIOGRAFÍA

- [1] J. Lee, S. Yoo, J. Kim, D. Song, and H. Jeong, “Improvements to the customer baseline load (CBL) using standard energy consumption considering energy efficiency and demand response,” *Energy*, vol. 144, pp. 1052–1063, Feb. 2018, doi: 10.1016/j.energy.2017.12.044.
- [2] K. Li, F. Wang, Z. Mi, M. Fotuhi-Firuzabad, N. Duić, and T. Wang, “Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation,” *Applied Energy*, vol. 253, p. 113595, Nov. 2019, doi: 10.1016/j.apenergy.2019.113595.
- [3] D. Gagne, E. Settle, A. Aznar, and R. Bracho, “Demand Response Compensation Methodologies: Case Studies for Mexico”, National Renewable Energy Laboratory, Denver West Parkway Golden, Tech. Rep. NREL/TP-7A40-71431, Jun. 2018.
- [4] S. Mohajeryami, M. Doostan, and P. Schwarz, “The impact of Customer Baseline Load (CBL) calculation methods on Peak Time Rebate program offered to residential customers,” *Electric Power Systems Research*, vol. 137, pp. 59–65, Aug. 2016, doi: 10.1016/j.epsr.2016.03.050.
- [5] E. Lee, K. Lee, H. Lee, E. Kim, and W. Rhee, “Defining virtual control group to improve customer baseline load calculation of residential demand response,” *Applied Energy*, vol. 250, pp. 946–958, Sep. 2019, doi: 10.1016/j.apenergy.2019.05.019.
- [6] G. Raman and J. C. -H. Peng, "A Hybrid Customer Baseline Load Estimator for Small and Medium Enterprises," *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, 2018, pp. 59-64, doi: 10.1109/IECON.2018.8591759.
- [7] I. Andrić et al., “A Baseline Load Estimation Approach for Residential Customer based on Load Pattern Clustering,” *Energy Procedia*, vol. 142, pp. 2042–2049, 2017, doi: 10.1016/j.egypro.2017.12.408.
- [8] M. Sun, Y. Wang, F. Teng, Y. Ye, G. Strbac and C. Kang, "Clustering-Based Residential Baseline Estimation: A Probabilistic Perspective," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6014-6028, Nov. 2019, doi: 10.1109/TSG.2019.2895333.
- [9] R. Sharifi, S. H. Fathi, and V. Vahidinasab, “Customer baseline load models for residential sector in a smart-grid environment,” *Energy Reports*, vol. 2, pp. 74–81, Nov. 2016, doi: 10.1016/j.egypro.2016.04.003.

- [10] O. Motlagh, A. Berry, and L. O’Neil, “Clustering of residential electricity customers using load time series,” *Applied Energy*, vol. 237, pp. 11–24, Mar. 2019, doi: 10.1016/j.apenergy.2018.12.063.
- [11] M. Charwand, M. Gitizadeh, P. Siano, G. Chicco, and Z. Moshavash, “Clustering of electrical load patterns and time periods using uncertainty-based multi-level amplitude thresholding,” *International Journal of Electrical Power & Energy Systems*, vol. 117, p. 105624, May 2020, doi: 10.1016/j.ijepes.2019.105624.
- [12] C.V. Chandran, M. Basu and K. Sunderland, Quot, “Demand Response and Consumer Inconvenience 2019” *International Conference on Smart Energy Systems and Technologies (SEST)*, Porto, Portugal, 2019, pp. 1-6, doi: 10.1109/SEST.2019.8849062.
- [13] F. Wang, K. Li, C. Liu, Z. Mi, M. Shafie-Khah, and J. P. S. Catalao, “Synchronous Pattern Matching Principle-Based Residential Demand Response Baseline Estimation: Mechanism Analysis and Approach Description,” *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6972–6985, Nov. 2018, doi: 10.1109/tsg.2018.2824842.
- [14] M. Paluszek and S. Thomas, *MATLAB Machine Learning*, 1st ed. Berkeley, CA Apress, 2017.
- [15] A. M. Ariza Ramírez, "Métodos utilizados para el pronóstico de demanda de energía eléctrica en sistemas de distribución", Tesis, UTP, Pereira, CO-RIS, 2013.
- [16] N. Chowdhury and R. Billinton, "A reliability test system for educational purposes-spinning reserve studies in isolated and interconnected systems," in *IEEE Transactions on Power Systems*, vol. 6, no. 4, pp. 1578-1583, Nov. 1991, doi: 10.1109/59.117005.
- [17] H. Om and S. Shukla, “Optimal Power Flow Analysis of IEEE-30 bus System using Soft Computing Techniques,” *International Journal of Engineering Research & Science (IJOER)*, vol. 1, no. 8, 2015.
- [18] N. Gonzalez Cabrera and G. Gutiérrez Alcaraz, "Estimation of customer base line and multi-period demand response," 2013 *IEEE International Autumn Meeting on Power Electronics and Computing (ROPEC)*, 2013, pp. 1-6, doi: 10.1109/ROPEC.2013.6702709.
- [19] A. Smola and S.V.N. Vishwanathan, *Introduction to Machine Learning*. 1st ed. United Kingdom: Cambridge University Press, 2008.
- [20] M. González, *Análisis de series temporales: Modelos ARIMA*. 1st ed. Euskadi: Facultad de Ciencias Económicas y Empresariales, 2009.

- [21] R. R. Pagano, *Estadística para las ciencias del comportamiento*. México (D.F.); Madrid: International Thomson, Imp, 2000. [22] H. Camdeviren, A. Yazici, Z. Akkus, R. Bugdayci, and M. Sungur, "Comparison of logistic regression model and classification tree: An application to postpartum depression data," *Expert Systems with Applications*, vol. 32, no. 4, pp. 987–994, May 2007, doi: 10.1016/j.eswa.2006.02.022.
- [23] E. Osuna, R. Freund and F. Girosi, "An improved training algorithm for support vector machines," *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop*, pp. 276-285, 1997, doi: 10.1109/NNSP.1997.622408.
- [24] Rasmussen C.E. (2004) *Gaussian Processes in Machine Learning*. In: Bousquet O., von Luxburg U., Rätsch G. (eds) *Advanced Lectures on Machine Learning. ML 2003. Lecture Notes in Computer Science*, vol 3176. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-28650-9_4
- [25] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37–66, Jan. 1991, doi: 10.1007/bf00153759.
- [26] S. Ledesma. "Las Redes Neuronales: implementación y consideraciones prácticas." Fifth Mexican International Conference on Artificial Intelligence, pp. 14-15, 2013.
- [27] N. González, L. Hernandez and J. Campos. "Técnicas estadísticas para el pronóstico de la demanda mediante Matlab", 2021 IEEE Reunión Internacional de Verano de Potencia, Acapulco, Guerrero, 2021.
- [28] A. Gabaldón, A. García-Garre, M. C. Ruiz-Abellón, A. Guillamón, C. Álvarez-Bel, and L. A. Fernandez-Jimenez, "Improvement of customer baselines for the evaluation of demand response through the use of physically-based load models," *Utilities Policy*, vol. 70, p. 101213, Jun. 2021, doi: 10.1016/j.jup.2021.101213.
- [29] J. Contreras, M. Asensio, P. M. de Quevedo, G. Muñoz-Delgado, and S. Montoya-Bueno, "Demand Response Modeling," *Joint RES and Distribution Network Expansion Planning Under a Demand Response Framework*, pp. 33–40, 2016, doi: 10.1016/b978-0-12-805322-5.00004-6.
- [30] Pennsylvania-New-Jersey-Meryand Interconnection, "Markets & Operations: Energy Market By Area", 2016. [Online]. Disponible en: <https://www.pjm.com/markets-and-operations/energy> [Fecha de acceso Feb. 26, 2022].
- [31] Campos-Romero J., Hernández-Núñez J., González-Cabrera N. (2022) Improving the Customer Baseline Technique Based on a Learning Machine Applied to a Power System. In: Flores Rodríguez K.L.,

Ramos Alvarado R., Barati M., Segovia Tagle V., Velázquez González R.S. (eds) Recent Trends in Sustainable Engineering. ICASAT 2021. Lecture Notes in Networks and Systems, vol 297. Springer, Cham. https://doi.org/10.1007/978-3-030-82064-0_4

[32] P. M. Subcommittee, "IEEE Reliability Test System," in IEEE Transactions on Power Apparatus and Systems, vol. PAS-98, no. 6, pp. 2047-2054, Nov. 1979, doi: 10.1109/TPAS.1979.319398.

APÉNDICE

Tabla 9. Datos de los nodos

Nodo	Tipo	Pd	Qd	Gs	Bs	Area	Vm	Va	Base kV	Zona	Vmax	Vmin
1	2	108	22	0	0	1	1	0	138	1	1.05	0.95
2	2	97	20	0	0	1	1	0	138	1	1.05	0.95
3	1	180	37	0	0	1	1	0	138	1	1.05	0.95
4	1	74	15	0	0	1	1	0	138	1	1.05	0.95
5	1	71	14	0	0	1	1	0	138	1	1.05	0.95
6	1	136	28	0	-100	2	1	0	138	1	1.05	0.95
7	2	125	25	0	0	2	1	0	138	1	1.05	0.95
8	1	171	35	0	0	2	1	0	138	1	1.05	0.95
9	1	175	36	0	0	1	1	0	138	1	1.05	0.95
10	1	195	40	0	0	2	1	0	138	1	1.05	0.95
11	1	0	0	0	0	3	1	0	230	1	1.05	0.95
12	1	0	0	0	0	3	1	0	230	1	1.05	0.95
13	3	265	54	0	0	3	1	0	230	1	1.05	0.95
14	2	194	39	0	0	3	1	0	230	1	1.05	0.95
15	2	317	64	0	0	4	1	0	230	1	1.05	0.95
16	2	100	20	0	0	4	1	0	230	1	1.05	0.95
17	1	0	0	0	0	4	1	0	230	1	1.05	0.95
18	2	333	68	0	0	4	1	0	230	1	1.05	0.95
19	1	181	37	0	0	3	1	0	230	1	1.05	0.95
20	1	128	26	0	0	3	1	0	230	1	1.05	0.95
21	2	0	0	0	0	4	1	0	230	1	1.05	0.95
22	2	0	0	0	0	4	1	0	230	1	1.05	0.95
23	2	0	0	0	0	3	1	0	230	1	1.05	0.95
24	1	0	0	0	0	4	1	0	230	1	1.05	0.95

Tabla 10. Datos de los generadores

Nodo	Pg	Qg	Qmax	Qmin	Vg	mBase	Status	Pmax	Pmin
1	10	0	10	0	1.035	100	1	20	0
1	10	0	10	0	1.035	100	1	20	0
1	76	0	30	-25	1.035	100	1	76	0
1	76	0	30	-25	1.035	100	1	76	0
2	10	0	10	0	1.035	100	1	20	0
2	10	0	10	0	1.035	100	1	20	0
2	76	0	30	-25	1.035	100	1	76	0
2	76	0	30	-25	1.035	100	1	76	0
7	80	0	60	0	1.025	100	1	100	0
7	80	0	60	0	1.025	100	1	100	0
7	80	0	60	0	1.025	100	1	100	0
13	95.1	0	80	0	1.02	100	1	197	0
13	95.1	0	80	0	1.02	100	1	197	0
13	95.1	0	80	0	1.02	100	1	197	0
14	0	35.3	200	-50	0.98	100	1	0	0
15	12	0	6	0	1.014	100	1	12	0
15	12	0	6	0	1.014	100	1	12	0
15	12	0	6	0	1.014	100	1	12	0
15	12	0	6	0	1.014	100	1	12	0
15	12	0	6	0	1.014	100	1	12	0
15	155	0	80	-50	1.014	100	1	155	0
16	155	0	80	-50	1.017	100	1	155	0
18	400	0	200	-50	1.05	100	1	400	0
21	400	0	200	-50	1.05	100	1	400	0
22	50	0	16	-10	1.05	100	1	50	0
22	50	0	16	-10	1.05	100	1	50	0
22	50	0	16	-10	1.05	100	1	50	0
22	50	0	16	-10	1.05	100	1	50	0
22	50	0	16	-10	1.05	100	1	50	0
22	50	0	16	-10	1.05	100	1	50	0
23	155	0	80	-50	1.05	100	1	155	0
23	155	0	80	-50	1.05	100	1	155	0
23	350	0	150	-25	1.05	100	1	350	0

Tabla 11. Datos de las líneas

Del bus	Al bus	r	x	b	rate A	rate B	rate C	ratio	angle	status	ang min	ang max
1	2	0.0026	0.0139	0.4611	175	250	200	0	0	1	-360	360
1	3	0.0546	0.2112	0.0572	175	208	220	0	0	1	-360	360
1	5	0.0218	0.0845	0.0229	175	208	220	0	0	1	-360	360
2	4	0.0328	0.1267	0.0343	175	208	220	0	0	1	-360	360
2	6	0.0497	0.192	0.052	175	208	220	0	0	1	-360	360
3	9	0.0308	0.119	0.0322	175	208	220	0	0	1	-360	360
3	24	0.0023	0.0839	0	400	510	600	1.03	0	1	-360	360
4	9	0.0268	0.1037	0.0281	175	208	220	0	0	1	-360	360
5	10	0.0228	0.0883	0.0239	175	208	220	0	0	1	-360	360
6	10	0.0139	0.0605	2.459	175	193	200	0	0	1	-360	360
7	8	0.0159	0.0614	0.0166	175	208	220	0	0	1	-360	360
8	9	0.0427	0.1651	0.0447	175	208	220	0	0	1	-360	360
8	10	0.0427	0.1651	0.0447	175	208	220	0	0	1	-360	360
9	11	0.0023	0.0839	0	400	510	600	1.03	0	1	-360	360
9	12	0.0023	0.0839	0	400	510	600	1.03	0	1	-360	360
10	11	0.0023	0.0839	0	400	510	600	1.02	0	1	-360	360
10	12	0.0023	0.0839	0	400	510	600	1.02	0	1	-360	360
11	13	0.0061	0.0476	0.0999	500	600	625	0	0	1	-360	360
11	14	0.0054	0.0418	0.0879	500	625	625	0	0	1	-360	360
12	13	0.0061	0.0476	0.0999	500	625	625	0	0	1	-360	360
12	23	0.0124	0.0966	0.203	500	625	625	0	0	1	-360	360
13	23	0.0111	0.0865	0.1818	500	625	625	0	0	1	-360	360
14	16	0.005	0.0389	0.0818	500	625	625	0	0	1	-360	360
15	16	0.0022	0.0173	0.0364	500	600	625	0	0	1	-360	360
15	21	0.0063	0.049	0.103	500	600	625	0	0	1	-360	360
15	21	0.0063	0.049	0.103	500	600	625	0	0	1	-360	360
15	24	0.0067	0.0519	0.1091	500	600	625	0	0	1	-360	360
16	17	0.0033	0.0259	0.0545	500	600	625	0	0	1	-360	360
16	19	0.003	0.0231	0.0485	500	600	625	0	0	1	-360	360
17	18	0.0018	0.0144	0.0303	500	600	625	0	0	1	-360	360
17	22	0.0135	0.1053	0.2212	500	600	625	0	0	1	-360	360
18	21	0.0033	0.0259	0.0545	500	600	625	0	0	1	-360	360
18	21	0.0033	0.0259	0.0545	500	600	625	0	0	1	-360	360
19	20	0.0051	0.0396	0.0833	500	600	625	0	0	1	-360	360
19	20	0.0051	0.0396	0.0833	500	600	625	0	0	1	-360	360
20	23	0.0028	0.0216	0.0455	500	600	625	0	0	1	-360	360
20	23	0.0028	0.0216	0.0455	500	600	625	0	0	1	-360	360
21	22	0.0087	0.0678	0.1424	500	600	625	0	0	1	-360	360

Tabla 12. Datos de demanda total del sistema

Horas	Lunes	Martes	Miércoles	Jueves	Viernes
1	255120.224	311426.446	312856.429	283858.089	262599.193
2	250742.957	308588.838	308542.226	277780.405	255834.761
3	250553.988	308137.483	307892.126	275304.896	252119.863
4	253126.162	310723.62	310053.921	275567.803	252098.672
5	261684.185	319717.768	318072.051	281163.572	257122.226
6	280266.095	338377.569	336074.682	298156.193	272429.924
7	309505.774	367907.267	364961.578	326274.92	299800.124
8	327848.406	384626.688	380298.366	343712.809	315884.009
9	328982.435	380281.97	372837.466	339861.738	316398.019
10	327113.795	368356.289	357238.852	332032.355	315161.101
11	325311.998	359105.165	342424.038	324132.986	314282.991
12	323003.464	349439.132	330307.12	315617.844	311230.765
13	321356.745	339283.41	319709.654	308642.039	307558.254
14	320453.38	332290.372	312495.597	302844.107	304925.515
15	320139.146	326184.89	307242.964	298766.258	301979.599
16	323629.021	325367.352	306721.3	297020.62	300241.116
17	335734.569	335105.023	315573.317	302934.753	305203.055
18	360549.773	361005.796	339795.942	322876.846	317259.67
19	370273.112	371904.053	349326.712	329047.814	316340.035
20	369140.33	371690.817	347976.9	326483.554	309462.554
21	365656.738	367838.289	343379.552	321480.574	301609.993
22	355782.047	356936.809	331478.555	310527.632	290749.722
23	338973.44	339445.337	313860.354	293073.996	275754.662
24	322829.07	323330.727	296271.723	275055.663	258532.399

Tabla 13. Datos de demanda en la región 1

Horas	Lunes	Martes	Miércoles	Jueves	Viernes
1	205050.497	252665.153	252517.864	229544.567	211154.914
2	201585.548	250122.352	249033.18	224634.35	205564.59
3	201479.741	249673.762	248549.208	222609.32	202553.959
4	203606.424	251789.647	250407.42	222813.691	202564.312
5	210698.662	259088.901	257031.306	227417.245	206789.733
6	226152.049	274463.014	272079.711	241532.601	219586.607
7	250297.949	298622.125	296010.385	264623.454	242265.193
8	265102.317	311880.071	308211.387	278570.915	255149.779
9	265897.435	307916.92	301694.11	275086.589	255310.732
10	264239.642	298140.535	288922.986	268710.603	254201.551
11	262388.306	290424.458	276859.39	262343.127	253353.172
12	260323.656	282516.888	267086.636	255548.72	250812.395
13	258854.32	274315.701	258593.894	249980.419	247863.47
14	258094.987	268748.094	252798.831	245448.239	245896.326
15	258092.575	264008.199	248753.111	242248.533	243662.809
16	261598.465	263691.591	248617.238	241134.017	242512.249
17	272289.17	272184.277	256334.095	246451.408	247038.149
18	293161.203	293596.899	276504.846	262788.153	257202.625
19	300836.188	301864.603	283805.504	266857.499	256131.458
20	299975.363	301573.751	282612.668	264273.642	250471.333
21	297168.173	298200.672	278772.616	259900.518	244049.035
22	288982.882	289070.612	268927.362	250728.675	235126.765
23	275064.606	274424.1	254275.627	236177.939	222785.467
24	261871.806	261076.726	239690.246	221401.385	208720.782

Tabla 14. Datos de demanda en la región 2

Horas	Lunes	Martes	Miércoles	Jueves	Viernes
1	50069.727	58761.293	60338.565	54313.522	51444.279
2	49157.409	58466.486	59509.046	53146.055	50270.171
3	49074.247	58463.721	59342.918	52695.576	49565.904
4	49519.738	58933.973	59646.501	52754.112	49534.36
5	50985.523	60628.867	61040.745	53746.327	50332.493
6	54114.046	63914.555	63994.971	56623.592	52843.317
7	59207.825	69285.142	68951.193	61651.466	57534.931
8	62746.089	72746.617	72086.979	65141.894	60734.23
9	63085	72365.05	71143.356	64775.149	61087.287
10	62874.153	70215.754	68315.866	63321.752	60959.55
11	62923.692	68680.707	65564.648	61789.859	60929.819
12	62679.808	66922.244	63220.484	60069.124	60418.37
13	62502.425	64967.709	61115.76	58661.62	59694.784
14	62358.393	63542.278	59696.766	57395.868	59029.189
15	62046.571	62176.691	58489.853	56517.725	58316.79
16	62030.556	61675.761	58104.062	55886.603	57728.867
17	63445.399	62920.746	59239.222	56483.345	58164.906
18	67388.57	67408.897	63291.096	60088.693	60057.045
19	69436.924	70039.45	65521.208	62190.315	60208.577
20	69164.967	70117.066	65364.232	62209.912	58991.221
21	68488.565	69637.617	64606.936	61580.056	57560.958
22	66799.165	67866.197	62551.193	59798.957	55622.957
23	63908.834	65021.237	59584.727	56896.057	52969.195
24	60957.264	62254.001	56581.477	53654.278	49811.617

Tabla 15. Datos de demanda en el nodo 1.

Horas	Lunes	Martes	Miércoles	Jueves	Viernes
1	80971.986	98215.629	98941.688	89708.522	83060.94
2	79572.803	97340.961	97606.781	87827.156	80941.945
3	79473.887	97230.152	97404.69	87075.235	79817.825
4	80241.121	98068.684	98079.742	87194.926	79786.956
5	82937.006	100909.13	100582.57	88953.416	81305.833
6	88745.616	106786.73	106142.93	94286.478	86120.924
7	97914.384	116073.96	115184.16	103133.78	94699.498
8	103643.47	121343.54	119995.42	108616.53	99744.15
9	103952.92	120026.27	117698.38	107413.19	99897.336
10	103385.75	116235.02	112788.75	104850.21	99513.023
11	102879.11	113281.7	108057.9	102220.24	99225.966
12	102144	110248.07	104141.81	99386.752	98243.434
13	101591.68	106937.4	100751.79	97079.956	97064.389
14	101277.11	104724.29	98417.616	95206.953	96182.619
15	101129.45	102715.03	96710.198	93861.358	95194.996
16	102106.71	102399.04	96469.524	93239.323	94607.994
17	105720.31	105381.4	99179.116	95002.715	96065.842
18	113401.52	113518.64	106745.34	101299.08	99775.588
19	116591.01	117200.61	109958.67	103610.6	99644.971
20	116222.74	117222.6	109605.8	102965.93	97518.769
21	115153.47	116068.28	108218.8	101479.01	95058.2
22	112091.83	112704.95	104540.67	98096.747	91718.094
23	106860.51	107266.96	99056.927	92639.853	87031.327
24	101777.13	102241.35	93570.743	86977.143	81649.148

Tabla 16. Datos de demanda en el nodo 2.

Horas	Lunes	Martes	Miércoles	Jueves	Viernes
1	69153.102	82953.516	82958.056	77201.496	71743.686
2	67961.045	82009.447	81573.041	75564.644	69880.599
3	67800.435	81632.215	81287.865	74875.252	68848.511
4	68277.372	82105.732	81684.336	74858.263	68778.614
5	70334.118	84215.889	83642.314	76312.19	69965.483
6	75039.612	88854.542	88277.179	80714.462	73897.135
7	82750.009	96590.255	95871.095	88154.222	81079.341
8	87978.027	101262.74	100284.26	93054.479	85538.651
9	88579.205	100511.72	98845.311	92229.252	85976.855
10	88294.654	97619.156	95460.592	90244.701	85760.834
11	88000.526	95506.744	92119.731	88091.514	85566.006
12	87486.675	93179.481	89171.866	85742.708	84750.025
13	87081.157	90692.211	86578.749	83894.623	83799.723
14	86831.427	88996.613	84754.534	82357.141	83097.385
15	86732.298	87367.869	83429.439	81257.075	82310.191
16	87414.702	87047.947	83178.942	80625.234	81680.712
17	90406.096	89374.853	85310.546	82061.969	82795.92
18	96636.535	96039.669	91577.815	87305.578	85917.137
19	99348.098	99101.776	94304.615	89200.417	85790.752
20	99027.228	98975.022	94025.224	88620.639	83945.546
21	98028.541	97955.805	92883.974	87350.824	81886.636
22	95360.497	95086.461	89799.481	84528.299	79075.698
23	90853.418	90438.796	85215.415	79913.183	74997.203
24	86370.898	85969.028	80554.848	75137.098	70299.271

Tabla 17. Datos de demanda en el nodo 7.

Horas	Lunes	Martes	Miércoles	Jueves	Viernes
1	3829.709	5479.58	5165.528	4846.423	4032.505
2	3905.004	5352.452	5184.605	4809.357	3970.356
3	4022.741	5281.078	5179.616	4809.25	3903.385
4	4127.553	5337.729	5217.962	4802.519	3919.047
5	4290.334	5471.821	5319.413	4951.258	4045.896
6	4613.294	5765.125	5631.811	5337.183	4296.498
7	5075.649	6258.456	6204.438	5850.017	4809.367
8	5365.37	6471.921	6398.773	6061.079	4924.92
9	5508.402	6322.116	6153.903	5911.085	4915.597
10	5477.451	6109.01	5929.403	5741.222	4872.515
11	5477.948	5951.308	5753.302	5567.511	4908.614
12	5502.571	5773.215	5596.657	5410.693	4873.451
13	5513.643	5591.888	5414.035	5294.074	4801.976
14	5471.494	5482.847	5272.924	5174.024	4778.433
15	5428.034	5384.565	5171.79	5092.781	4723.556
16	5507.731	5350.895	5119.693	5051.863	4618.469
17	5754.428	5521.381	5258.741	5155.15	4701.955
18	6131.281	5900.293	5588.213	5405.368	4891.704
19	6279.384	6001.765	5651.577	5088.943	4827.745
20	6326.362	5997.62	5627.425	4896.859	4742.244
21	6288.141	5859.533	5589.719	4746.21	4646.32
22	6125.913	5734.063	5451.079	4610.564	4481.842
23	5869.707	5467.152	5221.429	4358.396	4245.33
24	5680.468	5266.474	4984.968	4145.622	3933.283