



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Estudios de Postgrado
Maestría en Estadística Aplicada

**MODELO DE PRONÓSTICO PARA ESTIMAR LA DEMANDA DE ENERGÍA ELÉCTRICA EN
GUATEMALA**

Ing. Luis Humberto Duarte Rodríguez

Asesorado por el Mtro. Walter Arnoldo Bardales Espinoza

Guatemala, marzo de 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**MODELO DE PRONÓSTICO PARA ESTIMAR LA DEMANDA DE ENERGÍA ELÉCTRICA EN
GUATEMALA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

ING. LUIS HUMBERO DUARTE RODRÍGUEZ

ASESORADO POR EL MTRO. WALTER ARNOLDO BARDALES ESPINOZA

AL CONFERÍRSELE EL TÍTULO DE

MAESTRO EN ESTADÍSTICA APLICADA

GUATEMALA, MARZO DE 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martinez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANA	Inga. Aurelia Anabela Cordova Estrada
DIRECTOR	Ing. Edgar Darío Álvarez Cotí
EXAMINADOR	Ing. Edwin Adalberto Bracamonte Orozco
EXAMINADOR	Ing. William Eduardo Fagiani Cruz
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

MODELO DE PRONÓSTICO PARA ESTIMAR LA DEMANDA DE ENERGÍA ELÉCTRICA EN GUATEMALA

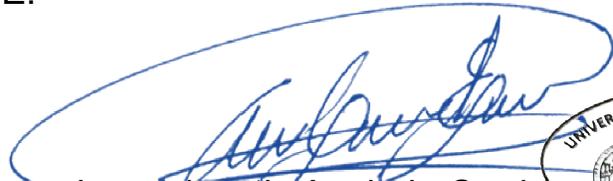
Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 3 de agosto de 2021.

Ing. Luis Humberto Duarte Rodríguez

LNG.DECANATO.OI.150.2022

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Posgrado, al Trabajo de Graduación titulado: **MODELO DE PRONÓSTICO PARA ESTIMAR LA DEMANDA DE ENERGÍA ELÉCTRICA EN GUATEMALA**, presentado por: **Luis Humberto Duarte Rodríguez**, que pertenece al programa de Maestría en artes en Estadística aplicada después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:



Inga. Aurelia Anabela Cordova

Decana

Guatemala, marzo de 2022

AACE/gaoc



Guatemala, marzo de 2022

LNG.EEP.OI.150.2022

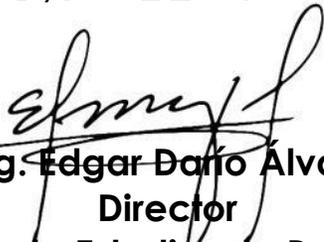
En mi calidad de Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor, verificar la aprobación del Coordinador de Maestría y la aprobación del Área de Lingüística al trabajo de graduación titulado:

“MODELO DE PRONÓSTICO PARA ESTIMAR LA DEMANDA DE ENERGÍA ELÉCTRICA EN GUATEMALA”

presentado por **Luis Humberto Duarte Rodríguez** correspondiente al programa de **Maestría en artes en Estadística aplicada**; apruebo y autorizo el mismo.

Atentamente,

“Id y Enseñad a Todos”


Mtro. Ing. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería





Guatemala 30 de octubre 2021.

M.A. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Presente

M.A. Ingeniero Álvarez Cotí:

Por este medio informo que he revisado y aprobado el Informe Final del trabajo de graduación titulado “**MODELO DE PRONÓSTICO PARA ESTIMAR LA DEMANDA DE ENERGÍA ELÉCTRICA EN GUATEMALA**” de la estudiante **Luis Humberto Duarte Rodríguez** quien se identifica con número de carné 202090558 del programa de Maestría en Estadística Aplicada.

Con base en la evaluación realizada hago constar que he evaluado la calidad, validez, pertinencia y coherencia de los resultados obtenidos en el trabajo presentado y según lo establecido en el *Normativo de Tesis y Trabajos de Graduación aprobado por Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014*. Por lo cual el trabajo evaluado cuenta con mi aprobación.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.

Atentamente,

MSc. Ing. Edwin Adalberto Bracamonte Orozco
Coordinador
Maestría en Estadística Aplicada
Escuela de Estudios de Postgrado



Guatemala, 30 de octubre 2021.

Maestro
Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería
Presente.

Estimado Señor Director:

De manera atenta hago constar que he revisado el Informe Final y Artículo Científico de la estudiante **Luis Humberto Duarte Rodríguez** con número de carné **202090558** de la Maestría en **Estadística Aplicada** titulado: **“MODELO DE PRONÓSTICO PARA ESTIMAR LA DEMANDA DE ENERGÍA ELÉCTRICA EN GUATEMALA”**.

Con base en la evaluación realizada hago constar la originalidad, calidad, coherencia según lo establecido en el Normativo de Tesis y Trabajos de Graduación aprobados por la Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014. Cumpliendo tanto en su estructura como en su contenido, por lo cual el trabajo y artículo evaluado cuenta con mi aprobación.

Atentamente,
“Id y Enseñad a Todos”

MSc. Ing. Edwin Adalberto Bracamonte Orozco
Coordinador
Maestría en Estadística Aplicada
Escuela de Estudios de Postgrado

Guatemala, 3 agosto de 2021.

M.A. Ing. Edgar Darío Álvarez Cotí

Director

Escuela de Estudios de Postgrado

Presente

Estimado M.A. Ing. Álvarez Cotí

Por este medio informo a usted, que he revisado y aprobado el Trabajo de Graduación y el Artículo Científico: **“MODELO DE PRONÓSTICO PARA ESTIMAR LA DEMANDA DE ENERGÍA ELÉCTRICA EN GUATEMALA”** del estudiante **Luis Humberto Duarte Rodríguez** del programa de Maestría en **Estadística Aplicada**, identificado con número de carné: **202090558**.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.



M.Sc. Ing. Walter Arnoldo Bardales Espinoza
Maestro en Recursos Hidráulicos Opción Hidrología
Ingeniero Agrónomo
Colegiado 4279

MSc. Ing. Walter Arnoldo Bardales Espinoza

Colegiado No. 4279

Asesor de Tesis

ACTO QUE DEDICO A:

Dios

Por todo lo que me ha dado.

Mi familia

Por ser todo lo que tengo.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por ser la institución que me permitió adquirir conocimientos.
Mis profesores	Por proporcionarme los conocimientos que me han permitido realizar este trabajo de graduación.
Mi asesor	Msc. Ing. Walter Bardales, por haberme guiado durante el trabajo de graduación.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	V
LISTA DE SÍMBOLOS	IX
GLOSARIO	XI
RESUMEN.....	XIII
PLANTEAMIENTO DEL PROBLEMA.....	XV
OBJETIVOS.....	XIX
RESUMEN DEL MARCO METODOLÓGICO	XXI
INTRODUCCIÓN	XXIX
1. MARCO REFERENCIAL.....	1
2. MARCO TEÓRICO.....	9
2.1. Series de tiempo.....	9
2.1.1. Componentes de la serie de tiempo	9
2.1.1.1. Tendencia secular	9
2.1.1.2. Variación estacional.....	11
2.1.1.3. Variación cíclica.....	12
2.1.1.4. Variación irregular.....	13
2.1.2. Clasificación descriptiva de las series de tiempo....	14
2.1.2.1. Estacionarias	14
2.1.2.2. No estacionarias	14
2.1.3. Métodos de pronósticos.....	14
2.1.3.1. Promedios móviles	14
2.1.3.2. Suavizamiento exponencial	15
2.1.3.3. Holt-Winters.....	16

2.1.3.4.	Autorregresivos	16
2.1.3.5.	Media móvil	17
2.1.3.6.	ARMA/ARIMA.....	17
2.1.3.7.	Regresión dinámica.....	18
2.1.4.	Correlación	19
2.1.4.1.	Coeficiente de Spearman	19
2.1.5.	Prueba de Kruskal-Wallis	20
2.1.6.	Análisis exploratorio	20
	A continuación, lo concerniente al análisis exploratorio.	20
2.1.6.1.	Media.....	20
2.1.6.2.	Mediana.....	21
2.1.6.3.	Moda	21
2.1.6.4.	Desviación estándar	21
2.1.6.5.	Máximo.....	21
2.1.6.6.	Mínimo.....	21
2.1.6.7.	Diagrama de cajas	22
2.1.7.	Autocorrelación	23
2.1.8.	Autocorrelación parcial	23
2.1.9.	Prueba de Dickey-Fuller	23
2.1.10.	Análisis residual	24
2.1.11.	Prueba de Kolmogorov-Smirnov	24
2.1.12.	Prueba de Ljung-Box.....	25
2.1.13.	Criterio de Información de Akaike	26
2.1.14.	Criterio de Información de Bayes	26
2.1.15.	Error absoluto porcentual medio	27
2.1.16.	Raíz del error cuadrado medio	27
2.2.	Redes neuronales	28
2.2.1.	Elementos básicos	28
2.2.1.1.	Capas.....	28

	2.2.1.2.	Función de entrada.....	29
	2.2.1.3.	Función de activación	29
	2.2.1.4.	Función de salida.....	30
2.3.		Demanda eléctrica.....	30
	2.3.1.	Factores que afectan la demanda eléctrica	30
		2.3.1.1. Clima	30
		2.3.1.2. Calendario	31
		2.3.1.3. Información económica.....	31
		2.3.1.4. Factores no predecibles	31
3.		PRESENTACIÓN DE RESULTADOS.....	33
	3.1.	Objetivo 1. Identificar las variables que influyen en el comportamiento de la demanda de energía eléctrica en Guatemala, por medio de un análisis de correlación multinomial	33
	3.2.	Objetivo 2. Construir un modelo de predicción con el mejor ajuste para la estimación de la demanda de energía eléctrica en Guatemala, por medio de la comparación de distintos modelos de series de tiempo y uno de redes neuronales	40
		3.2.1. Análisis de la serie de tiempo	41
		3.2.2. Modelo ARIMA.....	48
		3.2.3. Modelo regresión dinámica.....	57
		3.2.4. Redes neuronales.....	67
	3.3.	Objetivo 3. Evaluar la validez y confiabilidad del modelo	70
		3.3.1. ARIMA	70
		3.3.2. Regresión dinámica	72
		3.3.3. Redes neuronales.....	73
	3.4.	General. Determinar el modelo más confiable de pronóstico para estimar la demanda de energía eléctrica en Guatemala,	

entre distintos modelos de series de tiempo y uno de redes neuronales	74
4. DISCUSIÓN DE RESULTADOS	75
CONCLUSIONES.....	83
RECOMENDACIONES	85
REFERENCIAS	87

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Ejemplo de tendencia secular en serie de tiempo.....	10
2.	Ejemplo de variación estacional en serie de tiempo	11
3.	Ejemplo de variación cíclica en serie de tiempo.....	12
4.	Ejemplo de variación irregular en serie de tiempo	13
5.	Ejemplo de diagrama de cajas	22
6.	Comparación de demanda eléctrica con variables estudiadas	34
7.	Diagramas de dispersión para variables estudiadas.....	35
8.	Comparación de demanda eléctrica con variables estudiadas diferenciadas.....	37
9.	Diagramas de dispersión de variables estudiadas diferenciadas.....	38
10.	<i>Boxplot</i> de la demanda de energía eléctrica por mes	39
11.	Prueba de Kruskal-Wallis para la demanda de energía eléctrica por mes.....	40
12.	Demanda de energía eléctrica en Guatemala 1989-2019.....	41
13.	Demanda de energía eléctrica en Guatemala 1989-2019.....	42
14.	Descomposición de la serie de tiempo de la demanda de energía eléctrica en Guatemala 1989-2019	43
15.	Prueba de Box-Cox.....	44
16.	Demanda de energía eléctrica en Guatemala 1989-2019 con transformación logarítmica	45
17.	Descomposición de la serie de demanda de energía eléctrica en Guatemala 1989-2019 con transformación logarítmica.....	46

18.	Análisis de estacionalidad de la demanda de energía eléctrica en Guatemala 1989-2019	47
19.	Gráficos de ACF y PACF de la demanda de energía eléctrica en Guatemala 1989-2013	48
20.	Prueba de Dickey Fuller.....	49
21.	Demanda de energía eléctrica en Guatemala 1989-2013 diferenciada	50
22.	Prueba de Dickey Fuller.....	50
23.	Gráficos de ACF y PACF de la demanda de energía eléctrica en Guatemala 1989-2013 doblemente diferenciada	51
24.	Criterios de información BIC para modelos tentativos	52
25.	Criterios de información AIC para modelos tentativos	53
26.	MAPE para modelos tentativos.....	54
27.	Análisis residual del modelo ARIMA	56
28.	Prueba de Ljung-Box	57
29.	Prueba de Kolmogorov-Smirnov	57
30.	Prueba de Dickey Fuller (temperatura).....	58
31.	Prueba de Dickey Fuller (humedad)	58
32.	Coeficientes de regresión	58
33.	Residuos de regresión , ACF y PACF.....	59
34.	Prueba de Dickey Fuller.....	59
35.	Residuos de regresión diferenciados, ACF y PACF	60
36.	Prueba de Dickey Fuller.....	61
37.	Criterios de información BIC para modelos tentativos	62
38.	Criterios de información AIC para modelos tentativos	63
39.	MAPE para modelos tentativos.....	64
40.	Análisis residual del modelo de regresión dinámica	66
41.	Prueba de Ljung-Box	67
42.	Prueba de Kolmogorov-Smirnov	67

43.	Análisis residual del modelo de redes neuronales	69
44.	Prueba de Ljung-Box.....	70
45.	Prueba de Kolmogorov-Smirnov	70
46.	Pronóstico de la demanda de energía eléctrica mensual con ARIMA(2,1,3)(2,1,3) ₁₂	71
47.	Pronóstico de la demanda de energía eléctrica mensual con regresión dinámica (2,1,3)(1,1,2) ₁₂	72
48.	Pronóstico de demandad de energía eléctrica mensual con NNAR(15,4,10).....	73

TABLAS

I.	Variables del estudio	XXII
II.	Pruebas de correlación de las variables estudiadas contra la demanda de energía eléctrica	36
III.	Pruebas de correlación de las variables estudiadas diferenciadas contra la demanda de energía eléctrica	38
IV.	Resumen de los datos.....	41
V.	Resumen de los mejores modelos	55
VI.	Resumen de los mejores modelos	65
VII.	Modelos de redes neuronales tentativos.....	68
VIII.	Evaluación de modelo ARIMA(2,1,3)(2,1,3) ₁₂	71
IX.	Evaluación de modelo regresión con ARIMA(2,1,3)(2,1,1) ₁₂	72
X.	Evaluación de modelo NNAR(15,4,10) ₁₂	74
XI.	Comparación de modelos	74

LISTA DE SÍMBOLOS

Símbolo	Significado
°	Grados
°C	Grados celsius
=	Igual que
>	Mayor que
MW	Megavatio
<	Menor que
%	Porcentaje
W	Vatio

GLOSARIO

AIC	Criterio información Akaike.
AMM	Administrador del Mercado Mayorista.
ANOVA	Análisis de Varianza.
AR	Autorregresivo.
ARIMA	Autorregresivo integrado de media móvil.
BIC	Criterio información bayesiano.
BPNN	<i>Backpropagation artificial neural network.</i>
FNN	<i>Feedforward neural network.</i>
INSIVUMEH	Instituto Nacional de Sismología, Vulcanología, Meteorología e Hidrología.
MA	Media móvil.
MAE	Error absoluto medio.
MAPE	Error porcentual absoluto medio.

MM	Mercado Mayorista.
NNAR	Redes neuronales de retroalimentación con una sola capa oculta y entradas retrasadas para pronosticar series de tiempo univariadas.
PIB	Producto interno bruto.
RMSE	Raíz del error cuadrático medio.

RESUMEN

El propósito de esta investigación fue construir un modelo estadístico aceptablemente confiable y válido para la estimación de la demanda de energía eléctrica mensual en Guatemala, con el fin de realizar proyecciones acertadas a futuro que ayuden a mejorar las condiciones del mercado.

Para eso se planteó construir el modelo más confiable de pronóstico para estimar la demanda de energía eléctrica en Guatemala, utilizando una metodología con enfoque cuantitativo debido a que se analizó la demanda de energía eléctrica en Guatemala para el periodo comprendido entre 1989 y 2019 a través de la estimación de diferentes modelos de series de tiempo, siendo estos: un modelo ARIMA estacional, un modelo de regresión dinámica y uno de redes neuronales.

Obteniendo como principal resultado que el modelo más confiable para estimar la demanda de energía eléctrica en Guatemala era el modelo de regresión generado a partir de las variables humedad y temperatura con errores ARIMA (2,1,3) (1,1,2)₁₂ con un MAPE de 1.47 % y un RMSE de 15816.35MW. Por lo tanto, se recomienda evaluar el modelo obtenido contra modelo utilizado actualmente por el AMM, con el fin de comparar precisión de los resultados.

PLANTEAMIENTO DEL PROBLEMA

El mercado eléctrico en Guatemala funciona como un mercado libre desde el año 1996, este esquema permite la libre competencia en la generación y la comercialización de energía. El responsable de la operación del sistema y del mercado eléctrico es el Administrador del Mercado Mayorista (AMM), su labor es operar el Sistema Nacional Interconectado (SNI) y el Mercado Mayorista (MM) asegurando la continuidad y seguridad del suministro de energía eléctrica en el país, además de garantizar que se atienda la demanda de energía.

Debido a la importancia que tiene para el Administrador del Mercado Mayorista el cumplimiento de su misión, se hace necesario pronosticar la demanda de energía eléctrica; ya que esto permite realizar proyecciones a futuro del comportamiento del mercado, establecer planes de comercialización, optimizar la generación de energía, definir sus límites de exportación y construir planes de expansión a futuro.

Descripción del problema

Al ser la energía eléctrica un mercado que se maneja bajo oferta y demanda en Guatemala, una de las tareas de gran importancia para los agentes involucrados (generadores, distribuidores, transportistas, comercializadores y usuarios) en la generación de energía es el pronóstico de la demanda de electricidad, debido a que esta ayuda a mejorar la confiabilidad y la planificación de la red; el problema consiste en que esta demanda no presenta características lineales y se ve afectada por una gran cantidad de variables, entre las que se

encuentran las condiciones del tiempo, las estaciones del año, variables demográficas y económicas.

Es por ello por lo que se hace necesario construir un modelo estadístico confiable y válido para la estimación de demanda de energía eléctrica en Guatemala, con el fin de poder realizar proyecciones acertadas sobre la demanda energética que ayuden a su uso y a mejorar las condiciones del mercado.

Formulación del problema

Pregunta central

¿Entre la técnica de series de tiempo y redes neuronales, cuál es más confiable para predecir la demanda de energía eléctrica en Guatemala?

Preguntas auxiliares

- ¿Cuáles son las variables que influyen con mayor incidencia en el comportamiento de la demanda de energía eléctrica en Guatemala?
- ¿Cuáles son los modelos de series de tiempo y de redes neuronales que permiten estimar la demanda de energía eléctrica en Guatemala?
- ¿Cuál es el modelo más confiable y válido para el pronóstico de demanda de energía eléctrica en Guatemala?

Delimitación del problema

Para el análisis del problema se utilizará información histórica de las lecturas de carga horaria de demanda eléctrica a nivel de generación del Sistema Nacional Interconectado para Guatemala desde el año 1989 al 2019, además información sobre el PIB, índices poblaciones, temperatura, humedad, mes del año.

OBJETIVOS

General

Determinar el modelo más confiable de pronóstico para estimar la demanda de energía eléctrica en Guatemala, entre distintos modelos de series de tiempo y uno de redes neuronales.

Específicos

- Identificar las variables que influyen en el comportamiento de la demanda de energía en Guatemala, por medio de un análisis de correlación multinomial.
- Construir un modelo de predicción con el mejor ajuste para la estimación de la demanda de energía eléctrica en Guatemala, por medio de la comparación de distintos modelos de series de tiempo y uno de redes neuronales
- Evaluar la validez y confiabilidad del modelo, por medio del error absoluto porcentual medio, la raíz del error cuadrado medio y el análisis residual.

RESUMEN DEL MARCO METODOLÓGICO

Características del estudio

El enfoque del estudio fue cuantitativo, ya que se realizó una recolección de las lecturas de carga horaria de demanda eléctrica a nivel de generación del Sistema Nacional Interconectado de Guatemala, para estudiar y determinar sus patrones de comportamiento, con el fin de describir el fenómeno por medio de un modelo matemático.

El tipo de estudio y alcance fue correlacional, dado que se evaluó la relación o inferencia que tienen distintas variables (mes, temperatura, humedad, PIB, población) sobre la demanda eléctrica en Guatemala, con el fin de encontrar un modelo que permita explicarlo.

El diseño adoptado fue no experimental, pues la información de las lecturas de carga horaria de demanda eléctrica se analizó en su estado original sin ninguna manipulación; además fue transversal pues se estudió por medio de la serie de tiempo en el periodo 1989-2019 sus patrones de comportamiento y la influencia de otras variables; con el objetivo de encontrar un modelo válido y confiable que lo describiera y que proporcionará estimaciones a futuro.

Unidades de análisis

La población en estudio fue formada por los usuarios del Sistema Nacional Interconectado Guatemala, la cual se encuentra dividida en subpoblaciones dadas por su ubicación (municipio), que fueron estudiadas en su totalidad.

Variables

En la tabla I, se describen las variables que serán utilizadas en el desarrollo del estudio.

Tabla I. **Variables del estudio**

Variable	Definición teórica	Definición operativa
Mes	Es cada uno de los doce periodos de tiempo del año, contiene de 28 a 31 días. Variable cuantitativa discreta.	Número entero, adimensional. Escala de intervalo.
Demanda Energética	Es la cantidad de potencia consumida en un periodo determinado de tiempo. Variable cuantitativa continua.	Número real, en unidades Watts. Escala de razón.
Población	Es el conjunto de habitantes de un lugar.	Número entero, adimensional. Escala de razón.
Producto Interno Bruto (PIB)	Es un indicador económico que refleja el valor monetario de los bienes y servicios finales producido por un país, en un periodo de tiempo determinado. Variable cuantitativa continua.	Número real, en USD. Escala de razón.
Humedad	Es la cantidad de agua o cualquier otro liquido presente	Número real, porcentual. Escala de razón.

Continuación tabla I.

Variable	Definición teórica	Definición operativa
	en el aire. Variable cuantitativa continua.	
Temperatura (t)	Es el indicador de la cantidad de energía calorífica acumulada en el aire. Variable cuantitativa continua.	Número real en grados centígrados, Escala de intervalo.

Fuente: elaboración propia.

Fases del estudio

A continuación, se detallan las fases necesarias para el desarrollo del estudio.

Fase 1. Revisión de literatura

Esta tuvo como objetivo construir un marco teórico para contextualizar el estudio desarrollado, investigar los estudios similares para fundamentar la importancia de la investigación y analizar la metodología empleada, con el fin de seleccionar las técnicas y métodos que se adecuaban al cumplimiento de los objetivos. Entre las actividades realizadas se encuentran:

- Lectura de bibliografía referente a los temas de estudio.
- Elaboración de marco teórico y antecedentes.
- Definición del diseño de investigación, objetivos del estudio y metodología.

Fase 2. Recolección de la información

En esta fase se obtuvieron los datos para el análisis, se limpiaron y transformaron de acuerdo con las necesidades del estudio. Entre las actividades realizadas se encuentran:

- Recogida de información de las lecturas de carga horaria de demanda eléctrica a nivel de generación del Sistema Nacional Interconectado de Guatemala, recolección de datos climáticos por medio del INSIVUMEH y recolección de información de PIB y número de población por medio del Banco Mundial.
- Limpieza y transformación de los datos.

Fase 3. Análisis de información

En esta fase se realizó todo el análisis para cumplir los objetivos del estudio. Entre las actividades realizadas se encuentran:

- Análisis exploratorio de los datos, se realizaron gráficos de cajas, cálculos de la media, mediana, moda, desviación estándar, valores máximo y mínimo.
- Observación y análisis gráfico de las características de la serie de tiempo aleatoriedad, ciclos, tendencia y estacionalidad.
- Análisis de correlación para determinar las variables que influyen en la demanda de energética en Guatemala.
- Análisis de las técnicas de pronóstico para determinar el modelo de estimación para la demanda de energía eléctrica en Guatemala, tomando en cuenta las características de la serie de tiempo y las variables con mayor influencia, estudiando modelos ARIMA, (Regresión dinámica y redes neuronales).

- Evaluar la validez y confiabilidad del modelo por medio del error absoluto porcentual medio, la raíz del error cuadrado medio y un análisis residual.

Fase 4. Interpretación de información

En esta fase se realizó la discusión, conclusiones y recomendaciones de los resultados. Entre las actividades realizadas se encuentran:

- Construcción de discusión.
- Realización de conclusiones y recomendaciones.

Fase 5. Redacción de informe final

En esta fase se realizó la redacción del informe final, siguiendo el formato del documento. Entre las actividades realizadas se encuentran:

- Redacción de informe final.

Técnicas de análisis de información

Todas las técnicas de análisis de información listadas se realizaron utilizando el lenguaje de programación R.

Análisis de correlación: se utilizó un análisis de correlación, con todas las variables explicativas propuestas (humedad, temperatura, PIB, mes y poblacional) y el valor de la demanda eléctrica mensual, para medir el grado de asociación o dependencia entre las variables; con el fin de determinar las variables que influyen significativamente en el comportamiento de la demanda de energía en Guatemala, estas fueron definidas por medio de la significancia del coeficiente de Spearman; estas definieron las variables a utilizar en la

construcción de la red neuronal y las variables a tomar en cuenta en el modelo de regresión dinámica.

Análisis exploratorio: se utilizó un análisis exploratorio para comprender la naturaleza de los datos, verificar valores atípicos, extremos y valores perdidos; a través del cálculo de la media, mediana, moda, desviación estándar, máximo y mínimo; además de un diagrama de cajas. Además, se realizó un análisis gráfico de la serie de tiempo y descomposición de sus componentes para determinar sus características: aleatoriedad, ciclos, tendencia y estacionalidad.

Autocorrelación y autocorrelación parcial: se utilizó la autocorrelación y autocorrelación parcial en los datos de la demanda eléctrica, para explorar la serie de tiempo con el fin de identificar estacionalidad y tendencia; además a determinar el tipo de modelo de series de tiempo que se debía aplicar para su estimación.

Prueba de estacionariedad: se utilizó la prueba de Dickey-Fuller Aumentado para identificar si existía estacionariedad en los datos de demanda eléctrica en Guatemala, durante el periodo de 1989 a 2019; con el fin de aplicar modelos de tipo ARIMA.

Análisis de series de tiempo: la selección del modelo se definió a partir de las características identificadas en la serie de tiempo (tendencia y estacionalidad), al presentar tendencia y estacionalidad, se utilizó un modelo ARIMA estacional. Adicional se utilizó un modelo de regresión dinámica con las variables que se determinaron influían significativamente en el comportamiento de la demanda de energía eléctrica.

Redes neuronales: se utilizaron las variables que influían significativamente en el comportamiento de la demanda de energía eléctrica como entradas a la primera capa en conjunto con valores rezagados de la misma serie de tiempo; la selección del número idóneo de capas ocultas y de neuronas se realizó por medio de prueba y error; además se utilizó una función de activación de tipo sigmoide y una función lineal para la salida, utilizando el algoritmo de optimización de Broyden Fletcher Goldfarb Shaano.

Análisis residual: se realizó un análisis residual de los modelos para validar si los modelos eran adecuados. Se necesitó validar los supuestos de aleatoriedad, independencia y normalidad en los residuos, por lo cual se utilizó la prueba de Ljung-Box para determinar la aleatoriedad e independencia, acompañado de un análisis gráfico de autocorrelación y pruebas de Kolmogorov-Smirnov para evaluar la normalidad.

Validez de modelo: con los datos de prueba se evaluó la confiabilidad de los modelos para elegir la mejor opción, para esto se utilizó el error absoluto porcentual medio y la raíz del error cuadrado medio.

INTRODUCCIÓN

El presente trabajo es una sistematización para el pronóstico de la demanda de energía eléctrica en Guatemala. El problema es que el mercado eléctrico tiene características especiales, esta no puede ser almacenada y no tiene remplazante a corto plazo; por lo cual, los organismos encargados de regularlo deben garantizar el cumplimiento de la demanda por medio de una planeación exhaustiva.

Es allí donde un modelo pronóstico confiable tiene importancia ya que es una herramienta de suma utilidad que permite realizar predicciones a futuro sobre dicha demanda, dando la oportunidad de conocer previamente las necesidades de la red y de realizar una organización del sistema desde el punto de vista generacional.

Para realizar el estudio se analizó la demanda de energía eléctrica mensual en Guatemala, desde el año de 1989 hasta el 2019, con el fin de determinar el modelo más confiable de pronóstico para su estimación. Se identificaron las variables más influyentes en la demanda de energía eléctrica por medio de un análisis correlacional y las características de la serie de tiempo por medio de un análisis descriptivo y gráfico; luego de este análisis se propusieron tres modelos tentativos siendo estos: un modelo ARIMA estacional, un modelo de regresión dinámica y uno de redes neuronales; los modelos fueron evaluados por medio del análisis residual y de la exactitud de pronóstico por medio del MAPE y RMSE.

Como resultado se logró identificar que las variables humedad y temperatura influían en el comportamiento de la demanda de energía eléctrica,

además se determinó que el modelo más confiable para estimar la demanda de energía eléctrica en Guatemala era el modelo de regresión con errores ARIMA (2,1,3)(1,1,3)₁₂ con un MAPE de 1.47 % y un RMSE de 15,816.35MW

El estudio se dividió en 5 fases siendo estas: revisión de literatura, gestión o recolección de datos, análisis de la información, interpretación de información y redacción de informe final, adicional este estudio fue factible debido a que se contó con todos los recursos necesarios para su desarrollo.

El informe final del estudio se estructura en 4 capítulos que se discuten a continuación: en el primer capítulo se contextualizará el estudio por medio de antecedentes de estudios similares, fundamentando la importancia de la investigación. En el segundo capítulo se describen las técnicas estadísticas utilizadas para el análisis de series de tiempo y de redes neuronales, de igual manera se detalla el significado de la demanda eléctrica y los factores que influyen en esta. En el tercer capítulo se exponen los resultados de la investigación y en el cuarto capítulo se realiza la discusión detallada de los resultados, donde a través de conceptos estadísticos y del conocimiento teórico adquirido se concluye.

1. MARCO REFERENCIAL

La electricidad juega un papel importantísimo en nuestra actualidad; de ella dependen una gran cantidad de actividades que se extienden desde su uso industrial hasta el consumo en nuestros hogares. Es más, en el ámbito global toda economía en crecimiento esta potenciado por un sistema eléctrico de calidad con precios competitivos; dado a estas circunstancias se podría afirmar que vivimos en un mundo que funciona con electricidad (Gonzales, 2004).

Para entender el funcionamiento del mercado eléctrico se deben comprender dos características importantes, la primera es que la energía eléctrica no puede ser almacenada, independientemente de la fuente primaria que sea utilizada para su generación; y la segunda es que, en la mayoría de sus casos de uso, la electricidad juega el papel de un *input* específico, lo cual quiere decir que es un factor de producción no reemplazable a corto plazo (Gonzales, 2004).

Por lo tanto, en el mercado eléctrico un exceso de oferta, es decir, una cantidad de electricidad generada y no consumida, significa una pérdida económica; de la misma manera en el caso contrario, cuando existe un exceso de demanda, donde la cantidad generada no satisface las necesidades requeridas; y es que a diferencia del mercado tradicional en donde los demandantes tienen la posibilidad de consumir algo similar o de esperar a alguien que atienda su petición, la energía eléctrica no posee un sustituto competitivo en la actualidad (Gonzales, 2004).

Estas características especiales en el mercado eléctrico han motivado a las diferentes economías a la creación de organismos que actúen como reguladores, encargados de coordinar la oferta y demanda del mercado, con el fin de mantener y satisfacer la demanda, tal es el caso de nuestro país.

En Guatemala a partir del Decreto No. 93-96 en 1996 se emitió la Ley General de Electricidad, en la cual se establecía en el artículo 44 la creación de una entidad privada sin fines de lucro encargada de operar el sistema y el mercado eléctrico del país, siendo esta el Administrador del Mercado Mayorista (AMM). Su responsabilidad es operar el Sistema Nacional Interconectado y el Mercado Mayorista manteniendo la continuidad y seguridad en el suministro de la energía eléctrica (Administrador del Mercado Mayorista, 2020).

Dentro de sus obligaciones se encuentra la coordinación de la operación de centrales generadoras, interconexiones a nivel internacional y líneas de transporte para las operaciones del Mercado Mayorista, en un marco de libre contratación entre sus participantes, además debe garantizar la seguridad y el abastecimiento de energía eléctrica en Guatemala (Administrador del Mercado Mayorista, 2020).

Estos deberes únicamente pueden ser cumplidos por medio de una planeación exhaustiva, que dé la oportunidad de conocer el estado actual del sistema, y además permita conocer las medidas que se deben adoptar a futuro. Para realizar el planeamiento de un sistema eléctrico, una de las herramientas más útiles es la predicción de la demanda de energía eléctrica, la cual da la posibilidad de conocer de manera previa la necesidad de expansión del sistema, con el fin de mejorar el servicio. Su importancia surge ante la incertidumbre asociada a la magnitud de demanda a futuro; ya que esta ayuda a determinar si

existe la posibilidad de producir un déficit de capacidad generadora o en caso contrario, de existir un exceso de capacidad (García, 2009).

Aun así la importancia de realizar proyecciones a futuro en el mercado eléctrico va más allá, ya que una predicción correcta puede servir para la toma de decisiones sobre la cantidad y precios de conveniencia para realizar transacciones con otras empresas; de igual manera una subestimación de la demanda de energía, podría comprometer la confiabilidad y la seguridad del sistema, hasta el punto de tener que realizar un racionamiento de la demanda; en caso contrario si este fuera demasiado alto, los costos de oportunidad pueden ser muy elevados (Orellana, 2012). Además, Gil (2016), considera que un pronóstico de demanda de energía eléctrica permite al mercado establecer estrategias de comercialización, optimizar la generación de energía y realizar negociaciones.

Ahora bien, realizar el pronóstico de este tipo de demanda es una tarea complicada, debido a que las series suelen ser complejas y presentan diferentes niveles de estacionalidad, además existe una gran complejidad en la selección de variables influyentes ya que los mercados eléctricos suelen estar bajo influencia de un gran número de factores (Rueda, Velásquez y Franco, 2010). Según Gonzales (2004) se han identificado un conjunto de variables básicas que explican la evolución de la demanda de energía eléctrica a lo largo del tiempo, este conjunto de variables-causa abarca condiciones meteorológicas, de calendario y económicas. Para ser más específico en su estudio de predicción de la demanda eléctrica horaria mediante redes neuronales artificiales él utilizó las variables climáticas: temperatura y viento; y de calendario: hora, día, fin de semana y mes. Por su parte Lozano, Caiza y Zurita (2019) adicionan a estas variables la humedad, si es un día festivo, y la carga del día, hora y semana anterior para la construcción de su modelo. Además de esto los autores añaden

que en muchas ocasiones la base de datos utilizada suele ser corta e insuficiente, las cuales resultan incapaces de poder modelar el problema de investigación. Esto resalta la importancia de considerar variables externas para la construcción del modelo de predicción; así como la dificultad que existe para poder tomar en cuenta todos los factores y la necesidad de tener una base de datos extensa para poder realizar predicciones acertadas.

De igual manera, Velásquez, Dyner y Casto (2007), en su estudio a cerca de *¿Por qué es tan difícil obtener buenos pronósticos de los precios de la electricidad en mercados competitivos?*, mencionan algunos problemas teóricos, metodológicos y empíricos sobre el uso de modelos relacionados con la demanda energética; entre estos resalta que no existen juicios empíricos, metodológicos o teóricos sobre la preferencia de un modelo específico para el pronóstico, también hace énfasis en la falta de claridad sobre cuál es el procedimiento más adecuado para realizar la estimación y que no existe un criterio común para la selección de un modelo óptimo, ya que aún hay algunas discrepancias con la selección basada en medidas de error.

Esta complejidad ha llevado a la utilización de diferentes métodos para la realización de los pronósticos que van desde los más tradicionales hasta la utilización de métodos más sofisticados. Dentro de las técnicas utilizadas para pronosticar se encuentran: métodos de series de tiempo, los cuales predicen la demanda futura en función de los valores pasados; método de análisis regresión, estos se basan en la determinación de varios factores que influyen en el patrón de demanda y el cálculo de su correlación con la demanda real; y el método de redes neuronales el cual mapea las relaciones entre varios factores que afectan la demanda.

Grimaldo, Mendoza y Reyes (2017) utilizaron tres modelos para la estimación de la demanda eléctrica, el primero un modelo de regresión tradicional utilizando como variable regresora el PIB, el segundo utilizando un modelo de regresión con el PIB y población como variables regresoras y el último un modelo con distintas variables regresoras de distintos sectores de la economía; la selección del mejor modelo se realizó a través del coeficiente de Akaike; obteniendo que el mejor modelo fue el tercero obtenido por la eliminación de *backward*. A su vez, Ortiz (2015) en su estudio titulado *Pronóstico de la Demanda Eléctrica Residencial Basado en el Modelo de Regresión Adaptativa Multivariante Spline (MARS)* utilizó el método de regresión no paramétrica MARS, sus resultados reflejaron algunas limitaciones al no tomar en cuenta factores como el nivel socioeconómico, el sector, número de clientes y condiciones climáticas, además realizó énfasis en que se debe contar con suficiente información histórica para realizar un buen estudio estadístico y de predicción.

Los métodos de análisis de series de tiempo, también son muy utilizados como es el caso de Gil (2016), que en su estudio *Pronóstico de la demanda mensual de electricidad con series de tiempo*, utiliza tres modelos aditivos con diferentes tendencias: lineal, cuadrática y cúbica, utilizando el principio de parsimonia para la elección del mejor por medio del R cuadrado ajustado, el criterio de información Akaike y el criterio de información Bayesiano; además hace uso del error promedio de pronóstico, el error promedio absoluto de pronóstico y el error cuadrático medio de pronóstico para medir la precisión de los modelos; de dicho estudio concluyó que el modelo con tendencia cuadrática presentaba mejores resultados. En su caso Orellana (2012), en la tesis *Modelación y pronóstico de la demanda de energía eléctrica de mediano plazo de El Salvador* utilizó un modelo ARIMA estacional desarrollado por medio de la metodología descrita por Box Jenkins para la predicción de la demanda de

energía eléctrica de mediano plazo de El Salvador obteniendo valores de MAPE aceptables y más precisos que el modelo econométrico que se utilizaba.

Así mismo, las redes neuronales suelen tener popularidad en los pronósticos de demanda debido a su capacidad adaptativa a la complejidad de cualquier problema, Lozano *et al.* (2019), en su estudio *Predicción de demanda de energía eléctrica mediante redes neuronales artificiales* hacen referencia a distintos estudios entre estos *Electric load forecasting by using dynamic neural network* en el cual se desarrolló un modelo de predicción dinámico por medio de *backpropagations* obteniendo como resultados de error de predicción de entre 3.26 y 4.223 %. Mientras que Wang y Zhu, (2015) realizaron una comparación entre cuatro modelos de predicción basados en redes neuronales: redes neuronales *backpropagation* (BPNN), redes neuronales ELMAN, redes neuronales *fuzzy* (FNN) y redes neuronales *wavelet neural*, concluyendo que las BPNN representa el modelo más óptimo con un error de predicción entre el 2.23 y 4.53%; por su parte Lozano *et al.* (2019) con su modelo basado en una red neuronal *feedforward* y *backpropagation* obtuvieron un porcentaje de error medio absoluto del 2.63 % frente a un 4.56 % obtenido por la metodología tradicional utilizada por la empresa Eléctrica. Mientras que Gonzales (2004), en su estudio *Predicción de la demanda eléctrica horaria mediante redes neuronales artificiales* utilizó un modelo de redes neuronales de tipo perceptrón multicapa, obteniendo una capacidad de predicción del 98 % para la demanda eléctrica horaria.

Como se puede observar se han realizado distintos tipos de estudios para determinar la demanda energética, cada uno de ellos utilizando distintas técnicas y métodos de aplicación, esto habla de la flexibilidad que existe en el campo de pronóstico para desarrollar modelos y de la complejidad que un modelo puede tomar; cabe resaltar la importancia de los métodos utilizados para la selección de modelos como lo son los criterios AIC y BIC, y las medidas utilizadas para calcular

la precisión como: MAPE y MPE; además estos estudios hacen hincapié en la necesidad de utilizar más de un modelo para tener punto de comparación.

2. MARCO TEÓRICO

2.1. Series de tiempo

Se le llama serie de tiempo \bar{f} a un conjunto ordenado de valores, $\bar{f} \equiv \{f_j\} = \{\dots, f_{-2}, f_{-1}, f_0, f_1, f_2, \dots\}$; (Ec. 01) los valores f_i están asociados a puntos consecutivos y equitativamente espaciados por $t_j = j\Delta t$ de la variable del dominio t ; este valor de t siempre crece a medida que aumenta j , nunca disminuye o permanece constante. Aunque la variable del dominio puede no ser el tiempo, regularmente es el más utilizado para la selección de muestras, de donde se deriva su nombre.

Los valores de f_i pueden ser considerados resultados del muestreo de una función de tiempo $f(t)$ univaluada en todo el dominio de t , con un intervalo de muestreo Δt (Nava, 2013).

2.1.1. Componentes de la serie de tiempo

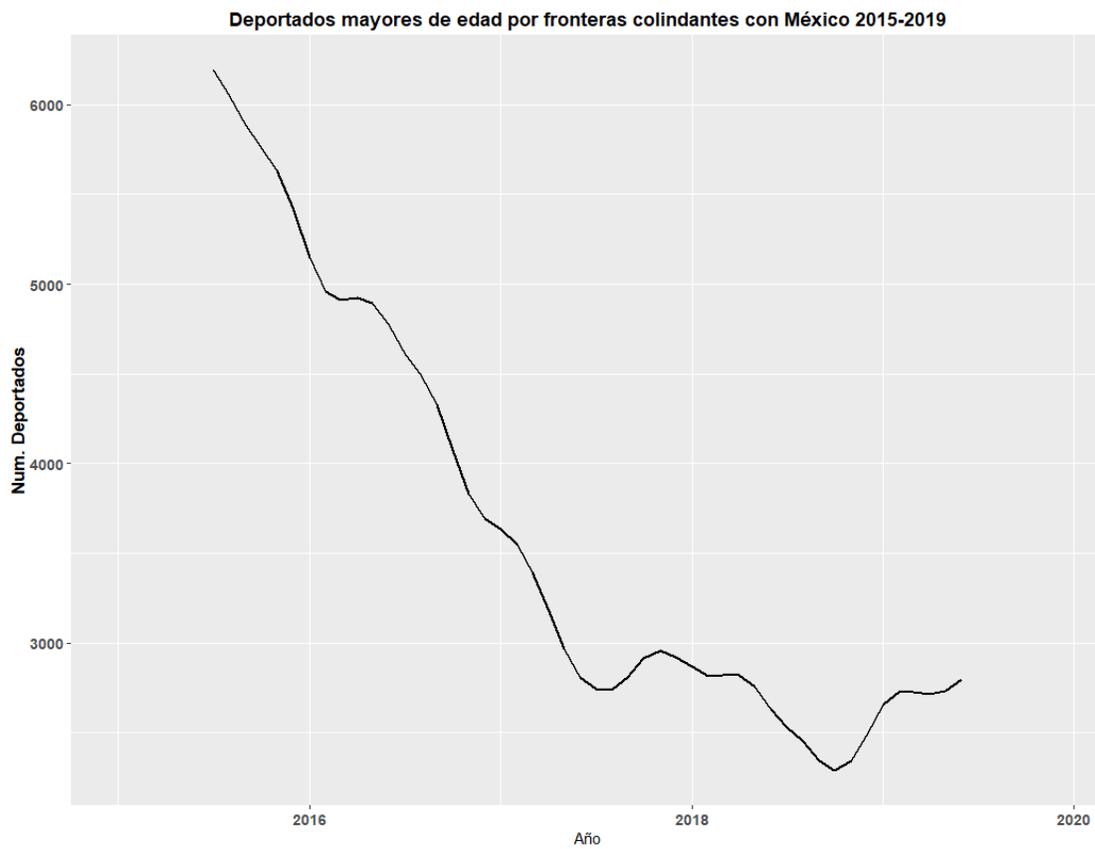
A continuación, se definen y describen los componentes de la serie de tiempo.

2.1.1.1. Tendencia secular

Este componente representa la trayectoria suavizada o patrón gradual que caracteriza la serie en el rango de variación del índice y se encuentra observando la forma funcional de la gráfica de la serie a largo plazo en el tiempo; esta puede

ser: constante, lineal, cuadrática, exponencial, entre otros. Estas variaciones de la serie se consideran consecuencias de aspectos como: cambios en la población, cambios en ingresos, en salud, entre otros (Polanco, 2020).

Figura 1. **Ejemplo de tendencia secular en serie de tiempo**



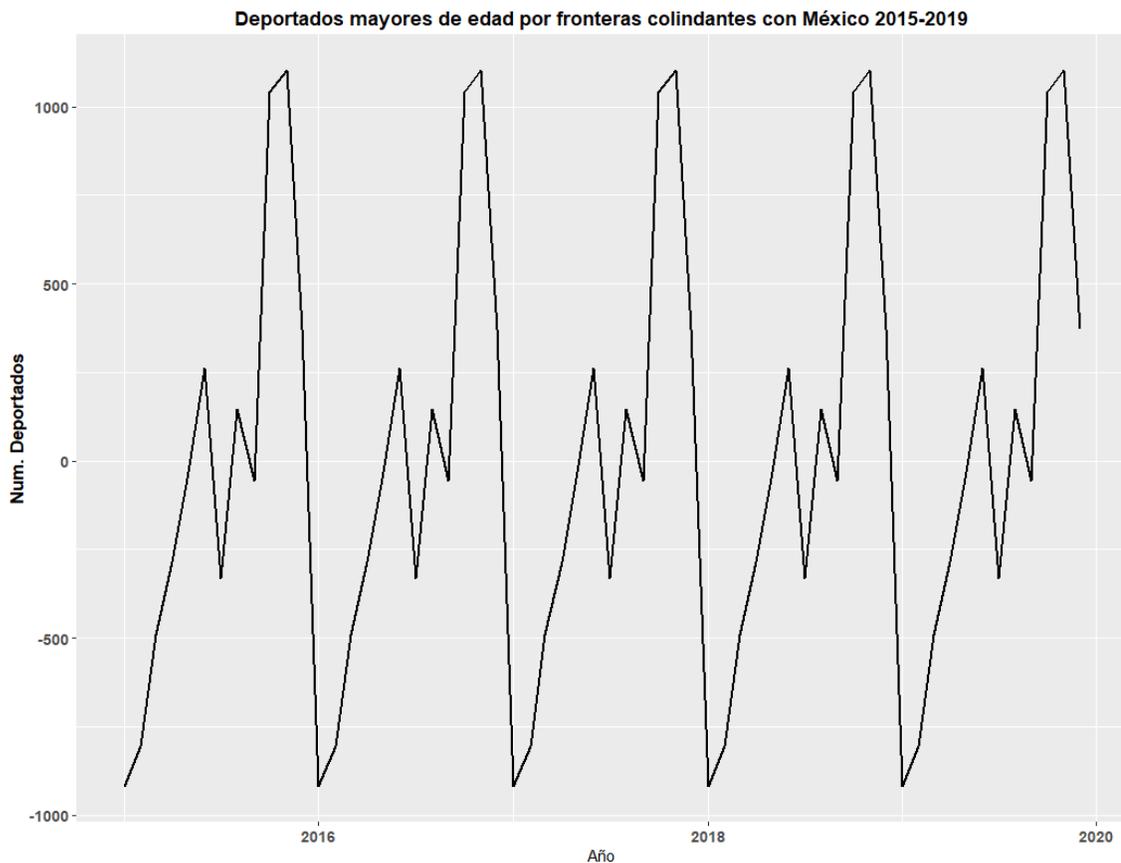
Fuente: elaboración propia.

2.1.1.2. Variación estacional

Este componente de la serie represente la variabilidad de los datos como influencia de las estaciones. Se trata de movimientos periódicos que se repiten un número determinado de ocasiones en un periodo de tiempo, usualmente un año, y este comportamiento permanece de periodo en periodo.

Este efecto puede ser aditivo o multiplicativo, el primero se caracteriza por ser constante y el segundo porque es proporcional a la tendencia (Polanco, 2020).

Figura 2. **Ejemplo de variación estacional en serie de tiempo**

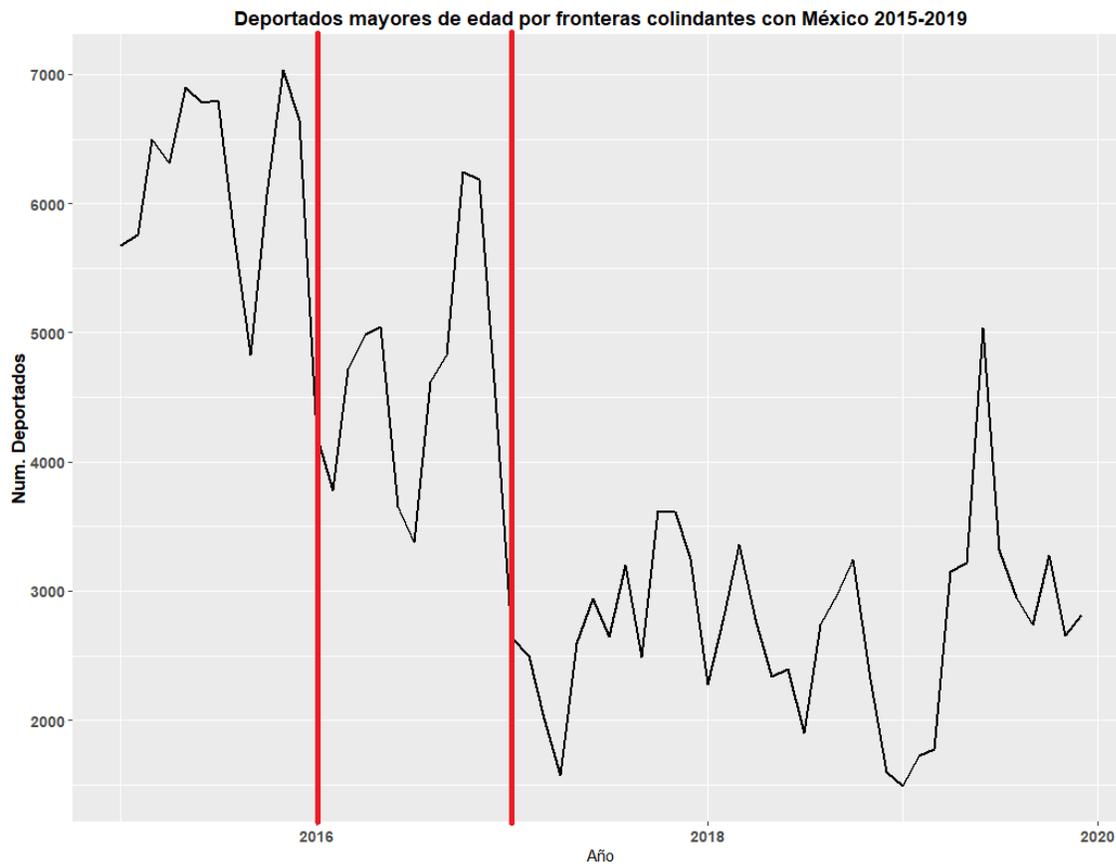


Fuente: elaboración propia.

2.1.1.3. Variación cíclica

Se refiere a movimientos oscilatorios de varios periodos de longitud, estos son regulares y mayores a un año, en otras palabras, con secuencias alternas de puntos tanto arriba como debajo de la línea de tendencia. Usualmente estos se presentan en series que son afectadas por fenómenos físicos o económicos en los cuales existe periodicidad de variable (Polanco, 2020).

Figura 3. **Ejemplo de variación cíclica en serie de tiempo**



Fuente: elaboración propia.

2.1.2. Clasificación descriptiva de las series de tiempo

A continuación, se presenta la clasificación descriptiva de las series de tiempo.

2.1.2.1. Estacionarias

Una serie de tiempo es estacionaria cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza permanecen constantes a través del tiempo. Gráficamente se observa cuando los valores tienden a permanecer alrededor de una media constante y con muy poca variabilidad (Villavicencio, 2012).

2.1.2.2. No estacionarias

A diferencia de las estacionarias acá las series poseen una tendencia y variabilidad cambiante en el tiempo. Los cambios en la media determinan la tendencia creciente o decreciente, la serie no permanece alrededor de la media (Villavicencio, 2012).

2.1.3. Métodos de pronósticos

A continuación, los métodos de pronósticos.

2.1.3.1. Promedios móviles

En esta técnica, los datos se atenúan al obtener la media aritmética de un subconjunto de los últimos datos históricos observados, y la observación o los datos históricos más antiguos se eliminan siempre que se dispone de nuevos datos. Se considera cuándo los valores de la serie oscilan o se mueven alrededor

del valor promedio. De la misma forma, asume que el desplazamiento del tiempo promedio es provocado por un número infinito de perturbaciones ocurridas en el pasado y está ponderado por un factor que se encarga de medirlas.

Un promedio móvil se construye reemplazando cada valor de la serie por la media obtenida con esa observación y algunos de los valores anteriores y siguientes (Camino, 2013).

2.1.3.2. Suavizamiento exponencial

Es el método de suavizamiento más simple, es adecuado para pronósticos en los cuales los datos no tengan tendencia clara o patrón estacional. El suavizado exponencial utiliza el promedio ponderado de la serie temporal pasada como pronóstico, es un caso especial del método de promedios ponderados en donde únicamente se selecciona un factor para la ponderación siendo el de la observación más reciente. El modelo básico de suavizamiento exponencial es el siguiente:

$$F_{i+1} = \alpha Y_t + (1 - \alpha)F_t \quad (\text{Ec. 02})$$

En donde F_{i+1} es el pronóstico de la serie de tiempo para el periodo t+1, Y_t es el valor real de la serie de tiempo en el periodo t, F_t es el pronóstico de la serie de tiempo para el periodo t y α es una constante de suavizamiento dentro del rango de 0 y 1 (Camino, 2013).

2.1.3.3. Holt-Winters

Holt y Winters crearon un modelo para modelar la estacionalidad y tendencia de una serie de datos; este método comprende la implementación de una serie de ecuaciones para incluir el nivel, la tendencia y la estacionalidad. Existen dos variaciones de este método que varían dependiendo de la naturaleza del componente estacional, cuando las variaciones estacionales son constantes a lo largo del tiempo se utiliza el método aditivo y cuando estas cambian proporcionalmente al nivel de la serie se utiliza el multiplicativo. Las ecuaciones para el modelo aditivo son las siguientes:

$$\text{Pronóstico: } y_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)} \text{ (Ec. 03)}$$

$$\text{Nivel: } \ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \text{ (Ec. 04)}$$

$$\text{Tendencia: } b_t = \beta * (\ell_t - \ell_{t-1} - 1) + (1 - \beta *)b_{t-1} \text{ (Ec. 05)}$$

$$\text{Estacional: } s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \text{ (Ec. 06)}$$

En donde k es la parte entera de $(h - 1)/m$, lo cual asegura que los índices estacionales estimados procedan del último año de la muestra. La ecuación de nivel muestra un promedio ponderado entre la observación ajustada estacionalmente $(y_t - s_{t-m})$ y el pronóstico no estacional $(\ell_{t-1} + b_{t-1})$ para un tiempo t . La ecuación estacional muestra un promedio ponderado entre el índice estacional actual $(y_t - \ell_{t-1} - b_{t-1})$ y el índice estacional de la misma temporada el año pasado (Hyndman y Athansopoulos, 2018).

2.1.3.4. Autorregresivos

Es un modelo de regresión múltiple, que utiliza una combinación lineal de predictores para realizar predicciones. En este tipo de modelos, utiliza una combinación lineal de valores pasados, se describe como:

$$y_t = a + by_{t-1} + \varepsilon_t \quad (\text{Ec. 07})$$

en donde a y b son constantes de la serie de tiempo y ε_t es el ruido blanco. Su modelo es muy parecido a un modelo de regresión múltiple, pero utilizando valores rezagados de y_t como predictores. Se refiere a este tipo de modelos como AR (p) donde p indica el orden del modelo (Montes, Calvete y Mantilla, 2016).

2.1.3.5. Media móvil

Es utilizado cuando los datos de la serie de tiempo dependen del ruido blanco asociado a cada una de las observaciones, es decir, que en lugar de usar valores pasados de la variable de interés como en los modelos AR utiliza errores de pronósticos pasados. La fórmula general es:

$$y_t = a + \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q} \quad (\text{Ec. 08})$$

En donde $a, \theta_1, \theta_2, \theta_q$ son parámetros que se deben determinar para el modelo, q define el orden del modelo, este es el número de valores pasados que afectan el cálculo del valor actual (Montes *et al.*, 2016).

2.1.3.6. ARMA/ARIMA

El análisis de modelos AR y MA únicamente considera una componente de la serie de tiempo, lo cual es muy poco común que ocurra pues, la mayoría de las series de tiempo, suelen poseer más de una característica.

Cuando se combina un modelo de autorregresión y un modelo de media móvil, se obtiene un modelo autorregresivo de media móvil, conocido

comúnmente como ARMA (p,q) , donde p y q representan, el orden de autorregresión y el orden de media móvil. La expresión de un modelo ARMA (1, 1) es la siguiente:

$$y_t = a + by_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \quad (\text{Ec. 09})$$

Que es una mezcla de los modelos vistos anteriormente, estos modelos son utilizados para modelar procesos estocásticos estacionados, donde la media y varianza permanece constante en el tiempo. Pero, en algunos casos esto no se cumple, razón por la cual Box y Jenkins plantea los modelos ARIMA, los cuales además de contener un componente autorregresivo y uno de media móvil, incluyen un componente de integración que pretende convertir una serie no estacionaria a estacionaria. El modelo se describe como:

$$y_t = a + by_{t-1} + \dots + b_q y_{t-q} + \varepsilon_t + \theta\varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (\text{Ec. 10})$$

Debido a la cantidad de datos a analizar, la complejidad de los cálculos es enorme y se debe poseer una gran experiencia en el manejo de los datos para poder realizar los cálculos necesarios (Montes *et al.*, 2016).

2.1.3.7. Regresión dinámica

Este tipo de modelos no toman en cuenta únicamente la variable a estudiar, sino que también la influencia de otras variables que pudieran ser relevante. Se considera como un modelo de regresión, en el cual sus errores contienen autocorrelación y siguen un modelo de tipo ARIMA. Su modelo se define como:

$$y_t = \beta_0 + \beta_1 X_{1,t} + \dots + \beta_k X_{k,t} + \eta_t \quad (\text{Ec. 11})$$

En donde los β_k , representan los coeficientes de la regresión y η_t son los errores que siguen un modelo ARIMA (Hyndman y Athansopoulos, 2018).

2.1.4. Correlación

Es una medida de relación de carácter lineal entre dos variables de tipo cuantitativas. Su forma más sencilla es la evaluación de su variación conjunta. Un aspecto importante que notar es que la covariación no necesariamente implica causalidad podría ser fortuita y generada por el efecto de una tercera variable.

En esencia es una medida normalizada de asociación entre dos variables, este índice de correlación r varía entre -1 y 1, ambos extremos indican una relación perfecta entre variables, un valor de 0 indica una nula relación. Los valores positivos indican que las variables varían al mismo sentido y los negativos que varían en sentidos opuestos (Vinuesa, 2016).

2.1.4.1. Coeficiente de Spearman

Es una prueba no paramétrica que mide el grado de asociación o interdependencia entre dos variables discretas, en donde al menos una de ellas es de escala ordinal. Se utiliza comúnmente cuando los datos poseen valores extremos o cuando los datos no poseen una distribución normal. Su valor puede oscilar entre -1 y +1, los valores próximos al 1 indican una correlación fuerte y positiva, mientras que los valores próximos a -1 indican una correlación fuerte y negativa. Los valores cercanos a 0 indican que no hay correlación (Tómas-Sabado, 2010).

2.1.5. Prueba de Kruskal-Wallis

La prueba de Kruskal-Wallis es una prueba no paramétrica, lo que significa que no asume normalidad en los datos; como la mayoría de las pruebas de este tipo, se realiza con datos clasificados, por lo que convierte las observaciones en rangos, el valor más pequeño tiene un rango de 1, mientras que el siguiente más pequeño tienen un rango de 2, y así sucesivamente. La hipótesis nula de la prueba es que los rangos medios de los grupos serán iguales.

Normalmente se utiliza cuando se tiene una variable nominal y una variable de medida, en un experimento en el cual se utilizaría un anova unidireccional, pero la variable de medida no cumple con el supuesto de normalidad. Para calcularlo se realiza la suma de los rangos para cada grupo y el estadístico de prueba H , el cual viene dado por una fórmula que representa básicamente la varianza de los rangos entre los grupos, con un ajuste por el número de empates (McDonald, 2014).

2.1.6. Análisis exploratorio

A continuación, lo concerniente al análisis exploratorio.

2.1.6.1. Media

Es el promedio de un conjunto de datos, es conocida como una medida de tendencia central. Se calcula como:

$$\bar{x} = \frac{\sum_{i=0}^n x_i}{N} \text{ (Ec. 12)}$$

En donde N es la cantidad de datos y x_i son los datos. (INEI, 2006)

2.1.6.2. Mediana

Es el valor que divide a los datos ordenados aproximadamente por la mitad (INEI, 2006).

2.1.6.3. Moda

Es una medida de tendencia central, es el valor que más se repite en la serie de datos, es la única variable que se puede estudiar como una variable cualitativa (INEI, 2006).

2.1.6.4. Desviación estándar

Es una medida de dispersión obtenida al sacar la raíz cuadrada de la varianza, se define como:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \text{ (Ec. 13)}$$

En donde n es la cantidad de datos, x_i es un dato y \bar{x} es la media de los datos. (INEI, 2006)

2.1.6.5. Máximo

Es el valor más grande en la serie de datos.

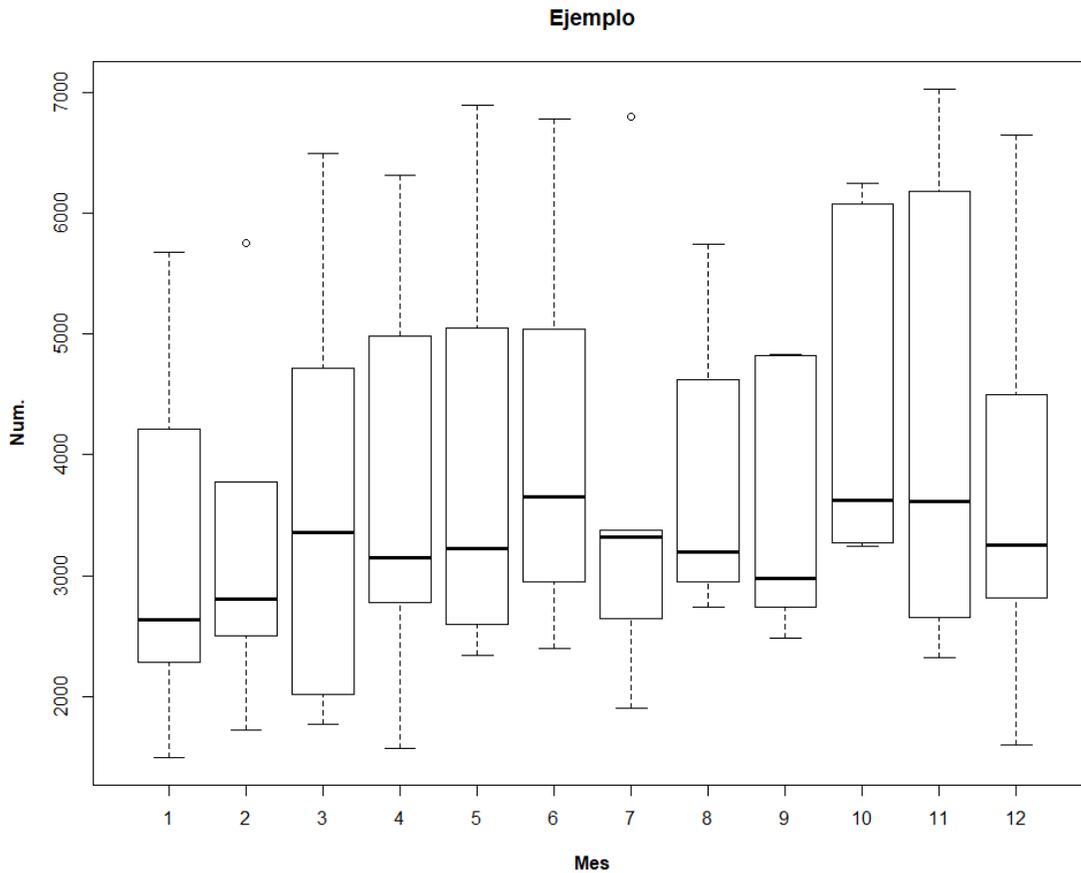
2.1.6.6. Mínimo

Es el valor más pequeño en la serie de datos.

2.1.6.7. Diagrama de cajas

También conocido como *boxplot*, es un gráfico que permite darse una idea visual de la distribución de los datos. Se puede observar si existe simetría, ver el grado de variabilidad y detectar datos atípicos; en resumen, es una gráfica que nos permite visualizar la distancia en la que se encuentran los datos y como se distribuyen (INEI, 2006).

Figura 5. Ejemplo de diagrama de cajas



Fuente: elaboración propia.

2.1.7. Autocorrelación

Villavicencio, (2012) la presenta como una forma de medir la dependencia entre dos valores, esta mide la correlación entre dos variables que están separadas por k periodos, se define como:

$$\rho_j = \text{corr}(X_j, X_{j-k}) = \frac{\text{cov}(X_j, X_{j-k})}{\sqrt{V(X_j)}\sqrt{V(X_{j-k})}} \text{ (Ec. 14)}$$

2.1.8. Autocorrelación parcial

Villavicencio, (2012) expone que la autocorrelación parcial evalúa la correlación entre dos variables que se encuentran separadas por k periodos, cuando no se toma en cuenta la dependencia que existe debido a los retornos intermedios entre valores; está dada por:

$$\pi_j = \text{corr}(X_j, X_{j-k}/X_{j-1}X_{j-2}\dots X_{j-k+1}) = \frac{\text{cov}(X_j - \hat{X}_j, X_{j-k} - \hat{X}_{j-k})}{\sqrt{V(X_j - \hat{X}_j)}\sqrt{V(X_{j-k} - \hat{X}_{j-k})}} \text{ (Ec. 15)}$$

2.1.9. Prueba de Dickey-Fuller

La prueba de Dickey-Fuller prueba si $\phi = 0$ en un modelo de tipo:

$$\Delta y_t = y_t - y_{t-1} = \alpha + \beta t + \gamma y_{t-1} + e_t \text{ (Ec. 16)}$$

En donde y_t es la información. Es un modelo del cual se puede hacer una regresión lineal de Δy_t entre t y y_{t-1} y probar si γ es diferente de 0, si $\gamma = 0$ entonces tenemos un proceso de caminata aleatoria. Si no y se encuentra entre

$-1 < 1 + \gamma < 1$, entonces tenemos un proceso estacionario (Holmes, Scheurell y Wardm, 2020).

2.1.10. Análisis residual

Un residual es la diferencia entre un valor observado real y el valor estimado por un modelo de pronóstico. El residual puede ser considerado como un error aleatorio observado. Este análisis permite cotejar ciertas suposiciones:

- Normalidad de los errores
- Valores anormales en la distribución de errores
- Existencia de varianza constante
- Independencia de errores (Díaz, 2005).

Dentro del análisis de series de datos, es necesario realizar la validación de los supuestos de aleatoriedad y normalidad en los residuos para determinar si el modelo encontrado es o no el adecuado para realizar pronósticos. Comúnmente se utiliza la prueba de Rachas para medir la aleatoriedad y la prueba de Smirnov Kolmogorv para determinar su normalidad (Bello y Martínez, 2007).

2.1.11. Prueba de Kolmogorov-Smirnov

La prueba de Kolmogorov-Smirnov fue nombrada así en honor a los estadísticos que la desarrollaron, se trata de un método no paramétrico utilizado para probar si existe una diferencia significativa entre una distribución observada y una distribución teórica.

Es decir, es una medida de la bondad de ajuste de una distribución de frecuencias teóricas, tal y como lo es la prueba de chi-cuadrado, sin embargo

esta prueba es más poderosa, y es más fácil de usar debido a que no requiere que los datos se agrupen de alguna manera.

El estadístico de prueba D_n , ofrece la facilidad de juzgar que tan cerca está la distribución observada de la esperada, porque la distribución de probabilidad D_n depende únicamente del tamaño de la muestra n , pero es totalmente independiente de la distribución de frecuencias esperadas. Es usualmente utilizada para realizar pruebas de normalidad (Levin y Rubin, 2004).

2.1.12. Prueba de Ljung-Box

Es una prueba que permite corroborar de manera conjunta si todos los coeficientes de autocorrelación son simultáneamente iguales a cero, lo que significa que existe independencia, se define como:

$$LB = n(n + 2) \sum_{k=1}^m \left(\frac{\hat{\rho}_k^2}{n-k} \right) \sim \chi^2_{(m)} \quad (\text{Ec. 17})$$

En donde n es el tamaño de la muestra y m la longitud del rezago.

Las hipótesis son:

- H_0 : Las autocorrelaciones son independientes
- H_a : Las autocorrelaciones no son independientes

Si el Q calculado sobrepasa el valor crítico de la tabla al nivel de significancia utilizado, se rechaza la hipótesis nula de que todos los coeficientes son iguales a cero (Villavicencio, 2020).

2.1.13. Criterio de Información de Akaike

Es una medida de la calidad del modelo, la cual presenta una estimación de la información perdida al utilizar el modelo planteado para describir los datos. Mientras más pequeño sea el valor del AIC, menor información se pierde, por lo tanto, el modelo será más apropiado. Su fórmula se define como:

$$AIC = 2k - 2\ln(L) \text{ (Ec. 18)}$$

En donde k es el número de parámetros en el modelo y L es el máximo valor de la función de verosimilitud (Montes *et al.*, 2016).

2.1.14. Criterio de Información de Bayes

Es una medida de la calidad del modelo, está relacionado con la desviación de los datos reales con respecto al modelo evaluado, lo cual incluye algunas correcciones bayesianas. Esta definición incluye la variación no explicada de la variable dependiente, por lo tanto, entre más pobre sea la descripción del modelo, mayor será su valor. Está dado por:

$$BIC = k \ln(n) - 2\ln(L) \text{ (Ec. 19)}$$

En donde k es el número de parámetros en el modelo, L es el máximo valor de la función de verosimilitud y n es el número de observaciones (Montes *et al.*, 2016).

2.1.15. Error absoluto porcentual medio

Es un método utilizado para evaluar las técnicas de pronósticos, se calcula al encontrar el error absoluto en cada periodo, dividiéndolo entre el valor real observado para ese periodo y luego realizando el promedio de los errores porcentuales absolutos. Es útil cuando la magnitud de las variables es importante para la evaluación del pronóstico, ya que este proporciona una indicación de que tan grandes son los errores de pronóstico con comparación de los valores reales en la serie. Está dado por:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{y}_t)}{Y_t} \text{ (Ec. 20)}$$

En donde Y_t es el valor real en el periodo t , n es la cantidad de datos y \hat{y}_t es el valor de pronóstico para el periodo de tiempo t (Hanke y Wichem, 2006).

2.1.16. Raíz del error cuadrado medio

Se utiliza para evaluar las técnicas de pronósticos este se calcula como la raíz cuadrada de la media de todos los errores al cuadrado. Está dado por:

$$RCEM = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{y}_t)^2}{n}} \text{ (Ec. 21)}$$

En donde Y_t es el valor real en el periodo t , n es la cantidad de datos y \hat{y}_t es el valor de pronóstico para el periodo de tiempo t (Soret y Gimenez, 2013).

2.2. Redes neuronales

Las redes neuronales artificiales, se puede decir que son un modelo matemático compuesto por un conjunto de elementos con la capacidad de procesar datos, este está organizado por niveles, básicamente son redes interconectadas de elementos simples, que poseen una organización jerárquica, las cuales tienen la capacidad de interacción entre sí, simulando el proceso que realizan las neuronas del sistema nervioso.

Estas brindan:

- Aprendizaje Adaptativo
- Auto-organización
- Tolerancia a fallos
- Operación en tiempo real (Matich, 2001).

2.2.1. Elementos básicos

A continuación, la definición de cada uno de los elementos básicos.

2.2.1.1. Capas

Está formada por neuronas interconectadas; la capa de inicio se conoce como capa de entrada, es por donde los datos ingresan; la capa oculta es la que realiza el procesamiento de los datos y la capa de salida es la capa que ofrece el resultado final (Matich, 2001).

2.2.1.2. Función de entrada

Una neurona trata a un gran número de valores de entrada como si se trataran de uno solo, a este suceso se le conoce como entrada global. El problema de cómo se pueden combinar estas entradas dentro de la entrada global; se soluciona por medio de la función de entrada, la cual se calcula a partir de un vector de entradas, se describe como:

$$\text{input } i = (in_{i1} \cdot w_{i2}) * (in_{i2} \cdot w_{i2}) * \dots (in_{in} \cdot w_{in}) \quad (\text{Ec. 22})$$

donde: * representa al operador apropiado, el cual puede ser: máximo, sumatoria, producto, entre otros; n el número de entradas a la neurona N_i y w_i al peso.

Los valores de entrada se multiplican por los pesos anteriores, lo que provoca que los pesos que no están restringidos cambian la medida de influencia que tienen dichos valores de entrada (Matich, 2001).

2.2.1.3. Función de activación

Las neuronas artificiales tienen la habilidad de estar activas o inactivas como es el caso de las neuronas del sistema nervioso, estas pueden tomar cualquier valor dentro de un conjunto determinado. La función de activación es la que calcula el estado de actividad de una neurona; transformando la entrada global en un valor de activación.

Las funciones de activación más comúnmente utilizadas son:

- Función lineal
- Función sigmoide
- Función tangente hiperbólica (Matich, 2001).

2.2.1.4. Función de salida

El resultado final de la función de salida es el resultado del procesamiento de la neurona; por lo que esta función determina qué valor se transfiere a las neuronas vinculadas. Si la función de activación permanece por debajo de un umbral fijado, no se producirá ninguna salida a la siguiente neurona. Usualmente los valores de salida están comprendidos en el rango de (0,1) o (-1,1). Entre las funciones más comunes están:

- Función de identidad
- Función binaria (Matich, 2001).

2.3. Demanda eléctrica

Es la cantidad de potencia consumida en un periodo determinado de tiempo. La demanda de una instalación eléctrica es tomada como un valor promedio en un intervalo determinado; este periodo de tiempo durante el cual se calcula el valor promedio es llamado intervalo de demanda; su duración depende del valor de demanda que se quiere conocer. La demanda se puede expresar en kW, MW, entre otros (Ariza, 2013).

2.3.1. Factores que afectan la demanda eléctrica

A continuación, se enumeran los factores con mayor impacto en la demanda de energía eléctrica.

2.3.1.1. Clima

Los factores climáticos suelen ser una fuente importante en el incremento del consumo de energía eléctrica; dentro de las características meteorológicas

se toman en cuenta aspectos como el viento, la pluviosidad, la temperatura, humedad, etc. El más relevante entre estas condiciones es la temperatura debido a que una gran influencia sobre las fuentes de consumo eléctrico como lo son los sistemas de calefacción, sistemas de congelación, refrigeradores, entre otros (Gonzales, 2004).

2.3.1.2. Calendario

La época del año, día, hora o mes puede variar el consumo de energía para un usuario. Ya que de forma evidente no se tiene el mismo consumo de energía al medio día que a las ocho de la noche. Por lo tanto, esta puede variar dependiendo de:

- Hora del día
- Día de la semana
- Fin de semana
- Descansos
- Mes del año (Gonzales, 2004).

2.3.1.3. Información económica

Debido a la importancia que tiene la energía eléctrica en la actividad económica de cualquier país existe una evidente relación entre los indicadores económicos y la demanda de energía eléctrica de un país (Gonzales, 2004).

2.3.1.4. Factores no predecibles

Siempre existe un grado aleatoriedad en cualquier aspecto evaluado, por lo cual se debe contemplar la existencia de una serie de factores no predecibles

que pueden afectar sensiblemente la demanda; como lo podrían ser: manifestaciones, ferias, apagones, catástrofes, entre otros (Gonzales, 2004).

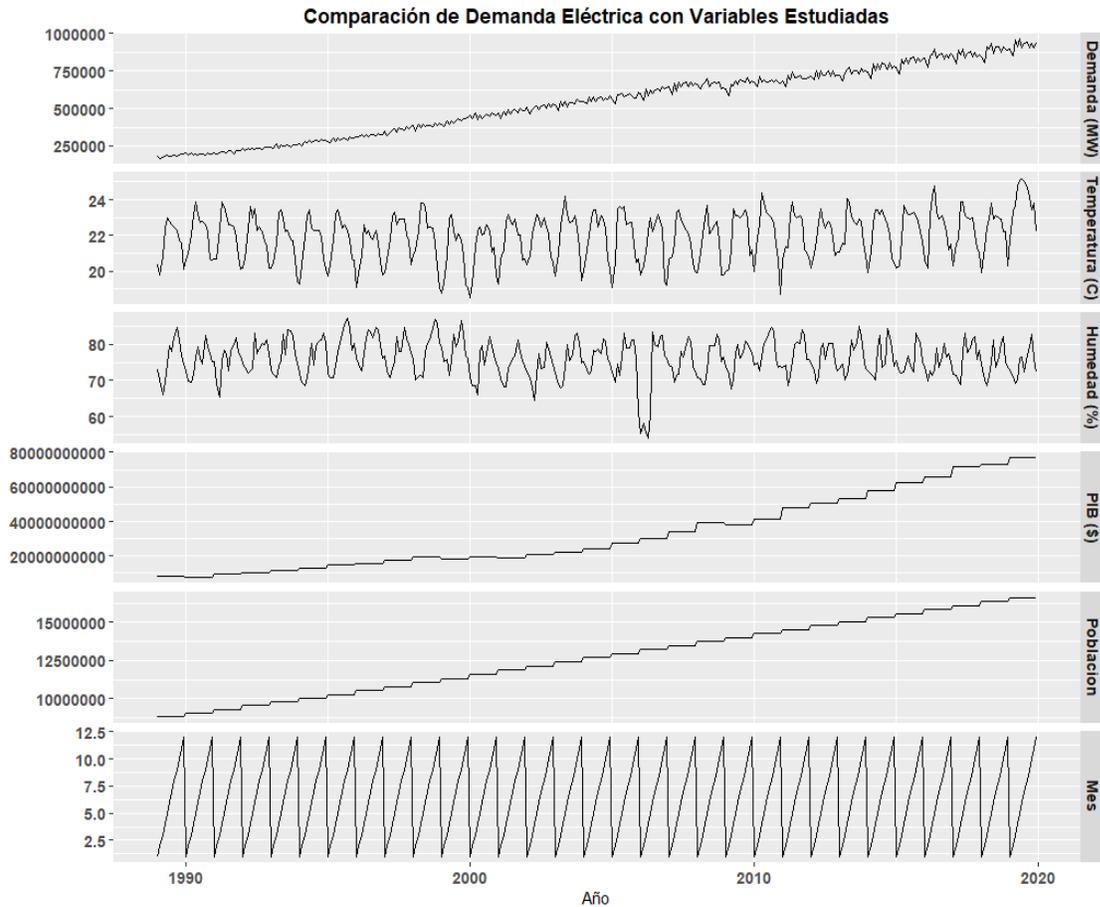
3. PRESENTACIÓN DE RESULTADOS

De acuerdo con los objetivos planteados se presentan los siguientes resultados.

3.1. Objetivo 1. Identificar las variables que influyen en el comportamiento de la demanda de energía eléctrica en Guatemala, por medio de un análisis de correlación multinomial

En el análisis se incluyeron distintos tipos de variables, como climatológicas se utilizó la temperatura y humedad promedio, el PIB (Producto interno bruto) como una variable económica, el número de población como una variable demográfica y además como variable calendario el mes del año; la selección de estas variables se determinó a partir de la revisión de antecedentes con estudios similares.

Figura 6. **Comparación de demanda eléctrica con variables estudiadas**



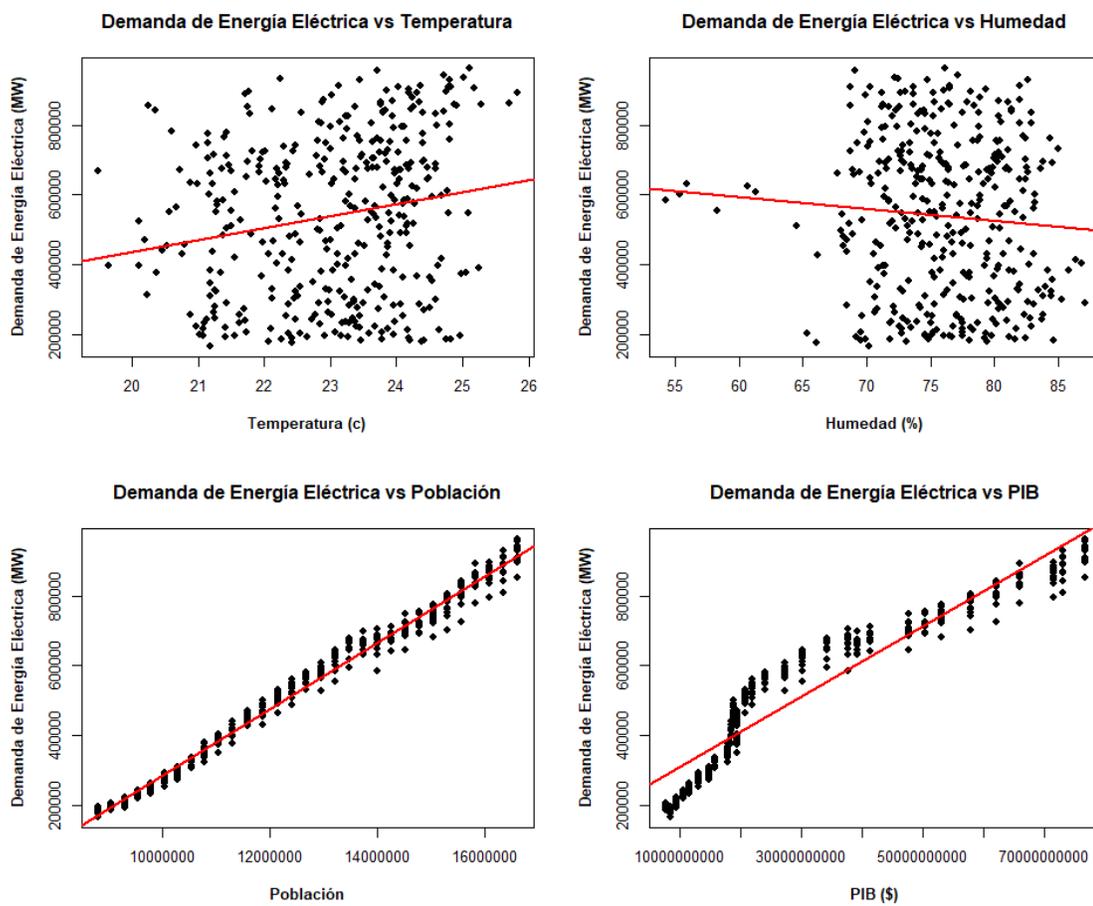
Fuente: elaboración propia.

Como se observa en la figura 6, vemos una tendencia creciente en la demanda de energía eléctrica, el PIB, y el número de población; mientras que las variables de temperatura y humedad parecen ser series estacionarias con cierta estacionalidad anual.

La influencia de las variables propuestas en la demanda de energía eléctrica se determinó por medio de un análisis correlacional para las variables cuantitativas utilizando el coeficiente de Spearman y por medio de un análisis de Kruskal-Wallis para la variable categórica (mes del año);

además se apoyó con un análisis gráfico para identificar tendencias y relaciones entre las variables estudiadas, con ayuda del gráfico de dispersión y el gráfico de *boxplot* para determinar las diferencias entre las medias de la demanda de energía eléctrica mensual.

Figura 7. **Diagramas de dispersión para variables estudiadas**



Fuente: elaboración propia.

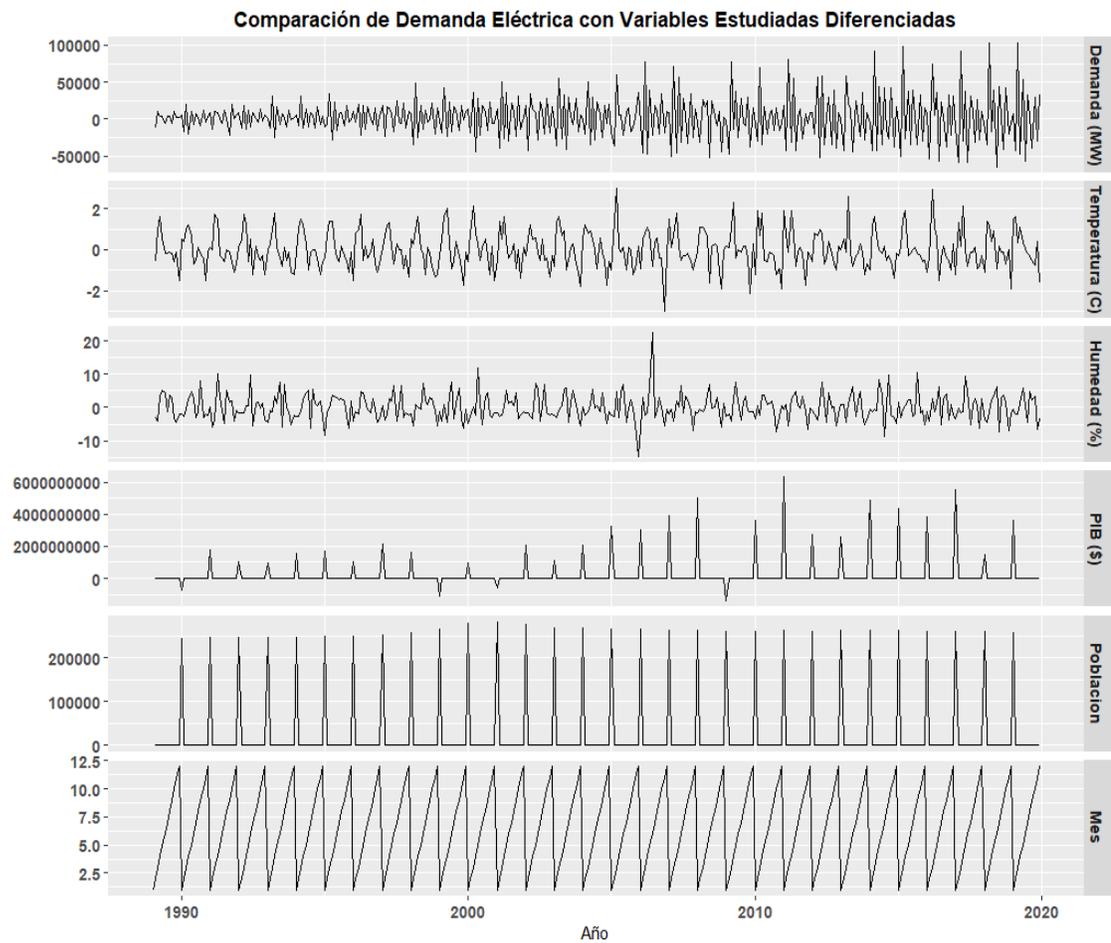
Tabla II. **Pruebas de correlación de las variables estudiadas contra la demanda de energía eléctrica**

Variable	Correlación	P-Valor
Temperatura	0.20	>0.00
Humedad	-0.06	0.21
Población	0.99	>0.00
PIB	0.99	>0.00

Fuente: elaboración propia.

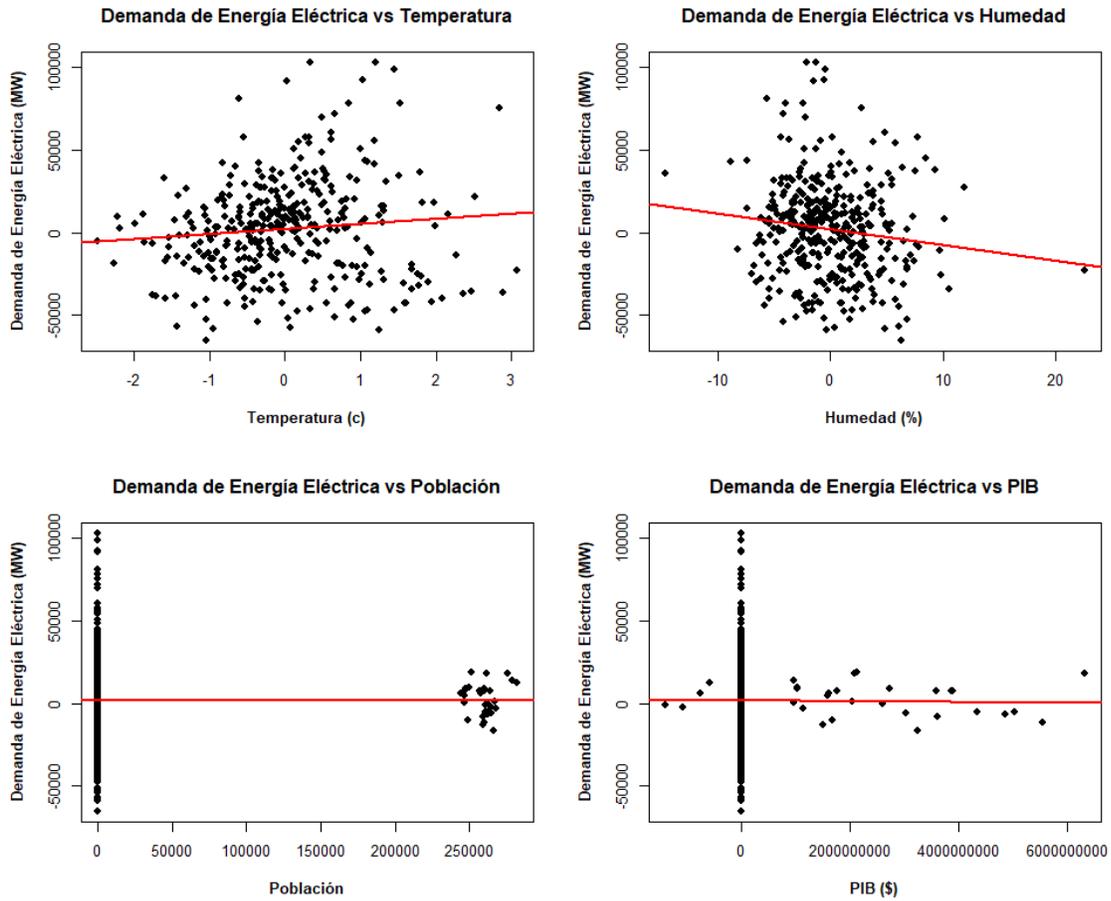
Como se observa en la figura 7, la demanda de energía eléctrica tiene una fuerte relación positiva con las variables de tendencia creciente (PIB y población), mientras que una relación muy baja con las variables de humedad y temperatura; las pruebas de correlación únicamente nos demuestran que la variable humedad no tiene una correlación significativa con nuestra variable de interés ($p\text{-valor} > 0.05$); para comprobar que las relaciones obtenidas no eran espurias se utilizó la diferenciación en las series para eliminar la tendencia y evaluar la correlación en las series diferenciadas.

Figura 8. **Comparación de demanda eléctrica con variables estudiadas diferenciadas**



Fuente: elaboración propia.

Figura 9. Diagramas de dispersión de variables estudiadas diferenciadas



Fuente: elaboración propia.

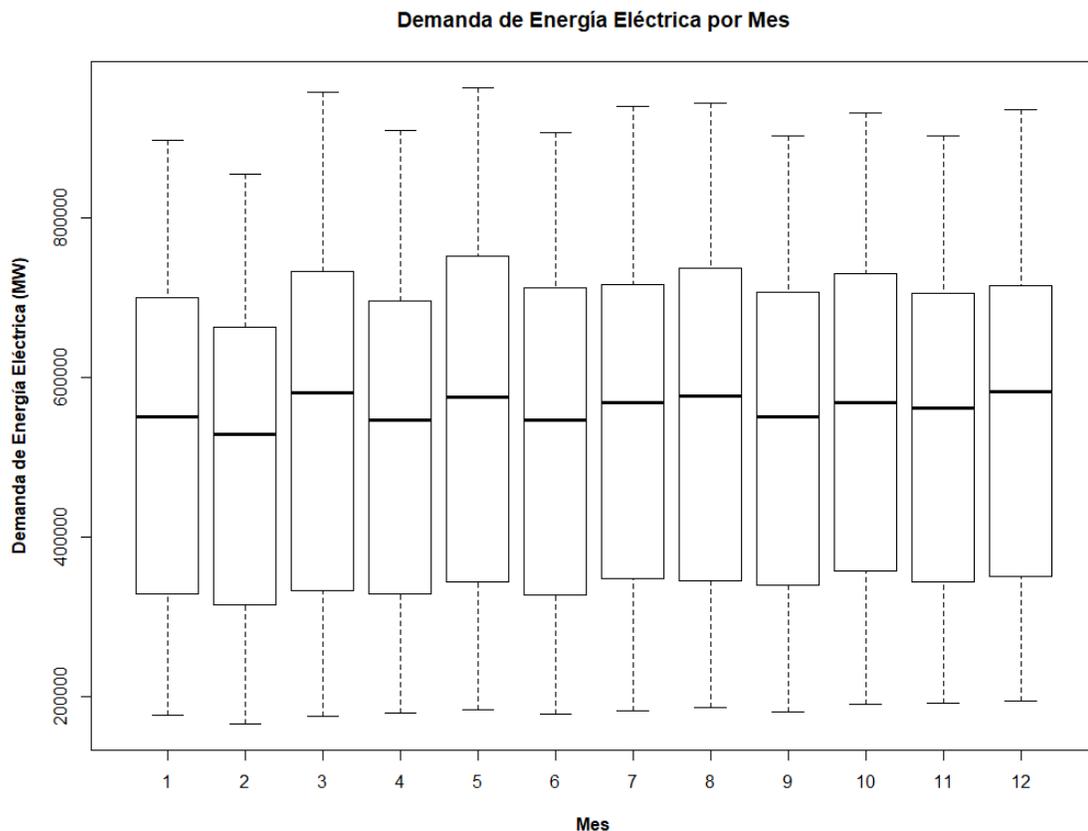
Tabla III. Pruebas de correlación de las variables estudiadas diferenciadas contra la demanda de energía eléctrica

Variable	Correlación	P-Valor
Temperatura	0.13	0.01
Humedad	-0.13	0.01
Población	>0.00	0.88
PIB	>0.00	0.99

Fuente: elaboración propia.

La diferenciación eliminó las fuertes correlaciones existentes entre la demanda de energía eléctrica y las variables de PIB y población, es decir, estas eran relaciones espurias debido a la tendencia en las series, mientras que las variables de temperatura y humedad mantuvieron sus relaciones débiles, positiva en el caso de la temperatura y negativa para la humedad; observando las pruebas de correlación de la tabla II, se determina que las variables PIB y población no tienen una correlación significativa con la demanda de energía eléctrica (p -valor >0.05).

Figura 10. **Boxplot de la demanda de energía eléctrica por mes**



Fuente: elaboración propia.

Figura 11. **Prueba de Kruskal-Wallis para la demanda de energía eléctrica por mes**

```
kruskal-wallis rank sum test
data: demanda by mes
kruskal-wallis chi-squared = 2.1468, df = 11, p-value = 0.9979
```

Fuente: elaboración propia.

En la figura 10, se observa un diagrama de *boxplot* para la demanda de energía eléctrica por mes, se logra identificar que existe una variación mensual en el consumo de energía eléctrica presentando un menor consumo en febrero y un mayor consumo en marzo, además no se observa a simple vista una diferencia significativa entre el promedio de consumo mensual para los diferentes meses. Por medio del análisis de Kruskal-Wallis de la figura 11, se determina que el factor mes no influye en la demanda de energía eléctrica, debido a que se comprobó que no existía diferencia significativa entre los promedios de consumo para cada mes ($p\text{-valor} > 0.05$).

3.2. Objetivo 2. Construir un modelo de predicción con el mejor ajuste para la estimación de la demanda de energía eléctrica en Guatemala, por medio de la comparación de distintos modelos de series de tiempo y uno de redes neuronales

A continuación, se presentan los elementos que se tomaron en cuenta para poder realizar la comparación de distintos modelos de series de tiempo y uno de redes neuronales.

3.2.1. Análisis de la serie de tiempo

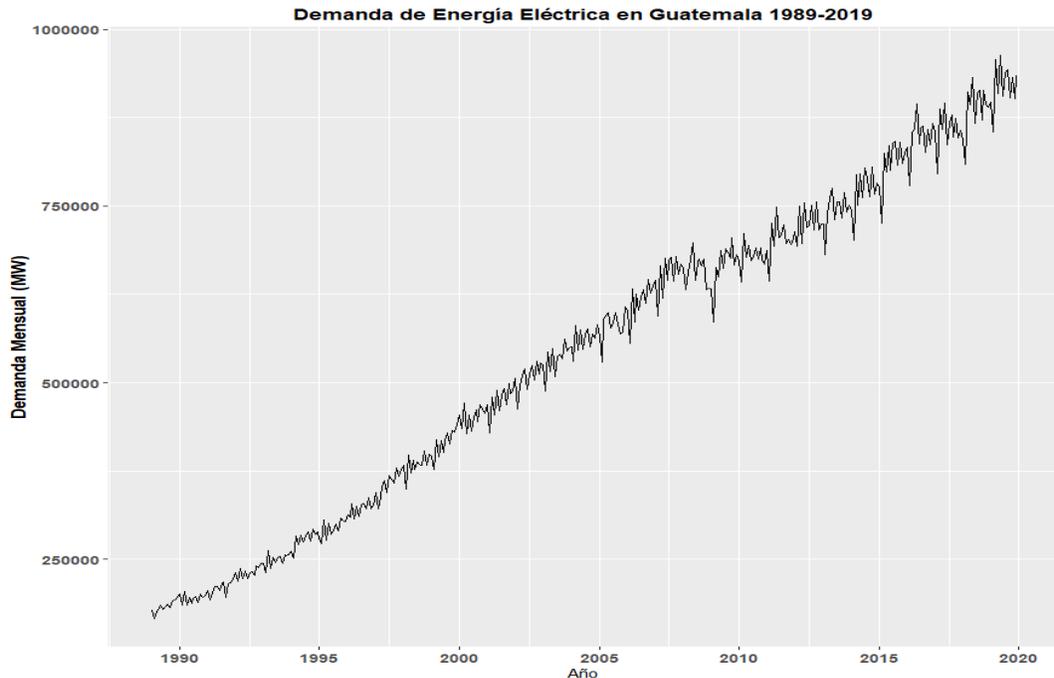
Se realizó un análisis gráfico y descriptivo de la serie de tiempo para identificar sus componentes (tendencia, estacionalidad, ciclos y aleatoriedad), con el fin de determinar el modelo más adecuado por utilizar.

Tabla IV. Resumen de los datos

Mínimo	166,114 MW
Primer Cuantil	327,387 MW
Mediana	558,165 MW
Media	540,836 MW
Tercer Cuantil	715,743 MW
Máximo	963,069 MW
Desviación Estándar	225,469.2 MW

Fuente: elaboración propia.

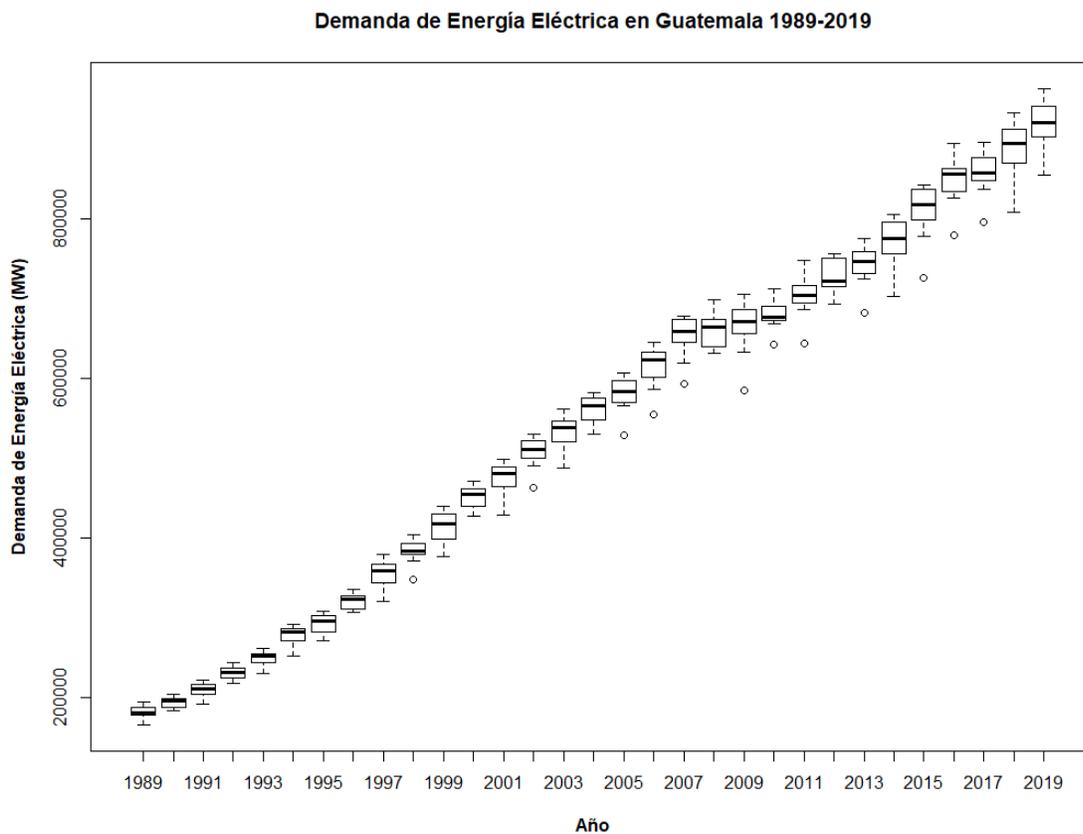
Figura 12. Demanda de energía eléctrica en Guatemala 1989-2019



Fuente: elaboración propia.

Como se puede observar en la figura 12, la demanda mensual de energía eléctrica presenta una tendencia creciente con estacionalidad anual, en el año 2009 se nota un salto abrupto en la tendencia; además encontramos que la serie de datos tienen un valor mínimo de 166,144 MW que se presenta en el primer año de la serie y un valor máximo de 963,069 MW con una desviación estándar de 225,469.2 MW.

Figura 13. **Demanda de energía eléctrica en Guatemala 1989-2019**



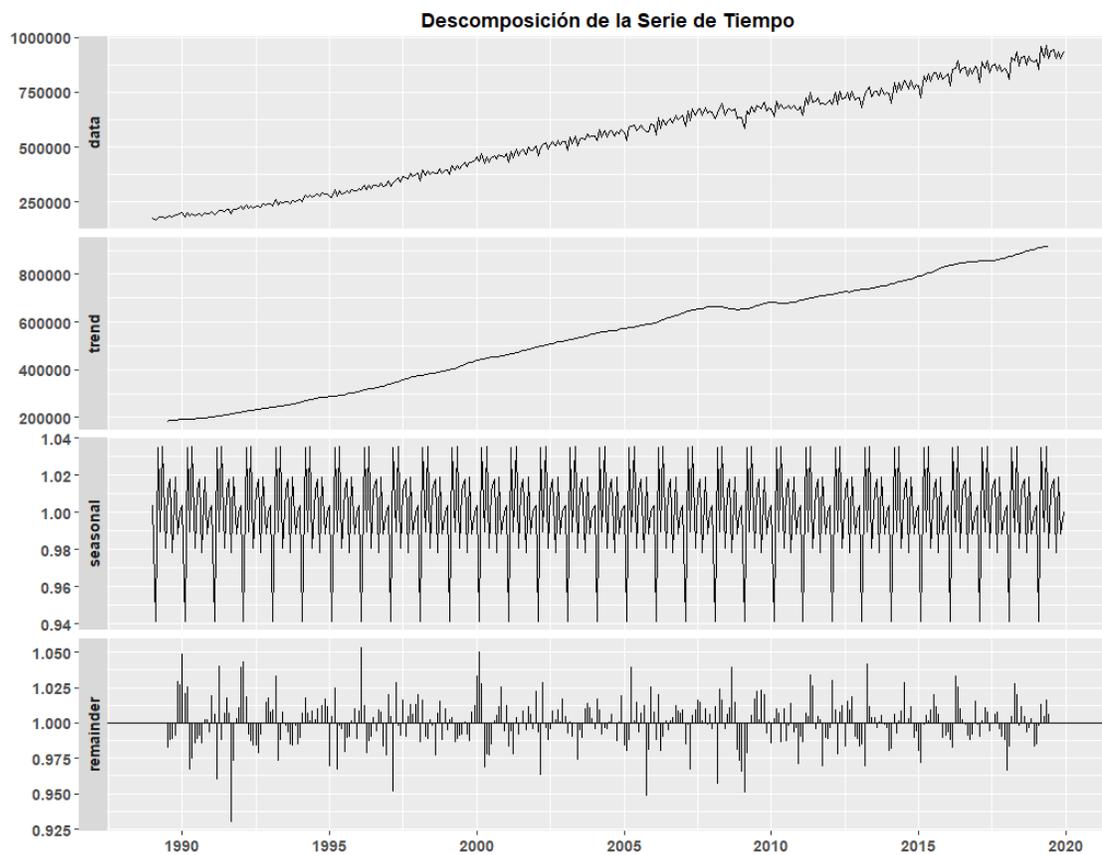
Fuente: elaboración propia.

En la figura 13 se observa un diagrama de *boxplot* para la demanda de energía eléctrica anual, podemos apreciar que la variación en los datos aumenta conforme pasan los años, esto debido a que las variaciones estacionales

incrementan con el tiempo, lo cual nos da la idea que tienen un comportamiento multiplicativo; de igual manera podemos apreciar ciertos valores atípicos para los distintos años.

Se realizó una descomposición de la serie de tiempo con la ayuda del *software* R, para identificar de mejor manera las características de tendencia, estacionalidad y aleatoriedad.

Figura 14. **Descomposición de la serie de tiempo de la demanda de energía eléctrica en Guatemala 1989-2019**

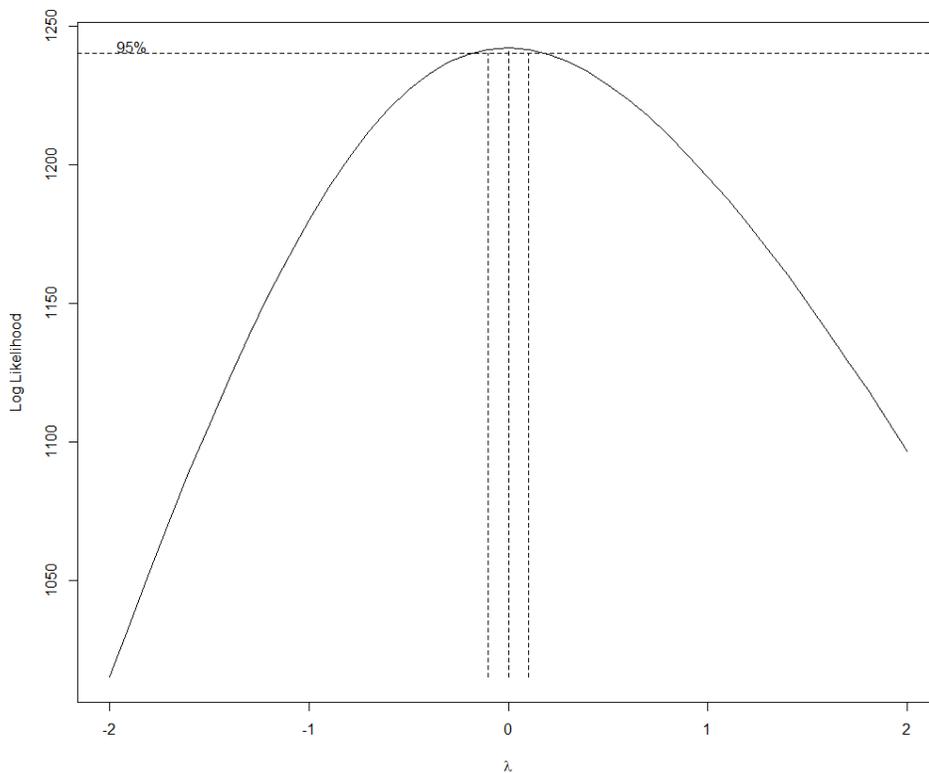


Fuente: elaboración propia.

Se observa en la figura 14 que la serie presenta una tendencia creciente con apariencia lineal, con una estacionalidad anual presentando los valores mínimos al comienzo del año y una componente aleatoria que se puede observar en el cuarto recuadro.

Se aplicó una prueba de *Box-Cox* para identificar si era necesario transformar la serie de tiempo, con el fin de estabilizar la varianza, e identificar por medio de la lambda cuál era la transformación idónea.

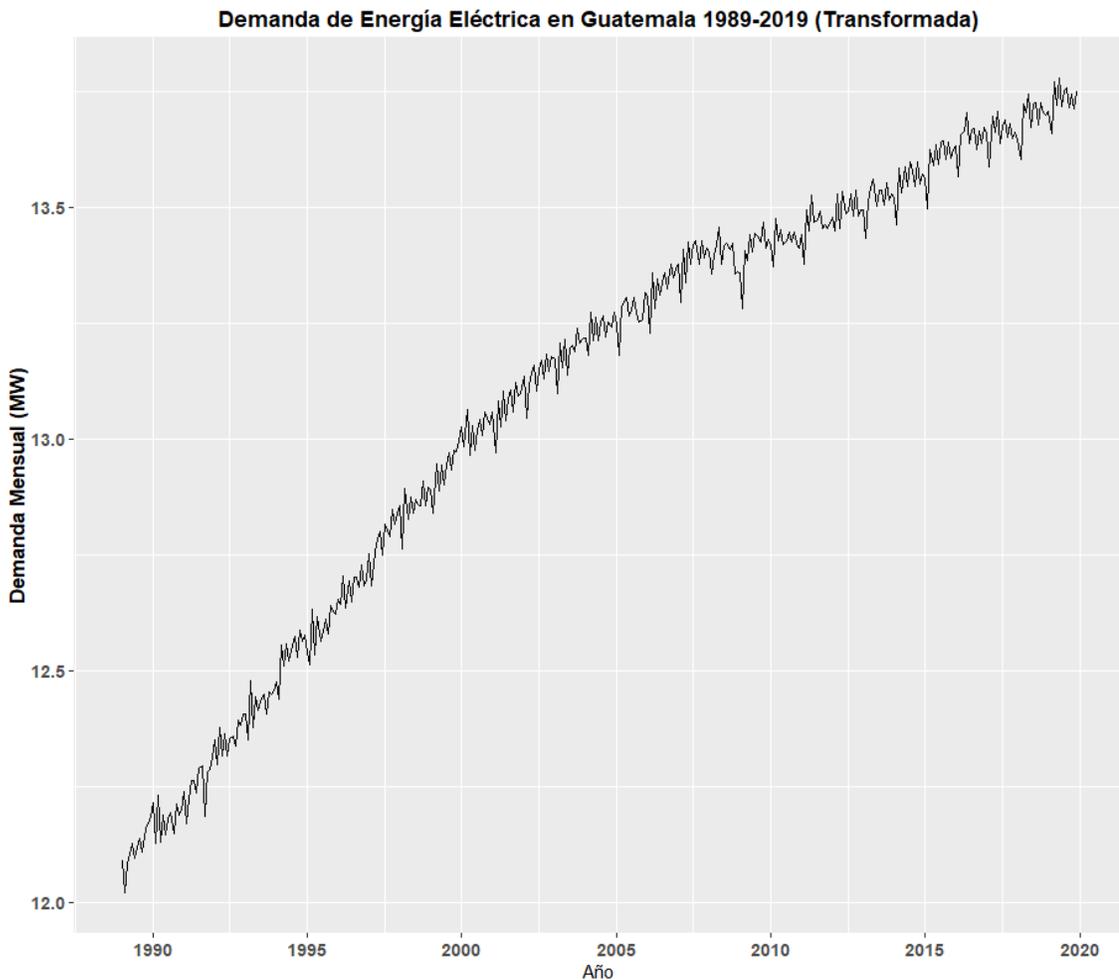
Figura 15. **Prueba de Box-Cox**



Fuente: elaboración propia.

El resultado de la prueba de Box-Cox muestra que la lambda se encuentra cercano a 0, por lo cual se realizó una transformación logarítmica para estabilizar la varianza.

Figura 16. **Demanda de energía eléctrica en Guatemala 1989-2019 con transformación logarítmica**

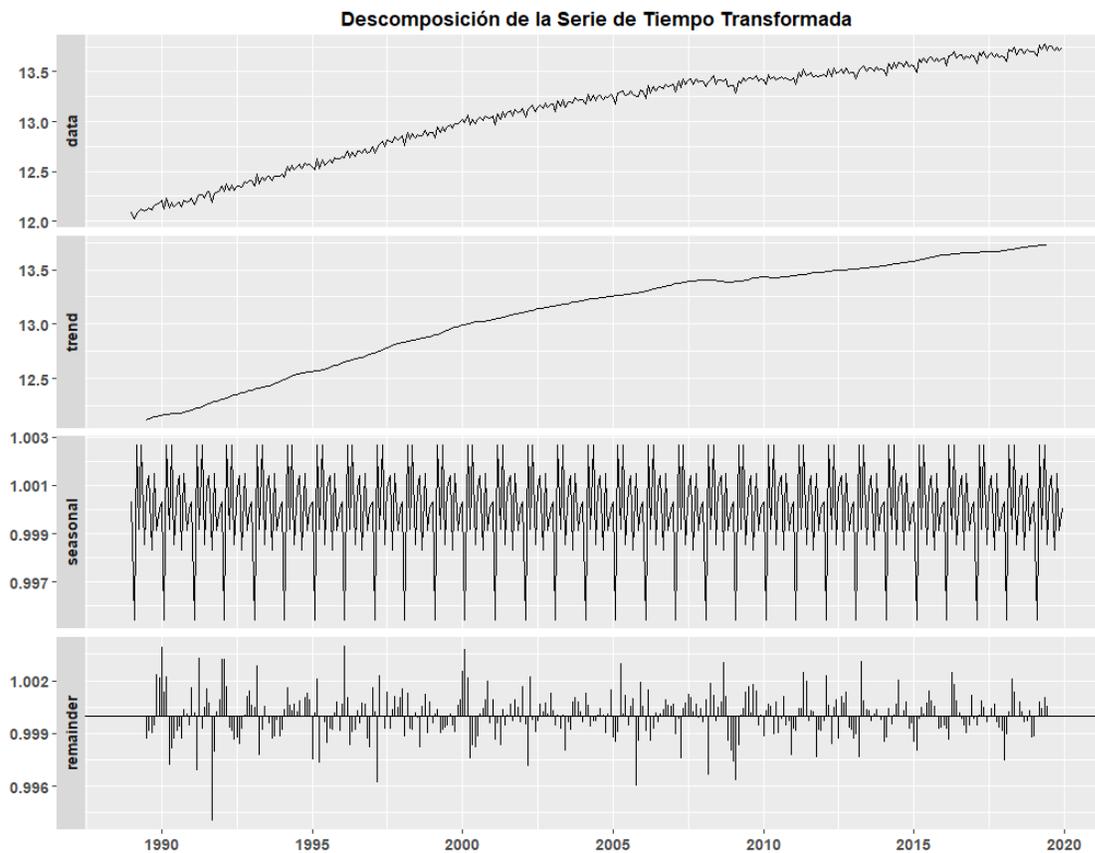


Fuente: elaboración propia.

Luego de la transformación logarítmica podemos apreciar como ahora se observa una varianza constante a lo largo de la serie de tiempo, con picos de

escala similar a lo largo de todo el tiempo; además se observa una curvatura en la tendencia a causa de la transformación.

Figura 17. **Descomposición de la serie de demanda de energía eléctrica en Guatemala 1989-2019 con transformación logarítmica**

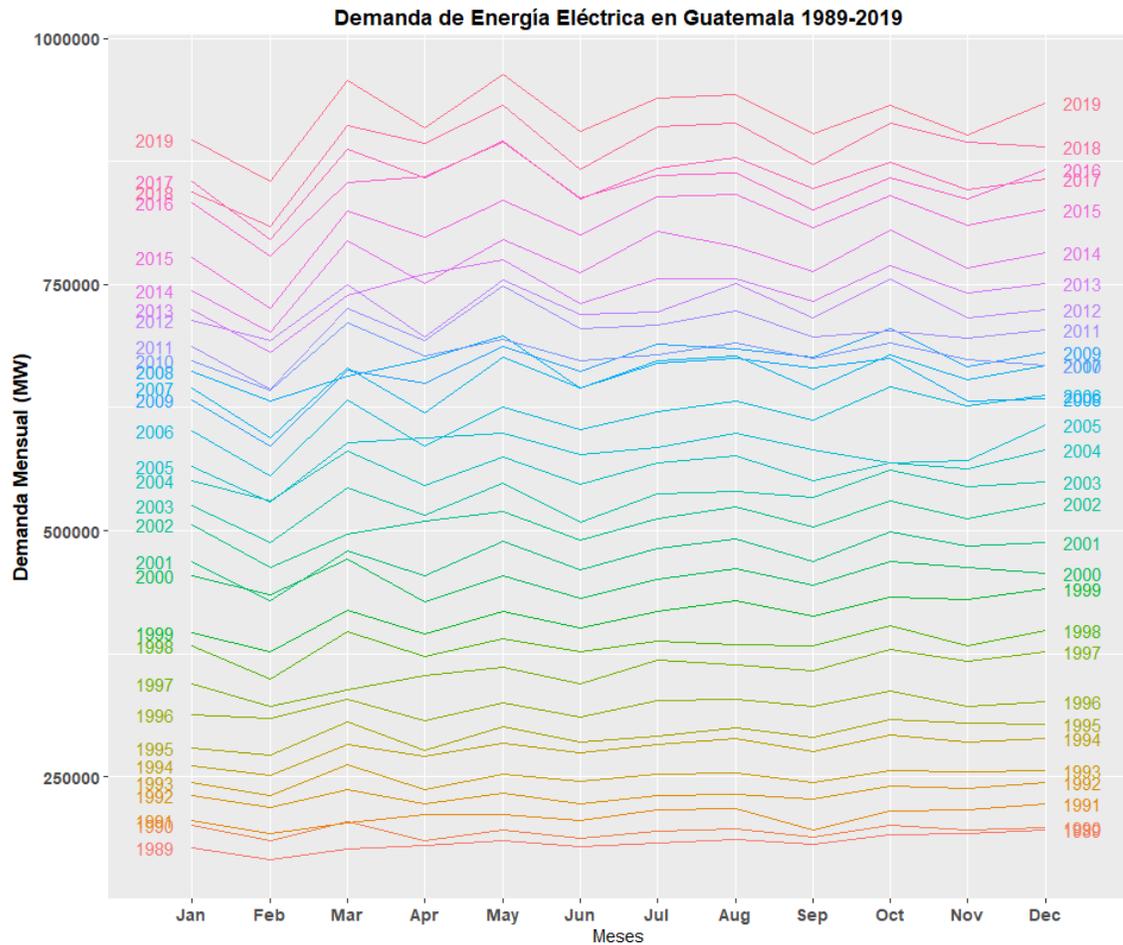


Fuente: elaboración propia.

En la descomposición de la serie de tiempo transformada se observa que guarda las características de la serie original, únicamente que ahora la tendencia lineal se transformó en una tendencia de tipo logarítmica.

De igual manera se utilizó un gráfico por periodo mensual, para identificar si existía estacionalidad en la serie de tiempo.

Figura 18. **Análisis de estacionalidad de la demanda de energía eléctrica en Guatemala 1989-2019**



Fuente: elaboración propia.

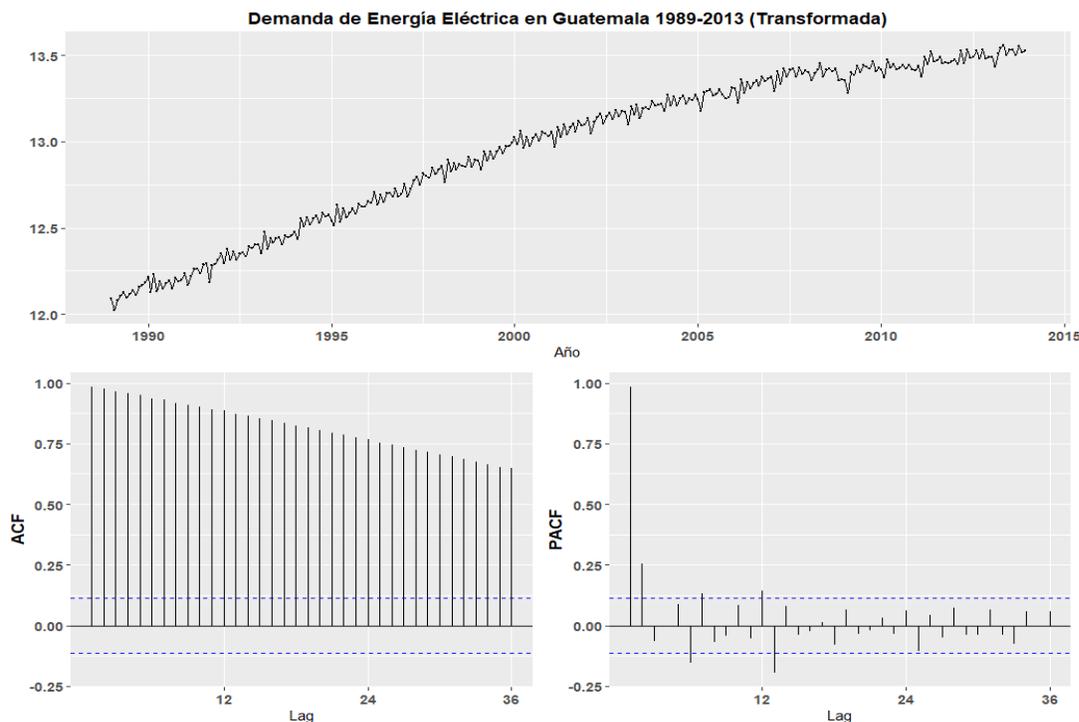
Como se observa en la figura 18 se identifica un patrón de demanda de energía eléctrica mensual que se repite año con año, observamos que febrero es el mes donde se consume menor energía esto ocasionado a que es el mes en donde se presenta menor cantidad de días, además observamos un pico en el consumo de marzo, una variabilidad de altas y bajas para los demás meses, terminando con un diciembre relativamente alto.

Luego de determinar las características de la serie de tiempo en donde se identificó que la serie de tiempo poseía una tendencia y estacionalidad, se procedió a seleccionar distintos modelos de acuerdo con las características presentadas; los datos fueron divididos en 80 % para la construcción y entrenamiento del modelo y 20 % para su validación.

3.2.2. Modelo ARIMA

Para la construcción de un modelo ARIMA es necesario que la serie estudiada sea estacionaria, por lo cual se realizó un análisis de las funciones de autocorrelación simple (FAC) y parcial (FACP) y una prueba de Dickey-Fuller para determinar si la serie era estacionaria o no.

Figura 19. **Gráficos de ACF y PACF de la demanda de energía eléctrica en Guatemala 1989-2013**



Fuente: elaboración propia.

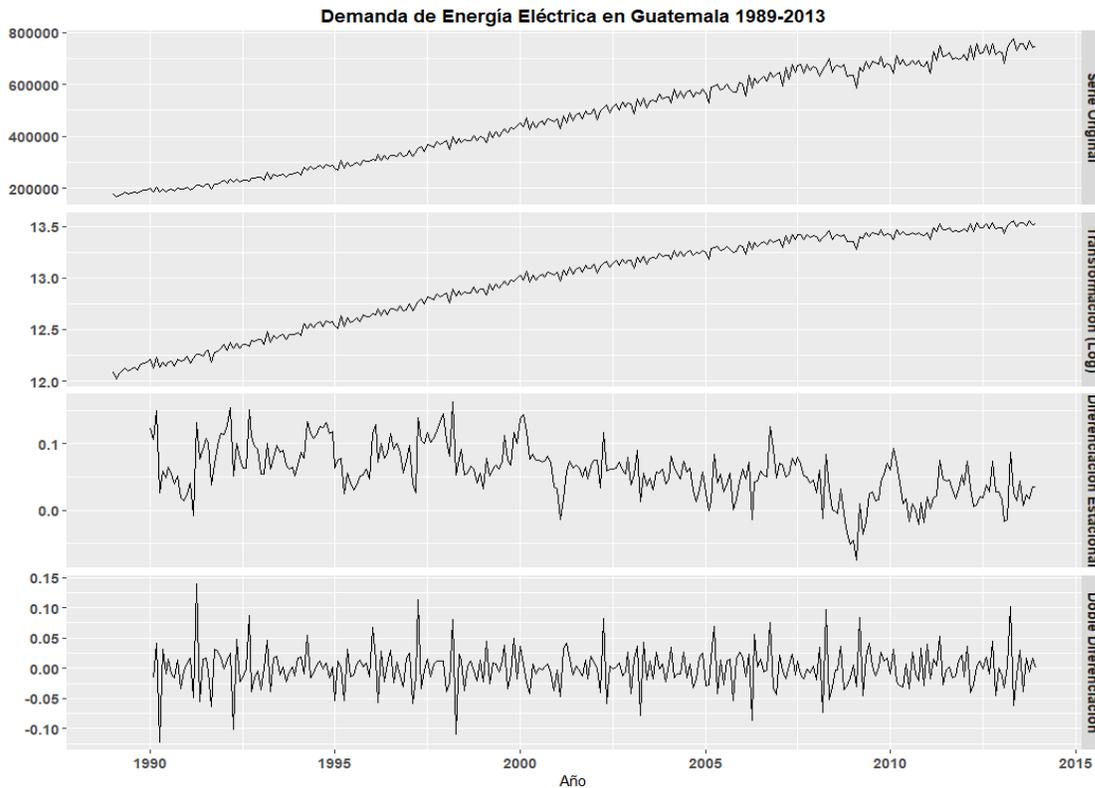
Figura 20. **Prueba de Dickey Fuller**

```
Augmented Dickey-Fuller Test  
data: log.demanda.train  
Dickey-Fuller = -0.3598, Lag order = 6, p-value = 0.9878  
alternative hypothesis: stationary
```

Fuente: elaboración propia.

En la figura 19, se observa para el ACF que los picos son significativos y descienden muy levemente lo que indica autocorrelación en los datos debido a la tendencia, además en el PACF observamos picos significativos en los primeros dos rezagos. Adicional por medio de la prueba de Dickey Fuller se comprueba que la serie no es estacionaria al aceptarse la hipótesis nula de no estacionariedad ($p\text{-valor} > 0.05$). Por lo tanto, se realizó una diferenciación estacional (rezago=12) para eliminar la estacionalidad y una segunda diferenciación para eliminar la tendencia.

Figura 21. **Demanda de energía eléctrica en Guatemala 1989-2013 diferenciada**



Fuente: elaboración propia.

En la comparación de las serie de tiempo (original, transformada, primera diferencia y doble diferencia), se observa como la doble diferenciación logra eliminar la tendencia y estacionalidad de la serie dejando una serie estacionaria cuyas propiedades no dependen del momento en el cual se observa.

Figura 22. **Prueba de Dickey Fuller**

```

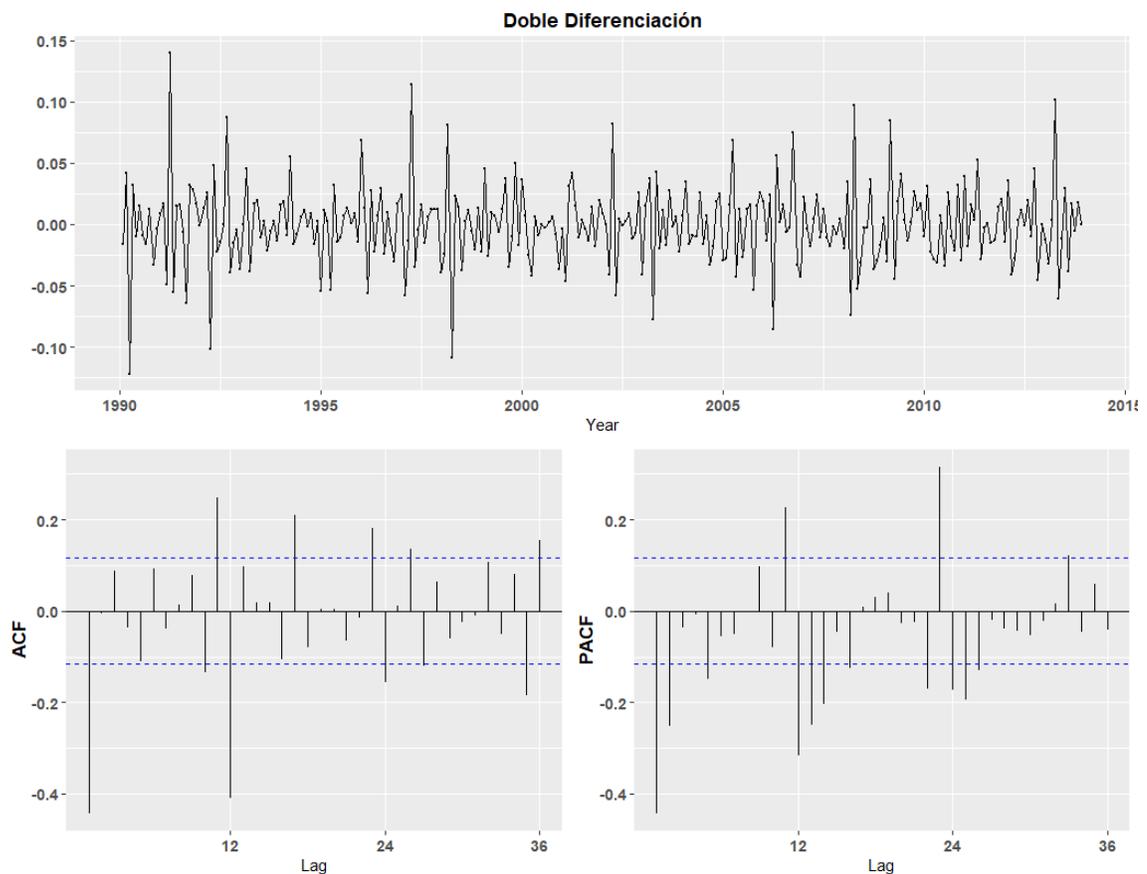
Augmented Dickey-Fuller Test
data: log.demanda.train.diff1
Dickey-Fuller = -8.4747, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
    
```

Fuente: elaboración propia.

Se comprobó estacionariedad de la serie luego de la diferenciación por medio de una prueba de Dickey Fuller con un p-valor <0.05 se rechaza la hipótesis nula de no estacionariedad.

Con la serie ya estacionaria se utilizó un análisis grafico por medio de la autocorrelación simple y parcial, para identificar los modelos tentativos de tipo ARIMA.

Figura 23. **Gráficos de ACF y PACF de la demanda de energía eléctrica en Guatemala 1989-2013 doblemente diferenciada**

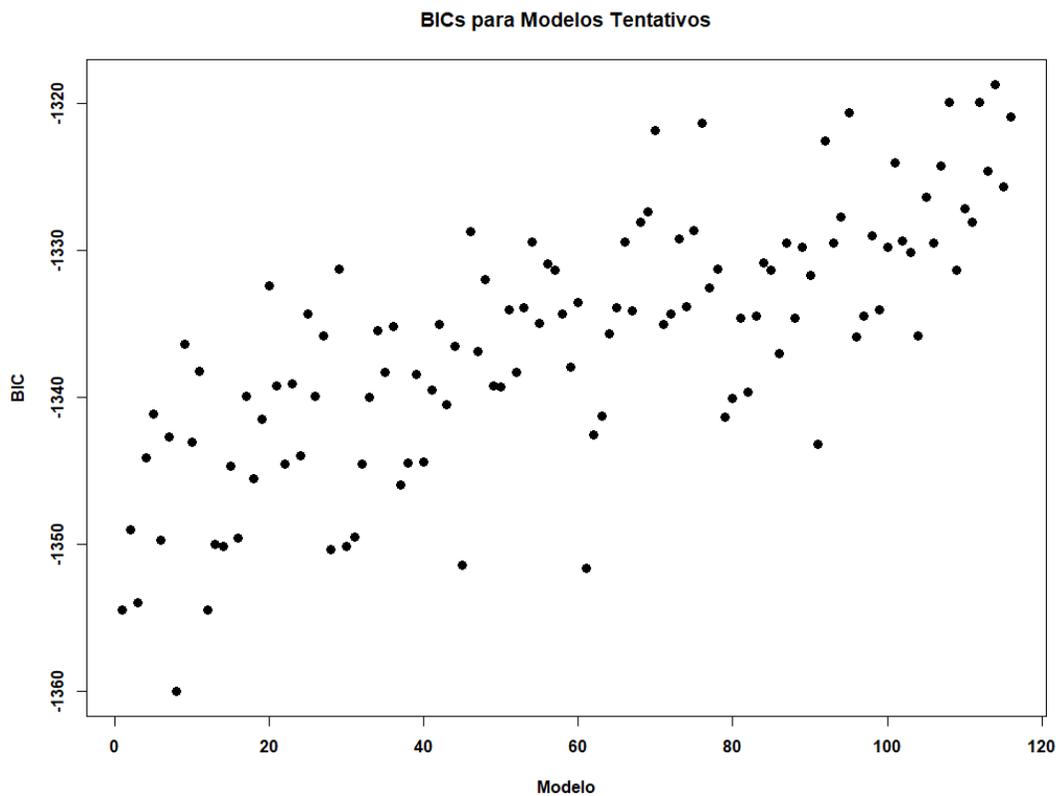


Fuente: elaboración propia.

Como se observa en la figura anterior, para el ACF se presentan correlaciones significativas en el primer rezago y en el rezago 12,24 y 36 lo que nos hace pensar en un MA(1) no estacional y un MA(3) estacional; mientras que para el PACF se presentan dos primeros rezagos significativos, y en los rezagos 12 y 24 por lo se contempla un AR(2) no estacional y un AR(2) estacional.

A partir de la premura anterior se seleccionaron distintos modelos tentativos y se calcularon sus estimadores; estos fueron evaluados por medio de los criterios de información (AIC, BIC) y el error MAPE para seleccionar el mejor.

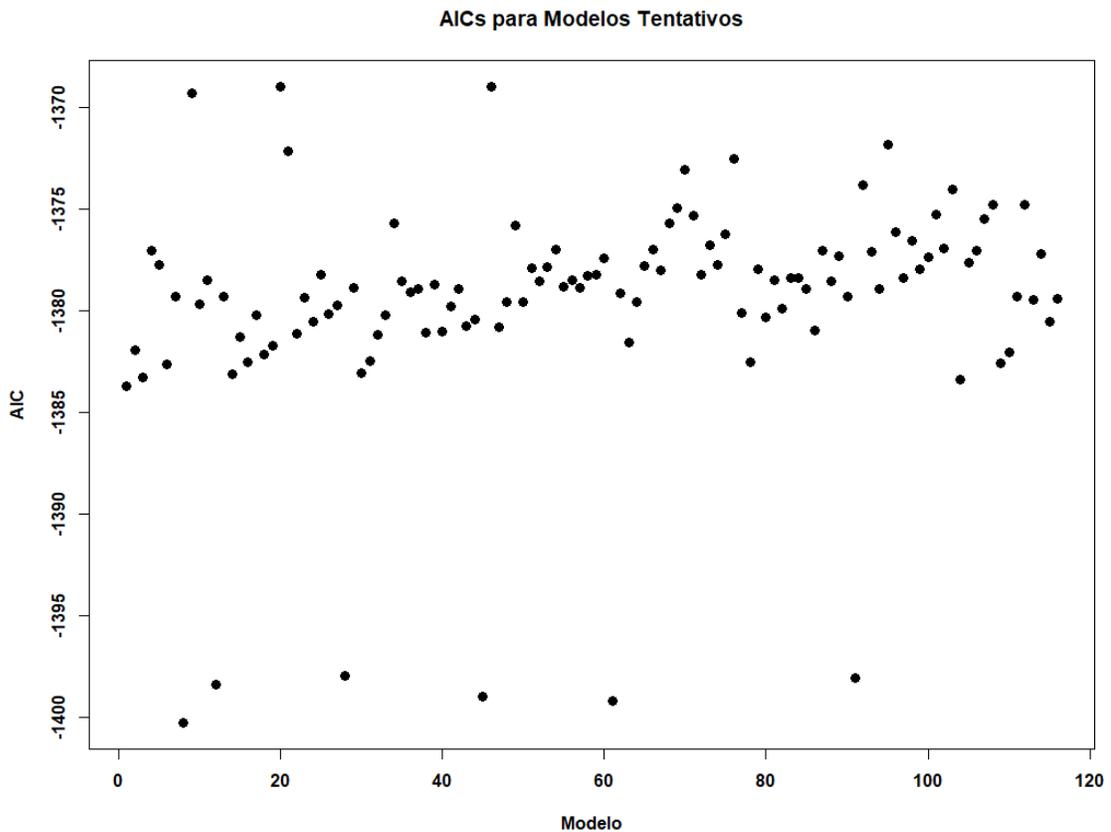
Figura 24. **Criterios de información BIC para modelos tentativos**



Fuente: elaboración propia.

Como se observa en la figura el modelo con el mejor BIC es el modelo [8] ARIMA(2,1,3)(2,1,3)₁₂, seguido por el modelo [12] ARIMA(2,1,3)(3,1,3)₁₂, modelo [1] ARIMA(2,1,3)(0,1,2)₁₂ y el número [3] ARIMA(2,1,3)(1,1,1)₁₂.

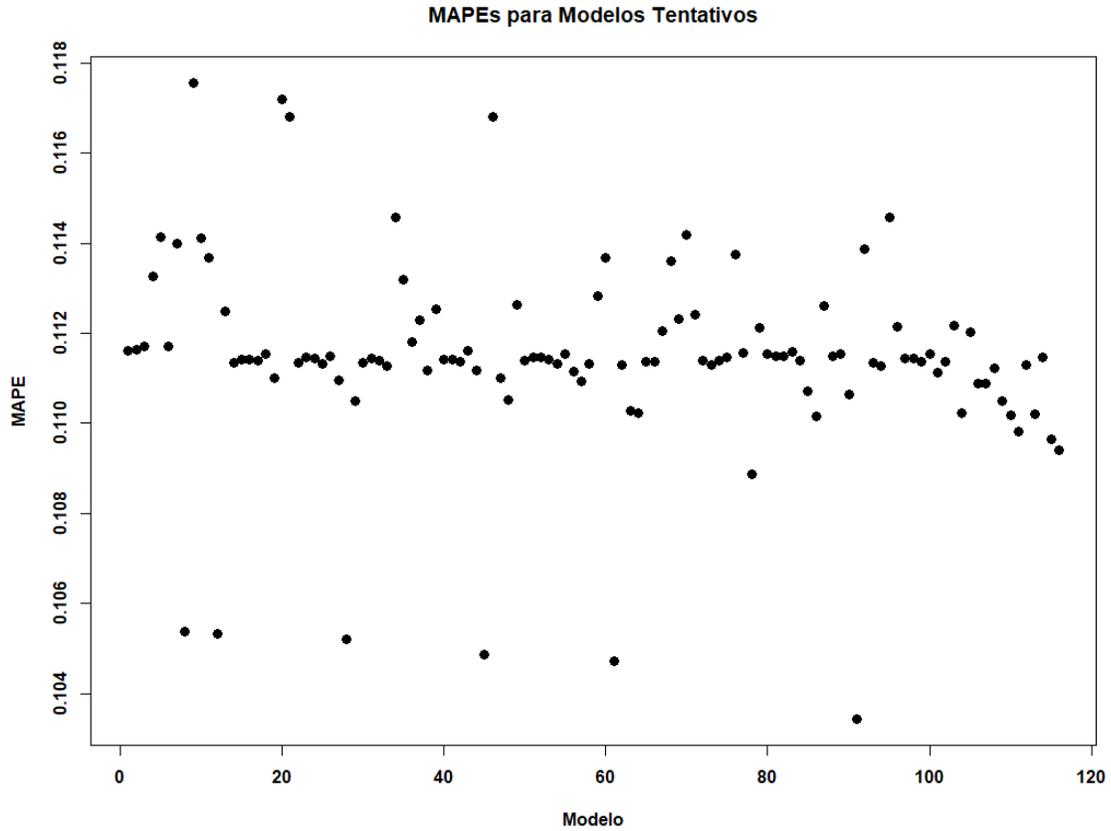
Figura 25. **Criterios de información AIC para modelos tentativos**



Fuente: elaboración propia.

En la figura se observa que los modelos con mejor AIC, son el modelo [8] ARIMA(2,1,3)(2,1,3)₁₂, el modelo [61] ARIMA(4,1,3)(2,1,3)₁₂, el modelo [45] ARIMA(3,1,4)(2,1,3)₁₂, el modelo [12] ARIMA(2,1,3)(3,1,3)₁₂, el modelo [28] ARIMA(2,1,5)(2,1,3)₁₂ y el modelo [91] ARIMA(5,1,4)(2,1,3)₁₂.

Figura 26. MAPE para modelos tentativos



Fuente: elaboración propia.

En la figura anterior se observan los modelos con los mejores MAPE, siendo estos los modelos [91] ARIMA(5,1,4)(2,1,3)₁₂, [61] ARIMA(4,1,3)(2,1,3)₁₂, [45] ARIMA(3,1,4)(2,1,3)₁₂, [28] ARIMA(2,1,5)(2,1,3)₁₂, [12] ARIMA(2,1,3)(3,1,3)₁₂ y [8] ARIMA(2,1,3)(2,1,3)₁₂.

En la tabla siguiente se resumen los mejores modelos con sus resultados.

Tabla V. **Resumen de los mejores modelos**

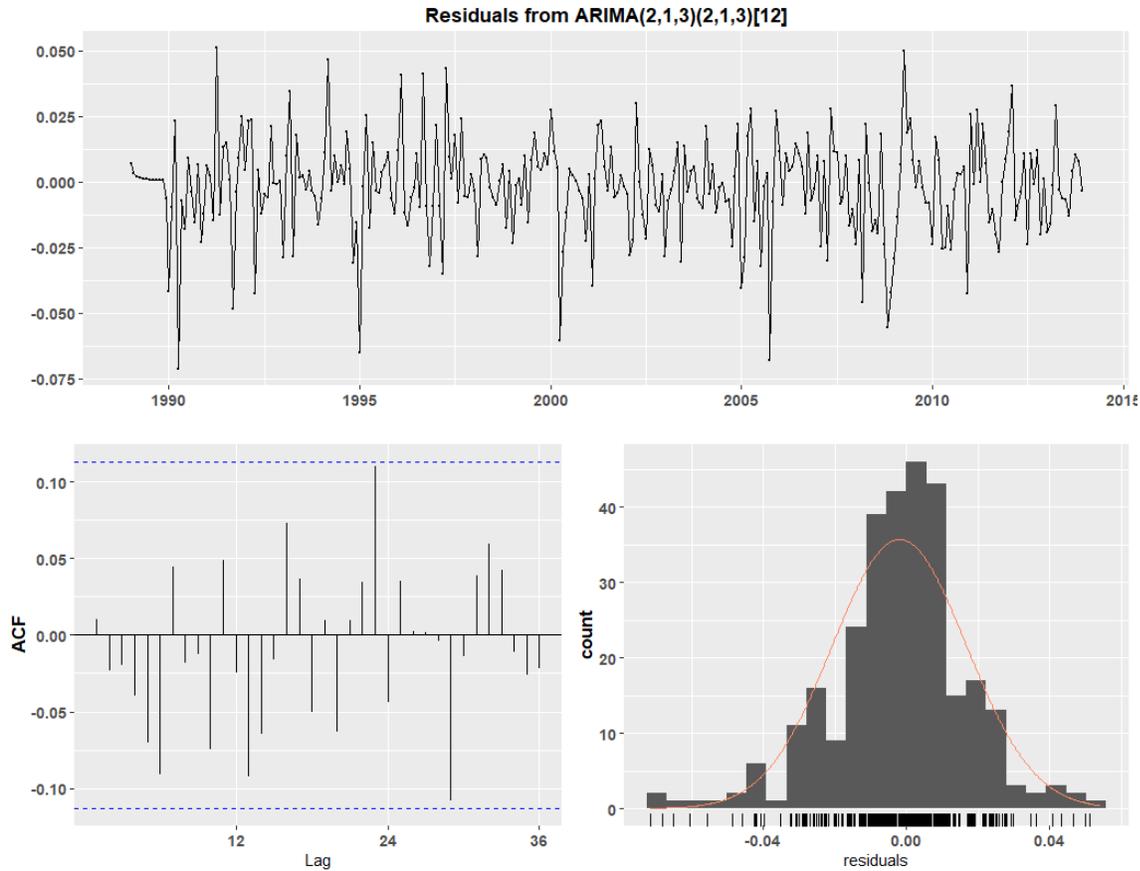
Modelo	BIC	AIC	MAPE
ARIMA(5,1,4)(2,1,3) ₁₂	-1343.16	-1398.05	0.1034
ARIMA(4,1,3)(2,1,3) ₁₂	-1351.61	-1399.18	0.1047
ARIMA(3,1,4)(2,1,3) ₁₂	-1351.4	-1398.98	0.1049
ARIMA(2,1,5)(2,1,3) ₁₂	-1350.38	-1397.95	0.1052
ARIMA(2,1,3)(3,1,3) ₁₂	-1354.48	-1398.39	0.1053
ARIMA(2,1,3)(2,1,3) ₁₂	-1360.03	-1400.28	0.1054

Fuente: elaboración propia.

Como se puede observar el modelo que presenta el mejor AIC y BIC, es el modelo ARIMA(2,1,3)(2,1,3)₁₂, aunque también es el modelo que presenta el MAPE más grande; pero realmente las diferencias de MAPE entre los modelos son mínimas; por lo que se seleccionó como el mejor modelo.

Posteriormente se verificó el modelo, analizando sus residuos, los cuales se esperan se comporten como ruido blanco, aleatorios y no correlacionados; para asegurar que cumpla con las características se realizó un análisis de residuos y la prueba de Box-Ljung.

Figura 27. **Análisis residual del modelo ARIMA**



Fuente: elaboración propia.

Como se observa en el análisis de los residuos, estos se comportan como ruido blanco sin correlaciones entre ellos, en el ACF todos los rezagos se encuentran dentro de los límites sin mostrar correlación significativa; adicional observamos que la distribución sí parece alejarse de la normalidad debido a que presenta un comportamiento más alargado en el centro y valores fuera de los parámetros en los extremos.

Figura 28. **Prueba de Ljung-Box**

```
Ljung-Box test
data: Residuals from ARIMA(2,1,3)(2,1,3)[12]
Q* = 21.736, df = 14, p-value = 0.08421
Model df: 10. Total lags used: 24
```

Fuente: elaboración propia.

La prueba de Ljung-Box nos confirma que los residuos se presentan como ruido blanco con un p-valor > 0.05.

Figura 29. **Prueba de Kolmogorov-Smirnov**

```
Lilliefors (Kolmogorov-Smirnov) normality test
data: modelo8$residuals
D = 0.079805, p-value = 0.00008896
```

Fuente: elaboración propia.

La prueba de Kolmogorov-Smirnov nos confirma que los residuos no presentan una distribución normal con p-valor < 0.05.

3.2.3. Modelo regresión dinámica

Se construyó un modelo de regresión dinámica a partir de las variables que resultaron significativas en el análisis del objetivo 1, en este caso las variables humedad y temperatura.

En la construcción de un modelo de regresión dinámica las series de tiempo deben ser estacionarias, por lo cual se realizaron pruebas de Dickey Fuller para comprobar el comportamiento de ambas variables.

Figura 30. **Prueba de Dickey Fuller (temperatura)**

```
Augmented Dickey-Fuller Test
data: temperatura
Dickey-Fuller = -10.91, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary
```

Fuente: elaboración propia.

Figura 31. **Prueba de Dickey Fuller (Humedad)**

```
Augmented Dickey-Fuller Test
data: humedad
Dickey-Fuller = -6.261, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary
```

Fuente: elaboración propia.

Ambas variables resultaron ser estacionarias con p-valor<0.05.

Figura 32. **Coefficientes de regresión**

```
Series: log.demanda.train
Regression with ARIMA(0,0,0) errors

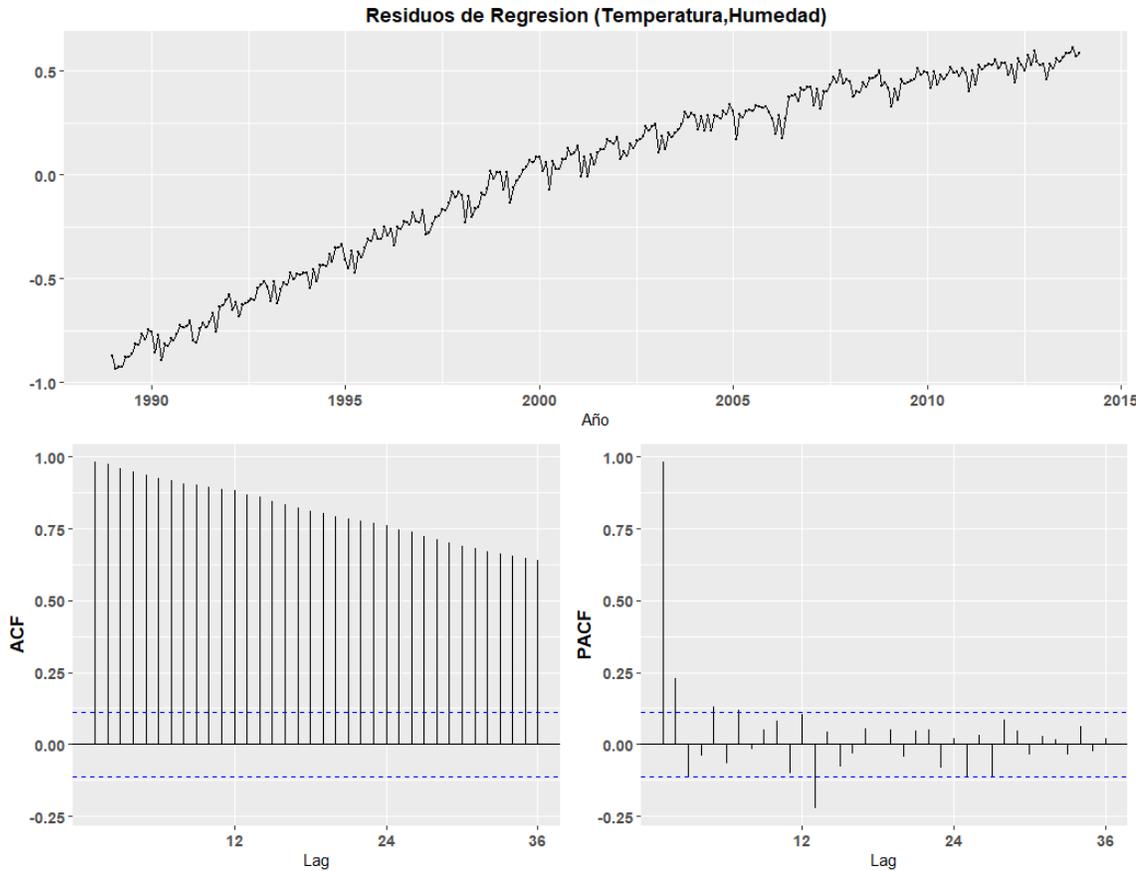
Coefficients:
    intercept  Temperatura  Humedad
    12.8955      0.0209     -0.0054
s.e.          0.5875      0.0213     0.0049

sigma^2 estimated as 0.2006:  log likelihood=-183.23
AIC=374.47  AICC=374.6  BIC=389.28
```

Fuente: elaboración propia.

En la figura anterior se muestran los coeficientes del modelo de regresión, este modelo fue construido inicialmente para estudiar el comportamiento de sus residuos e identificar qué tipo de modelo ARIMA se acoplaba a sus características.

Figura 33. Residuos de regresión, ACF y PACF



Fuente: elaboración propia.

Figura 34. Prueba de Dickey Fuller

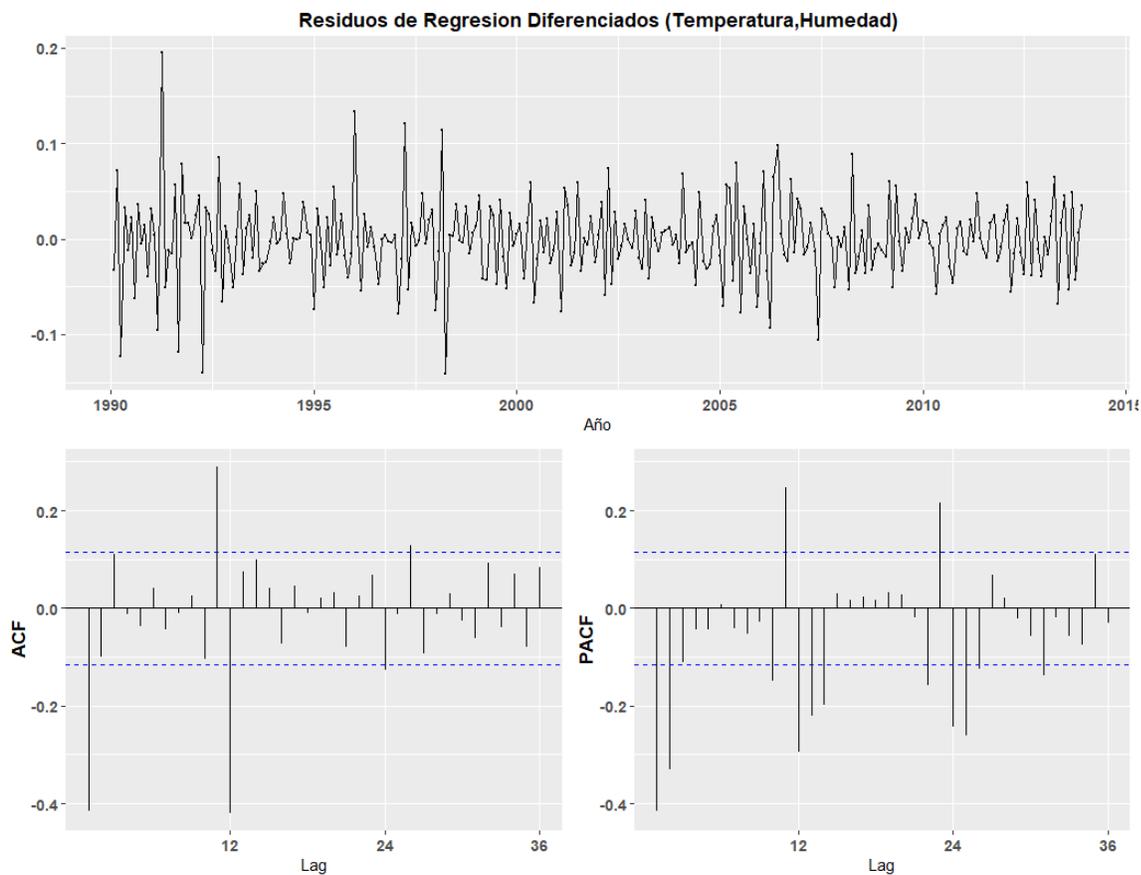
```
Augmented Dickey-Fuller Test
data: a3$residuals
Dickey-Fuller = -1.6548, Lag order = 6, p-value = 0.7218
alternative hypothesis: stationary
```

Fuente: elaboración propia.

En la figura 33, se observa que para el ACF de los residuos los picos son significativos y descienden muy levemente lo que indica autocorrelación en los datos debido a cierta tendencia, además en el PACF observamos picos

significativos en los primeros dos rezagos, lo cual indica que la serie no es estacionaria. Adicional por medio de la prueba de Dickey Fuller se comprueba que la serie no es estacionaria al aceptarse la hipótesis nula de no estacionariedad ($p\text{-valor} > 0.05$). Por lo tanto, se realizó una diferenciación estacional (rezago=12) para eliminar la estacionalidad y una segunda diferenciación para eliminar la tendencia características identificadas con el análisis previ6.

Figura 35. Residuos de regresión diferenciados, ACF y PACF



Fuente: elaboraci6n propia.

En la figura anterior, para el ACF se observan correlaciones significativas en el primer rezago y en el rezago 12 y 24 lo que nos hace pensar en un MA(1) no estacional y un MA(2) estacional; mientras que el PACF presenta dos primeros rezagos significativos, y en los rezagos 12 y 24 por lo se contempla un AR(2) no estacional y un AR(2) estacional; como se observan ambos diagramas tienen mucha similitud a los presentados en el análisis del modelo ARIMA.

Figura 36. **Prueba de Dickey Fuller**

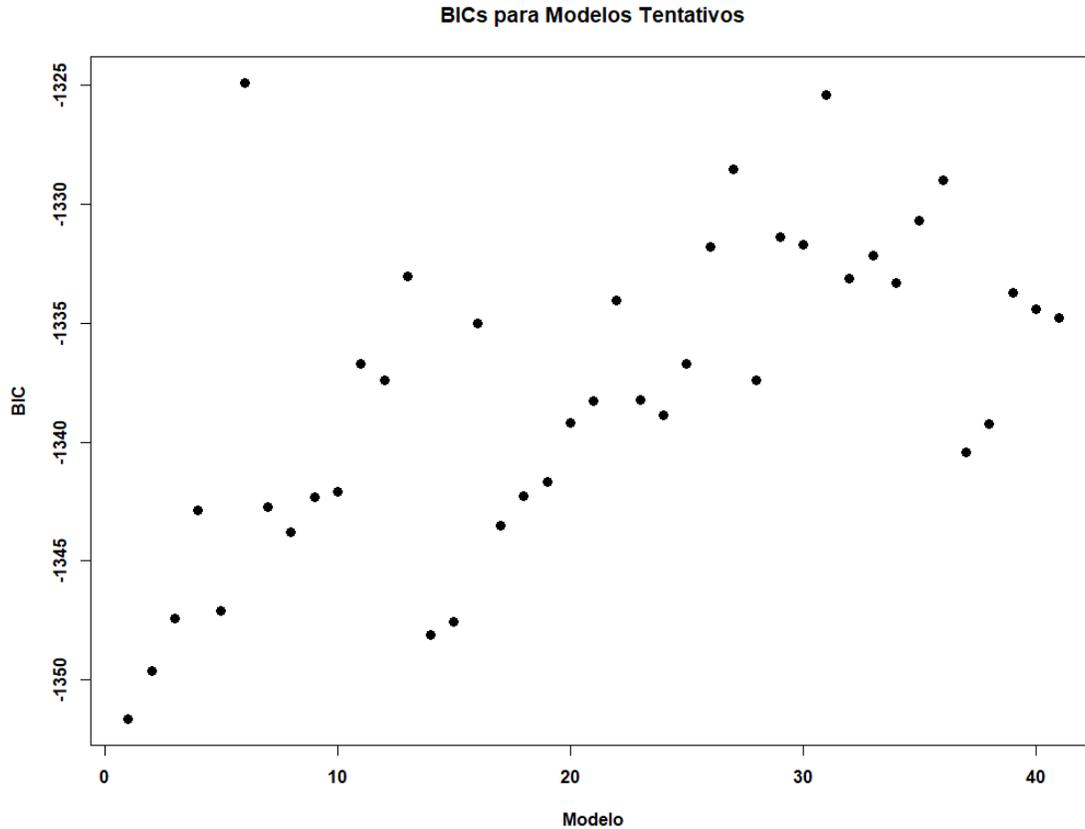
```
Augmented Dickey-Fuller Test
data: diff(diff(a3$residuals, 12))
Dickey-Fuller = -7.903, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Fuente: elaboración propia.

Se comprobó la estacionariedad de la serie luego de la diferenciación por medio de una prueba de Dickey Fuller con un p-valor < 0.05 se rechaza la hipótesis nula de no estacionariedad.

A partir del análisis anterior se seleccionaron una serie de posibles modelos, los cuales fueron evaluados por medio de los criterios de información (AIC, BIC) y el error MAPE para seleccionar el mejor.

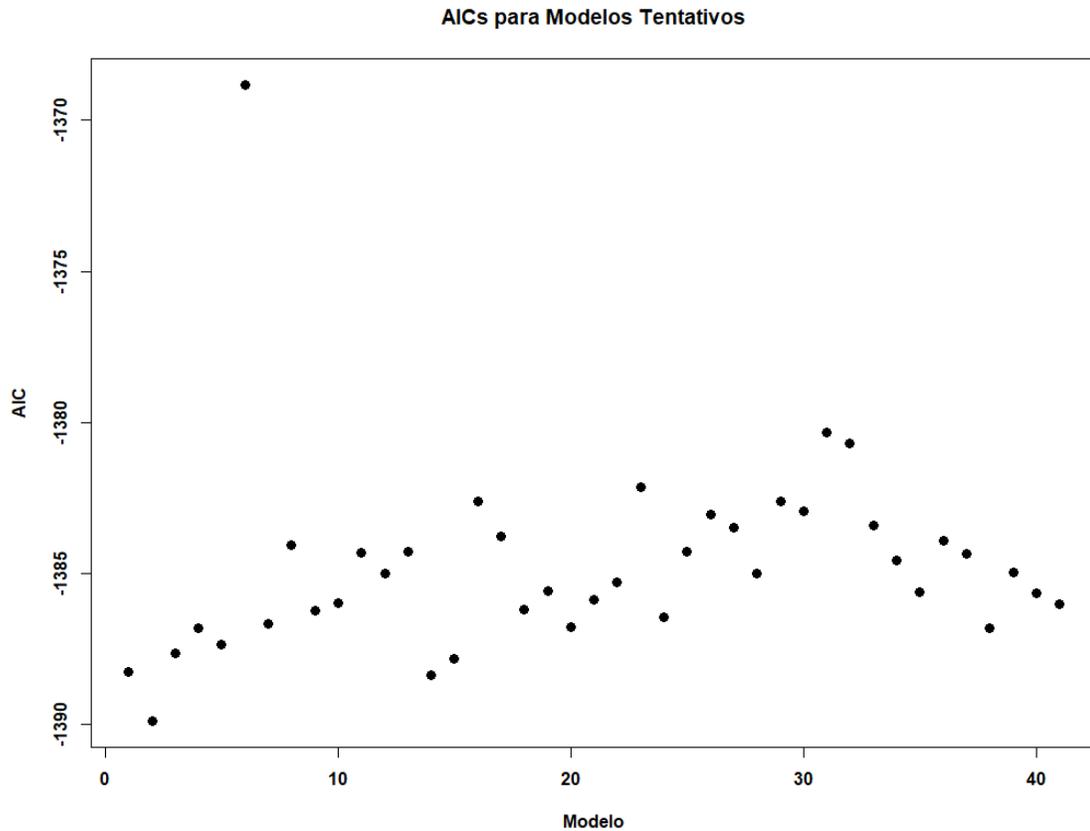
Figura 37. Criterios de información BIC para modelos tentativos



Fuente: elaboración propia.

Como se observa en la figura anterior el modelo con el mejor BIC es el modelo [1] de regresión con errores $ARIMA(2,1,3)(1,1,1)_{12}$, seguido por el modelo [2] de regresión con errores $ARIMA(2,1,3)(1,1,2)_{12}$, el modelo [14] de regresión con errores $ARIMA(3,1,3)(0,1,2)_{12}$, el modelo [15] de regresión con errores $ARIMA(3,1,3)(1,1,1)_{12}$ y el modelo [3] de regresión con errores $ARIMA(2,1,3)(2,1,1)_{12}$.

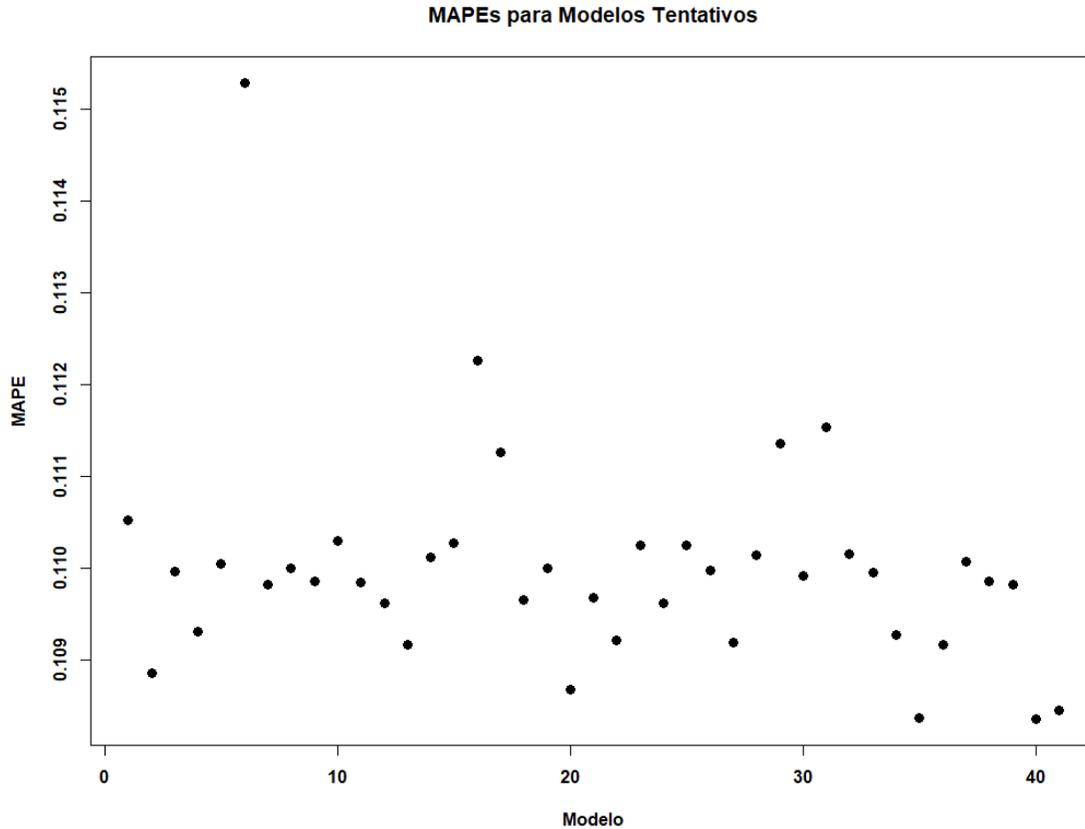
Figura 38. **Criterios de información AIC para modelos tentativos**



Fuente: elaboración propia.

En la figura anterior se observa que los modelos con mejor AIC son el modelo [2] de regresión con errores $ARIMA(2,1,3)(1,1,2)_{12}$, el modelo [14] de regresión con errores $ARIMA(3,1,3)(0,1,2)_{12}$, el modelo [1] de regresión con errores $ARIMA(2,1,3)(1,1,1)_{12}$, el modelo [15] de regresión con errores $ARIMA(3,1,3)(1,1,1)_{12}$ y el modelo [3] de regresión con errores $ARIMA(2,1,3)(2,1,1)_{12}$.

Figura 39. MAPE para modelos tentativos



Fuente: elaboración propia.

En la figura anterior se presentan los errores MAPE para los modelos tentativos, como se observa los modelos que presentaron los mejores más bajos fueron el modelo [40] de regresión con errores $ARIMA(5,1,4)(0,1,2)_{12}$, el modelo [35] de regresión con errores $ARIMA(4,1,5)(1,1,2)_{12}$, el modelo [41] de regresión con errores $ARIMA(5,1,4)(1,1,1)_{12}$, el modelo [20] de regresión con errores $ARIMA(3,1,4)(1,1,2)_{12}$ y el modelo [2] de regresión con errores $ARIMA(2,1,3)(1,1,2)_{12}$.

En la tabla siguiente se resumen los mejores modelos con sus resultados.

Tabla VI. **Resumen de los mejores modelos**

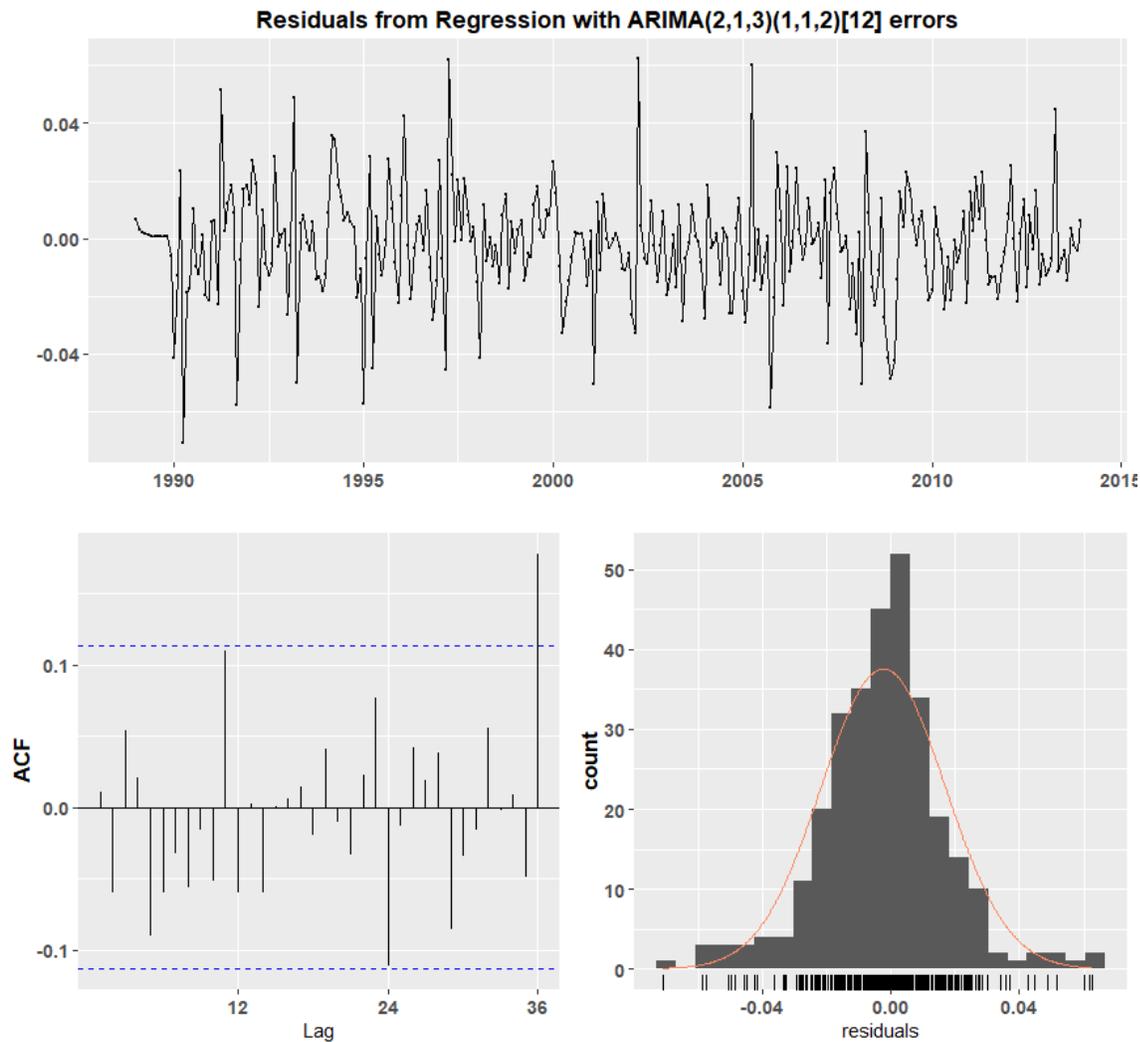
Modelo	BIC	AIC	MAPE
Regresión con ARIMA(2,1,3)(1,1,1) ₁₂	-1351.65	-1388.24	0.1105
Regresión con ARIMA(2,1,3)(1,1,2) ₁₂	-1349.64	-1389.89	0.1088
Regresión con ARIMA(3,1,3)(0,1,2) ₁₂	-1348.12	-1388.37	0.1101
Regresión con ARIMA(3,1,3)(1,1,1) ₁₂	-1347.55	-1387.80	0.1102
Regresión con ARIMA(2,1,3)(2,1,1) ₁₂	-1347.39	-1387.65	0.1099

Fuente: elaboración propia.

Como se puede observar en la tabla VI, el modelo con el menor AIC y MAPE es el de regresión con ARIMA(2,1,3)(1,1,2)₁₂ mientras que este mismo modelo presenta el segundo mejor BIC, por lo cual se seleccionó como el mejor modelo de regresión dinámica.

Posteriormente se verificó el modelo, analizando sus residuos, los cuales se esperan se comporten como ruido blanco, aleatorios y no correlacionados; para asegurar que cumpla con las características se realizó un análisis de residuos y la prueba de Box-Ljung.

Figura 40. **Análisis residual del modelo de regresión dinámica**



Fuente: elaboración propia.

Como se muestra en la figura anterior los residuos parecieran comportarse como ruido blanco sin correlaciones entre ellos debido a que en el ACF todos los rezagos se encuentran dentro de los límites sin mostrar correlación significativa; de igual manera observamos que la distribución de los residuos parece alejarse de la normalidad ya que presenta una gran concentración de datos en el centro del gráfico y ciertos valores en los extremos.

Figura 41. **Prueba de Ljung-Box**

```
Ljung-Box test
data: Residuals from Regression with ARIMA(2,1,3)(1,1,2)[12] errors
Q* = 21.323, df = 14, p-value = 0.09363
Model df: 10. Total lags used: 24
```

Fuente: elaboración propia.

La prueba de Ljung-Box nos confirma que los residuos se presentan como ruido blanco con un p-valor > 0.05.

Figura 42. **Prueba de Kolmogorov-Smirnov**

```
Lilliefors (Kolmogorov-Smirnov) normality test
data: modelo2$residuals
D = 0.056144, p-value = 0.02342
```

Fuente: elaboración propia.

La prueba de Kolmogorov-Smirnov nos confirma que los residuos no presentan una distribución normal con p-valor < 0.05.

3.2.4. Redes neuronales

Se construyó un modelo de redes neuronales a partir de las variables que resultaron significativas en el análisis del objetivo 1, en este caso las variables humedad y temperatura, además se utilizaron distintos valores rezagados de la serie de tiempo; el modelo de redes neuronales contempló una capa oculta y el número de neuronas se escogió como la media de las neuronas en la capa de entrada y salida, para todos los modelos evaluados se utilizaron 1000 repeticiones para encontrar el mejor ajuste.

Se evaluaron distintos modelos variando la cantidad de valores rezagados en la capa de entrada, y la cantidad de neuronas en la capa oculta, los resultados se presentan a continuación.

Tabla VII. **Modelos de redes neuronales tentativos**

Cant. Rezagos No Estacionales	Cant. Rezados Estacionales	Número de Neuronas	MAPE	RMSE
1	1	2	0.1695	0.0284
1	2	2	0.1509	0.0249
1	3	3	0.1372	0.0225
1	4	3	0.1335	0.0220
3	1	3	0.1466	0.0241
3	2	3	0.1328	0.0172
3	3	4	0.1208	0.0201
3	4	4	0.1180	0.0197
5	1	4	0.1380	0.0229
5	2	4	0.1233	0.0208
5	3	5	0.1096	0.0185
5	4	5	0.1060	0.0180
7	1	5	0.1274	0.0211
7	2	5	0.1138	0.0190
7	3	6	0.1030	0.0172
7	4	6	0.0958	0.0160
9	1	6	0.1160	0.0194
9	2	6	0.1057	0.0177
9	3	7	0.0961	0.0159
9	4	7	0.0864	0.0146
15	4	10	0.0626	0.0107

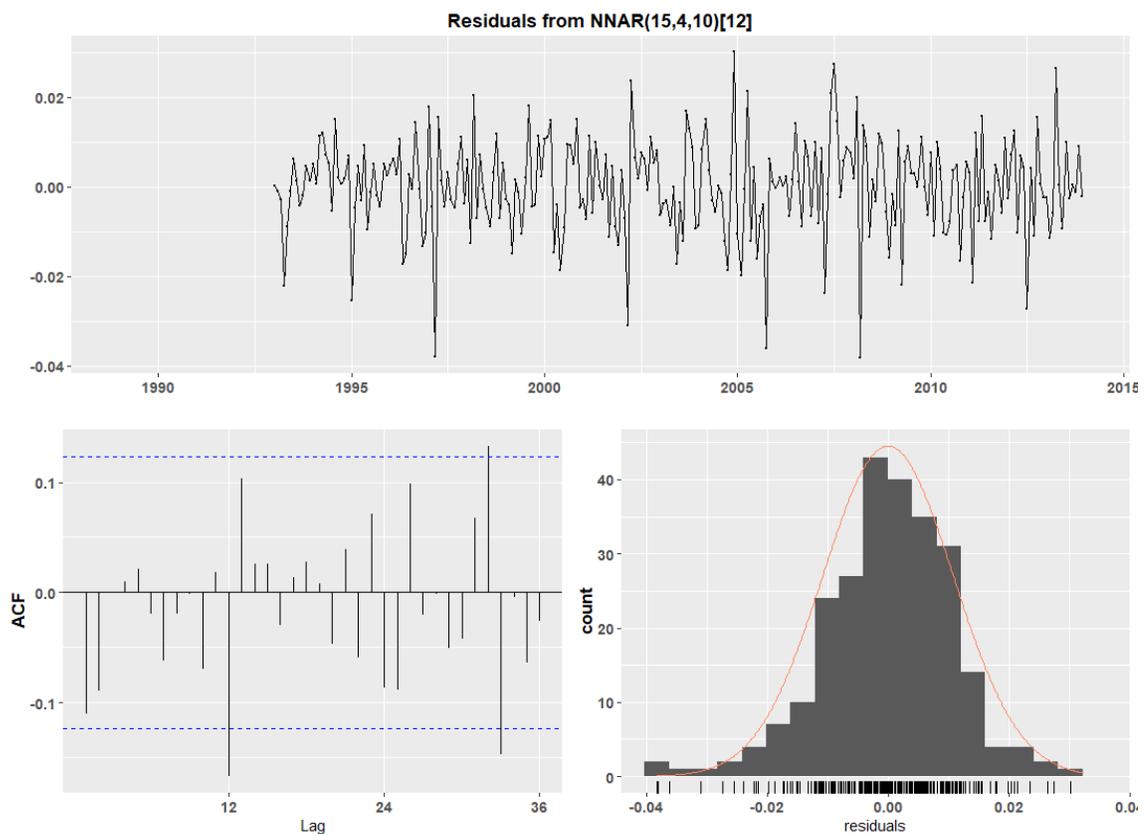
Fuente: elaboración propia.

Como se aprecia en la tabla anterior la incorporación de más valores rezagados disminuye el error, pero en conjunto con la cantidad de neuronas en la capa de entrada, se deben ir aumentando el número de neuronas en la capa oculta; en este caso se escogió como modelo uno con 15 valores rezagados no

estacionales y 4 valores rezagados estacionales y una capa oculta con 10 neuronas, debido a que fue el que presento menor error de MAPE y RMSE.

A continuación, se presenta el análisis de residuos para el modelo.

Figura 43. **Análisis residual del modelo de redes neuronales**



Fuente: elaboración propia.

En la figura anterior se observa que los errores se comportan como ruido blanco sin correlaciones entre ellos, aunque en el ACF se presentan tres rezagos con correlación significativa, esto se mantiene entre el parámetro aceptable del 5% de rezagos que pueden ser significativos; en la distribución de los residuos en comparación con la campana normal se aprecian algunos valores en los

extremos, aun así, los datos se distribuyen la mayoría por debajo de la curva normal.

Figura 44. **Prueba de Ljung-Box**

```
Box-Ljung test
data: fit2$residuals
X-squared = 3.1222, df = 1, p-value = 0.07723
```

Fuente: elaboración propia.

La prueba de Ljung-Box nos confirma que los residuos se presentan como ruido blanco con un $p\text{-valor} > 0.05$.

Figura 45. **Prueba de Kolmogorov-Smirnov**

```
Lilliefors (kolmogorov-smirnov) normality test
data: fit2$residuals
D = 0.054871, p-value = 0.0636
```

La prueba de Kolmogorov-Smirnov nos confirma que los residuos se distribuyen de manera normal con $p\text{-valor} > 0.05$.

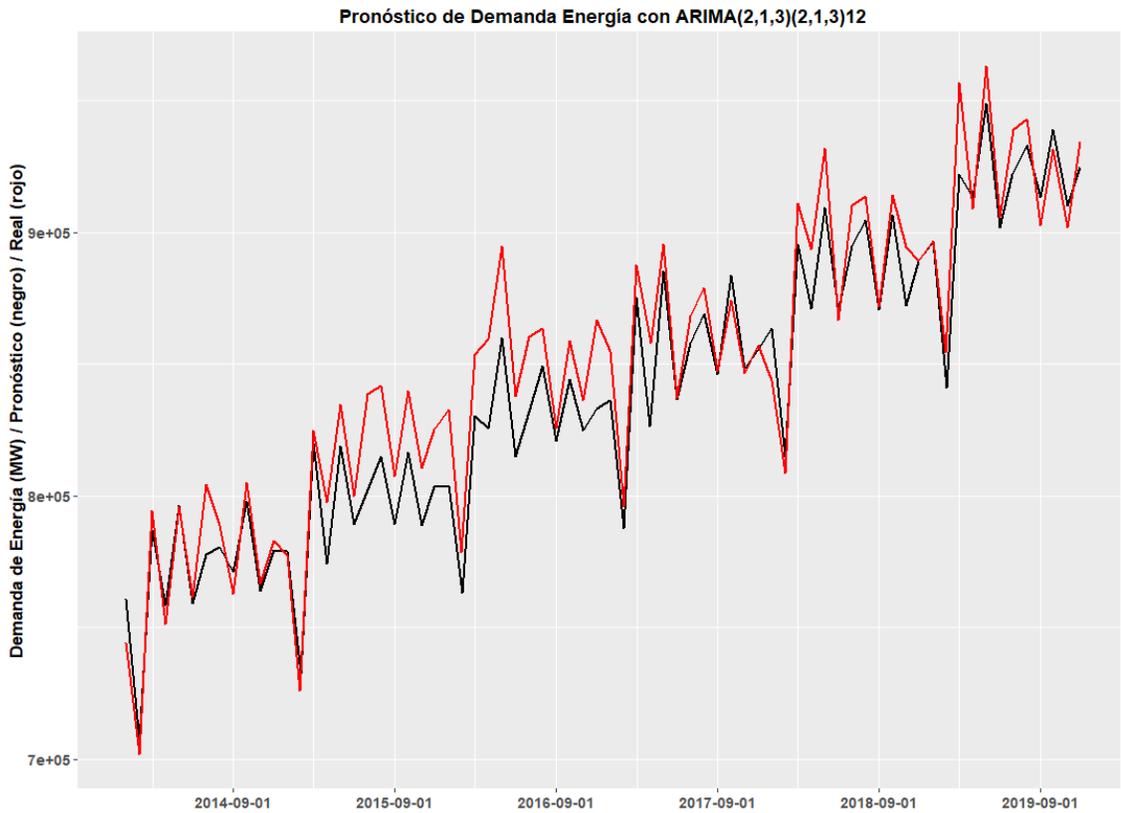
3.3. **Objetivo 3. Evaluar la validez y confiabilidad del modelo**

Para evaluar la validez y confiabilidad del modelo se utilizaron los datos de prueba los cuales representaban el 20 % de los datos totales, los resultados se presentan a continuación.

3.3.1. **ARIMA**

Es un modelo estadístico que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro.

Figura 46. **Pronóstico de la demanda de energía eléctrica mensual con ARIMA(2,1,3)(2,1,3)₁₂**



Fuente: elaboración propia.

Tabla VIII. **Evaluación de modelo ARIMA(2,1,3)(2,1,3)₁₂**

Modelo	RMSE	MAPE
ARIMA(2,1,3)(2,1,3) ₁₂	16674.62	1.57

Fuente: elaboración propia.

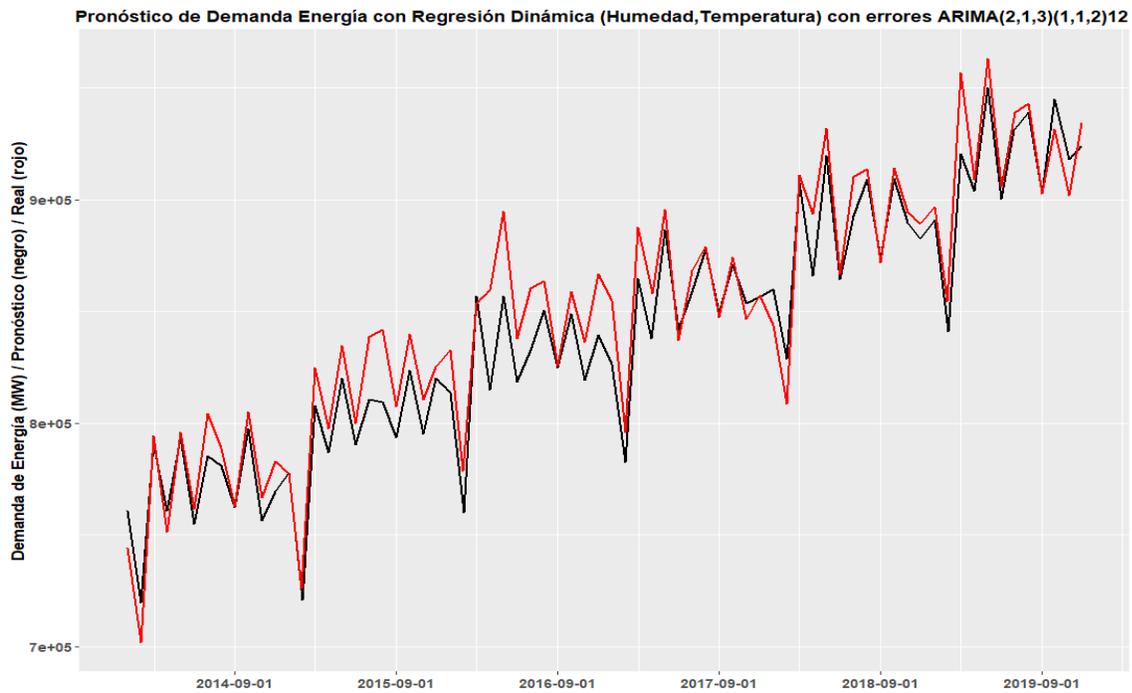
Como se puede apreciar en la figura anterior el pronóstico (negro) distingue muy bien la tendencia y estacionalidad de la serie de tiempo logrando producir un excelente pronóstico; únicamente se observa que para el periodo 2015-2016 el pronóstico se quedó corto en los picos los cuales fueron más altos

de lo esperado; el error MAPE es del 1.57 % y el RMSE de 16674.62 MW los cuales son valores muy bajos y confiables.

3.3.2. Regresión dinámica

En la siguiente figura se presenta el pronóstico de la demanda de energía eléctrica mensual con regresión dinámica.

Figura 47. **Pronóstico de la demanda de energía eléctrica mensual con regresión dinámica (2,1,3)(1,1,2)₁₂**



Fuente: elaboración propia.

Tabla IX. **Evaluación de modelo regresión con ARIMA(2,1,3)(2,1,1) 12**

Modelo	RMSE	MAPE
Regresión con ARIMA(2,1,3)(2,1,1) 12	15816.35	1.47

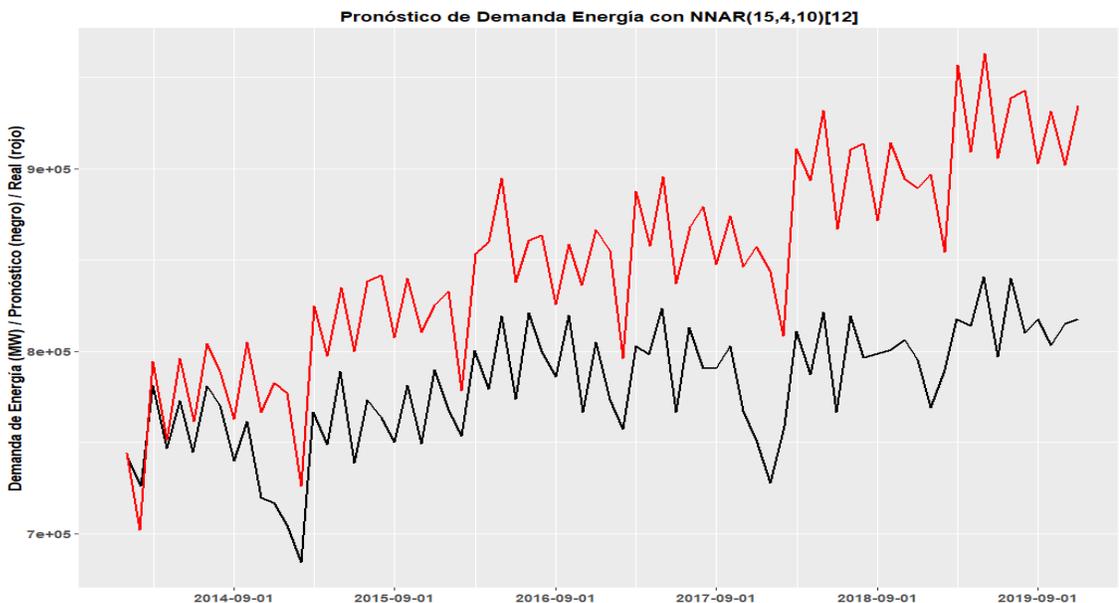
Fuente: elaboración propia.

En la figura anterior se observa el comparativo del pronóstico (negro) *versus* el valor real (rojo) para el periodo 2014-2019, se observa que el pronóstico identifica muy bien la estacionalidad de la serie de tiempo y la tendencia, mostrando un comportamiento muy cercano al real; el MAPE es del 1.47 %, con un RMSE de 15816.35 MW, lo cual nos indica que el pronóstico es muy bueno.

3.3.3. Redes neuronales

En la siguiente figura se puede observar este modelo.

Figura 48. **Pronóstico de demanda de energía eléctrica mensual con NNAR(15,4,10)**



Fuente: elaboración propia.

Tabla X. **Evaluación de modelo NNAR(15,4,10)₁₂**

Modelo	RMSE	MAPE
NNAR(15,4,10) ₁₂	76678.25	7.98

Fuente: elaboración propia.

Como se observa en el comparativo del pronóstico (negro) *versus* el valor real (rojo) del periodo 2014-2019 para el modelo de redes neuronales, el modelo logró identificar muy bien la tendencia y estacionalidad de la serie para los primeros tres años utilizando la información histórica, a partir del cuarto año con la utilización de la información pronosticada, observamos que el modelo pierde exactitud. A pesar de eso su MAPE es de 7.98 % y su RMSE 76678.25 MW, valores relativamente aceptables.

3.4. General. Determinar el modelo más confiable de pronóstico para estimar la demanda de energía eléctrica en Guatemala, entre distintos modelos de series de tiempo y uno de redes neuronales

En la tabla anterior, se observan los resultados obtenidos en la validez de los modelos tentativos utilizando los datos de prueba; de la comparación se identificó que el modelo más confiable es el de regresión con errores ARIMA(2,1,3)(2,1,1)₁₂ con un MAPE de 1.47 % , seguido por el modelo ARIMA(2,1,3)(2,1,3)₁₂ con un MAPE de 1.57 % y el de redes neuronales con 7.98 %.

Tabla XI. **Comparación de modelos**

Modelo	RMSE	MAPE
ARIMA(2,1,3)(2,1,3) ₁₂	16674.62	1.57
Regresión con ARIMA(2,1,3)(2,1,1) ₁₂	15816.35	1.47
NNAR(15,4,10) ₁₂	76678.25	7.98

Fuente: elaboración propia.

4. DISCUSIÓN DE RESULTADOS

Análisis interno

El objetivo de este estudio era determinar el modelo más confiable de pronóstico para estimar la demanda de energía eléctrica en Guatemala, comparando modelos tradicionales de series de tiempo con modelos más sofisticados como lo son las redes neuronales; para esto se planteó identificar las variables que influyen en el comportamiento de la demanda de energía, construir un modelo de predicción con el mejor ajuste y evaluar su validez y confiabilidad.

La selección de las variables para el estudio se realizó a partir de la revisión de antecedentes, se tomaron en cuenta las variables: cantidad de población, PIB, mes del año, humedad y temperatura; la influencia de las variables propuestas en la demanda de energía eléctrica se determinó por medio de un análisis correlacional para las variables cuantitativas utilizando el coeficiente de Spearman y por medio de un análisis con Kruskal-Wallis para la variable categórica (mes del año), además fue apoyado por medio de un análisis gráfico utilizando diagramas de dispersión para identificar los tipos de relaciones entre las variables y un *boxplot* para medir las diferencias entre las medias por mes del año.

Inicialmente se había determinado una relación lineal positiva muy fuerte entre la demanda de energía eléctrica y las variables de población y PIB, mientras que una relación muy débil para la variable temperatura (figura 7) y una nula según la prueba de correlación para la variable humedad (tabla II); es muy común que en este tipo de análisis para las series de tiempo existan relaciones espurias,

es decir, relaciones entre variables que no guardan una conexión lógica entre ellas y únicamente se relacionan debido a que ambas presentan una tendencia, para poder excluir este tipo de resultados se realizó un análisis correlacional con las series diferenciadas eliminando la tendencia que estas pudieran presentar, las pruebas de correlación mostradas en tabla III, muestran que la diferenciación elimino las fuertes correlaciones existentes entre la demanda de energía eléctrica y las variables de PIB y población, resultando no poseer una correlación significativa con p-valor mayor al 0.05; mientras que las variables humedad y temperatura mostraron tener una correlación significativa pero débil (-0.13 y 0.13 respectivamente).

Por otro lado, por medio del análisis de Kruskal-Wallis (figura 11), se determinó que el factor mes no influía en la demanda de energía eléctrica, esto significa que no existía diferencia significativa entre los promedios de consumo para cada mes del año, esto calculado con una significancia del 5 %.

Para la construcción del modelo de predicción, se realizó un análisis gráfico y descriptivo de la demanda de energía eléctrica con el fin de identificar sus componentes (sección 3.2.1). A partir de este análisis se determinó que la serie de tiempo poseía una tendencia creciente (figura 17) con estacionalidad anual (figura 18); otro aspecto identificado en el análisis de la serie de tiempo fue la falta de varianza constante, debido a que existe un aumento en la variación anual de los datos como se aprecia en la figura 13, por lo cual se necesitó realizar una transformación logarítmica (figura 15).

La identificación de las características y variables influyentes determinó la utilización de distintos modelos tentativos, siendo estos un ARIMA estacional, un modelo de regresión dinámica y uno de redes neuronales. El primer modelo se plantea como un $ARIMA(2,1,3)(2,1,3)_{12}$, este es un modelo de doble

diferenciación con orden 2 autorregresivo y orden 3 de media móvil tanto para la parte estacional como para la parte no estacional; este modelo presentó un BIC de -1360.03, un AIC de -1400.28 y un MAPE de 0.1054 % (tabla V).

El segundo modelo se plantea como un modelo de regresión con errores ARIMA(2,1,3)(1,1,2)₁₂, este es un modelo generado a partir de las variables influyentes humedad y temperatura, con presencia en los errores de un modelo de doble diferenciación con orden 2 autorregresivo y orden 3 de media móvil para la parte no estacional, y orden 1 autorregresivo y orden 2 de media móvil para la parte estacional, el modelo presentó un BIC de -1349.64, AIC de -1389.89 y un MAPE de 0.1088 % (tabla VI).

El tercer modelo se plantea como un modelo de redes neuronales con una capa oculta de 10 neuronas, utilizando como valores de entrada las variables influyentes humedad y temperatura, 15 valores rezagados no estacionales y 4 valores rezagados estacionales; con una capa de salida de una sola neurona. El entrenamiento se hizo utilizando el algoritmo de optimización de Broyden Fletcher Goldfarb Shaano y la función *sigmoid* como función de activación, este modelo presentó un MAPE de 0.0626 % y un RMSE de 0.0.107 (tabla VII).

Los valores de BIC y AIC comprueban que son modelos que pierden muy poca información y presentan un grado aceptable de parsimonia, por otro lado, los errores porcentuales absolutos medios se encuentran muy bajos para los tres modelos es decir los tres presentan muy buena exactitud, aunque el modelo de redes neuronales presenta un mejor valor en comparación con los otros dos.

Para evaluar la validez y confiabilidad de los modelos se realizó un análisis residual y además se utilizaron los datos de prueba con los cuales se evaluó el

pronóstico de 72 meses desde enero de 2014 a diciembre de 2019 por medio del MAPE y RMSE.

Por medio del análisis residual se determinó que los residuos para los tres modelos se comportaban como ruido blanco, los gráficos de ACF no mostraron más del 5 % de correlaciones significativas y los p-valores para las pruebas de Ljung-Box se encontraron por encima del 0.05 de significancia, por lo tanto los modelos se ajustan muy bien a los datos ya que en los residuos no existen indicios de información no explicada; por otra parte, las pruebas de Kolmogorov-Smirnov sobre los residuos comprobaron la falta la normalidad con probabilidades menores al 0.05 de significancia en el modelo ARIMA y de regresión dinámica lo cual agregará mayor grado de incertidumbre a los intervalos de confianza; mientras que el modelo de redes neuronales si presentó normalidad en sus residuos.

En la evaluación del pronóstico se encontró que el de regresión con errores ARIMA(2,1,3)(1,1,2)₁₂ era el que presentaba mejor exactitud con un MAPE de 1.47 % y un RMSE de 15816.35 MW, seguido por el modelo ARIMA(2,1,3)(2,1,3)₁₂ con un MAPE de 1.57 % y un RMSE de 16674.62 MW y por último el modelo de redes neuronales con un MAPE de 7.98 % y un RMSE de 76678.25 MW; a pesar de que el modelo de redes neuronales presentaba mayor exactitud con los datos de entrenamiento al evaluarlo contra los datos de prueba se identificó que el modelo no logró captar idealmente la estacionalidad y tendencia de la serie (figura 48), mientras que el modelo de regresión con errores ARIMA si lo logró (figura 47) al igual que el modelo ARIMA estacional (figura 46); adicionalmente se comprueba que la incorporación de variables en el pronóstico si ayuda a mejorar la exactitud del modelo. Por lo tanto podemos decir que el modelo Regresión con ARIMA(2,1,3)(2,1,1)₁₂ es el modelo más confiable de pronóstico para estimar la demanda de energía eléctrica en Guatemala.

De los resultados obtenidos es importante considerar que la variable PIB y población son estimadores que se calculan anualmente, mientras que el modelo de pronóstico fue planteado con una frecuencia mensual, por lo tanto es posible que planteando una frecuencia anual para el estudio las correlaciones presenten mejores resultados; con respecto a la construcción del modelo de redes neuronales únicamente se trabajó con modelos de redes neuronales multicapa, por lo cual podría ser conveniente modelar distintos tipos para tener un mejor resultado; de igual manera la incorporación de los valores rezados en el modelo se realizó a prueba y error, por lo cual es posible que diferentes configuraciones brinden mayor exactitud.

Análisis externo

Velásquez, Dyner y Casto (2007), en su estudio *¿Por qué es tan difícil obtener buenos pronósticos de los precios de la electricidad en mercados competitivos?*, hacen énfasis en que no existe una serie de juicios experimentales, metodológicos o teóricos sobre la selección de un modelo específico para el pronóstico sobre otras opciones, también hace ver la falta de claridad sobre cuál es el procedimiento más adecuado para realizar la estimación y que no existe un criterio común para la selección de un modelo óptimo; aun así Gil (2016) en su estudio *Pronóstico de la demanda mensual de electricidad con series de tiempo*, hace uso del principio de parsimonia para la elección del mejor modelo por medio del coeficiente de determinación ajustado, el criterio de información Akaike y el criterio de información Bayesiano; además hace uso del error promedio de pronóstico, el error promedio absoluto de pronóstico y el error cuadrático medio de pronóstico para medir la precisión de los modelos; con esto se pudo comprobar la validez de la metodología utilizada para la selección del

modelo de cada tipo, utilizando AIC, BIC y el error MAPE para elegir la mejor opción; además de la evaluación del pronóstico por medio del MAPE y el RMSE.

Con respecto a los resultados obtenidos, observamos similitud con el estudio realizado por Lozado, Morocho y Zurita (2019), en donde por medio de un análisis de correlación identificaron a las variables humedad y temperatura como variables de influencia en la demanda de energía eléctrica incorporándolas junto con otras variables de calendario para la construcción del modelo de predicción. Por otro lado Grimaldo, Mendoza y Reyes (2017) utilizaron el PIB y la población para la construcción de distintos modelos de regresión para la estimación de la demanda eléctrica, aunque en el mejor modelo estas no se incorporaron se comprobó cierta relación entre la demanda eléctrica y ambas variables, eso sí en un modelo que contemplaba periodicidad anual a diferencia de este estudio en donde se consideró la demanda mensual; Orellana (2012) hace referencia a esto en la tesis *Modelación y pronóstico de la demanda de energía eléctrica de mediano plazo de El Salvador*, en donde critica la utilización de la variable PIB en modelos econométricos para la estimación de la demanda debido a que son indicadores anuales y estos no son valores definitivos hasta 5 años después, lo cual complica las estimaciones con frecuencia diaria, semanal o mensual.

Gonzales (2004), en su estudio *Predicción de la demanda eléctrica horaria mediante redes neuronales artificiales* al igual que en este estudio realiza una comparación entre distintos modelos para la estimación de la demanda eléctrica siendo estos: uno redes neuronales, un modelo ingenuo, un ARMA y un alisado exponencial, en los resultados obtenidos el modelo de redes neuronales presenta mayor exactitud, seguido por el ARMA y el modelo de alisado exponencial; esto a diferencia de este estudio donde el mejor modelo fue el de regresión con errores ARIMA. De esta comparación podemos sacar la conclusión que no existe un

mejor método para realizar el modelo de estimación de la demanda de energía eléctrica, puesto que cada una de las series de tiempo estudiadas presenta diferentes características y retos.

CONCLUSIONES

1. Las variables identificadas que probablemente influyen en mayor medida en el comportamiento de la demanda de energía eléctrica en Guatemala fueron la humedad y temperatura.
2. Para construir el mejor modelo de predicción, con el mejor ajuste para la estimación de la demanda de energía eléctrica en Guatemala se construyeron y compararon los modelos: ARIMA(2,1,3)(2,1,3)₁₂, de regresión con errores ARIMA(2,1,3)(1,1,2)₁₂ y un modelo de redes neuronales con una capa oculta de 10 neuronas y 15 valores rezagados no estacionales y 4 estacionales.
3. Se evaluó la validez y confiabilidad de los modelos obteniendo para el modelo ARIMA(2,1,3)(2,1,3)₁₂ un MAPE de 1.57 % y un RMSE de 16,674.62MW, para el modelo de regresión con errores ARIMA(2,1,3)(1,1,2)₁₂ un MAPE de 1.47 % y un RMSE de 15,816.35MW y para el modelo de redes neuronales con un MAPE de 7.98 % y un RMSE de 76,678.25 MW, adicional se comprobó que los residuos se comportan como ruido blanco por medio de la prueba de Box-Ljung.
4. Se infiere que el modelo más confiable para estimar la demanda de energía eléctrica en Guatemala es el modelo de regresión con errores ARIMA(2,1,3)(1,1,3)₁₂ con un MAPE de 1.47 % y un RMSE de 15,816.35MW.

RECOMENDACIONES

1. En la identificación de variables influyentes en el comportamiento de la demanda de energía se sugiere adicionar indicadores económicos y demográficos porcentuales, ya que este tipo de indicadores son más propensos a presentar un comportamiento estacionario, lo cual facilita su incorporación en el estudio de modelos de series de tiempo. Adicional se podrían incorporar factores tecnológicos para estudiar el comportamiento que han tenido los avances en la tecnología en la demanda energética.
2. En la construcción de modelos de predicción para la estimación de la demanda de energía, se propone analizar otros tipos y diseños de redes neuronales con el fin de determinar el tipo y modelo que mejor se ajuste al problema.
3. En la evaluación de la validez y confiabilidad del modelo, se sugiere evaluar el pronóstico para los años actuales, con el fin de tener más información para la evaluación de los modelos.
4. Se propone evaluar el modelo de regresión con errores $ARIMA(2,1,3)(1,1,3)_{12}$ con el modelo de estimación de la demanda de energía utilizado por el AMM, con el fin de comparar precisión en los resultados.

REFERENCIAS

1. Administrador del Mercado Mayorista. (2020). *Mercado Eléctrico de Guatemala*. Guatemala: Administrador del Mercado Mayorista. Recuperado de [https://www.amm.org.gt/portal/?wpfb_dl=610Guía%20para%20inversiones%20mercado%20eléctrico%20de%20Guatemala%20-VF-\[4\]%20\(1\).pdf](https://www.amm.org.gt/portal/?wpfb_dl=610Guía%20para%20inversiones%20mercado%20eléctrico%20de%20Guatemala%20-VF-[4]%20(1).pdf)
2. Ariza, A. (2013). *Métodos utilizados para el pronóstico de demanda de energía eléctrica en sistemas de distribución* (Tesis de pregrado). Universidad Tecnológica de Pereira, Pereira. Colombia.
3. Bello L. y Martínez, S. (2007). Una metodología de series de tiempo para el área de la salud; caso práctico. *Revista Facultad Nacional de Salud Pública*, 25(2), 177-122.
4. Camino, F. (marzo 2013). *Tema: Series de tiempo*. [Mensaje en un blog]. Recuperado de: <https://es.calameo.com/read/002332565a6be51cc89bb>
5. Díaz, J. (agosto 2005). *Análisis Residuales*. [Mensaje en un blog]. Recuperado de <http://math.uprag.edu/residuales1.pdf>
6. García, Y. (2009). *Predicción de la demanda eléctrica empleado redes neuronales artificiales* (Tesis de pregrado). Universidad Central “Marta Abreu” de Las Villas, Santa Clara. Cuba.

7. Gil, V. (2016). Pronóstico de la demanda mensual de electricidad con series de tiempo. *Revista EIA*, 13(26), 111-120.
8. Gonzales, C. (2004). Predicción de la demanda eléctrica horaria mediante redes neuronales artificiales. *Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, 5 (1), 5-28.
9. Grimaldo, J., Mendoza, M. y Reyes, W. (2017). Modelo para pronosticar la demanda de energía eléctrica utilizando los productos internos brutos sectoriales: Caso Colombia. *Revista Espacios*, 38 (22). 3-15.
10. Hanke, J. y Whichern, D. (2006). *Pronósticos en negocios*, México, México: Pearson Educación México, S.A. de C.V.
11. Holmes, E., Scheurell, M. y Wardm E. (mayo 2020). *Dickey-Fuller and Augmented Dickey-Fuller tests*. *EEUU: Applied Time Series Analysis*. [Mensaje en un blog]. Recuperado de <https://nwfsc-timeseries.github.io/atsa-labs/sec-boxjenkins-aug-dickey-fuller.html>
12. Hyndman, R. y Athansopoulos, G. (2018). *Forecasting: principles and practice*, (2a. edición. Melbourne, Australia: OTexts. Recuperado de <https://otexts.com/fpp2/>.
13. INEI. (2006). *Glosario básico de términos estadísticos*. INEI. Recuperado de https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib0900/Libro.pdf

14. Levin, R. y Rubin, D. (2004). *Estadística para Administración y Economía*, México, México: Pearson Educación México, S.A. de C.V.
15. Lozano I., Morocho, A. y Márquez, C. (2019). Predicción de demanda de energía eléctrica mediante redes neuronales artificiales. *Revista Ibérica de Sistemas e Tecnologías de Información*, Lousada (25), 505-519.
16. Matich, D. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*. [Mensaje en un blog]. Recuperado de <ftp://decsai.ugr.es/pub/usuarios/castro/Material-Redes-Neuronales/Libros/matich-redesneuronales.pdf>
17. McDonald, J. (2014). *Hanbook of Biological Statistics*, Maryland, Estados Unidos: Sparky House Publising.
18. Montes, E., Calvete, F. y Mantilla, C. (2016) Aplicación de series de tiempo en la realización de pronósticos de producción. *Revista Fuentes: El Reventón Energético*, 14(1), 79-88.
19. Mordjaoui, M., Haddad, S., Medoued, A., y Laouafi, A. (1 de diciembre 2017). Electric load forecasting by using dynamic neural network. *International Journal of Hydrogen Energy*, 1–9. <https://doi.org/10.1016/j.ijhydene.2017.03.101>
20. Nava, N. (2013). *Procesamiento de series de tiempo*, México, México: Fondo de cultura económica.

21. Olleh, D. (junio 2019). *Package 'seastests'* [Mensaje en un blog]. Recuperado de <https://cran.r-project.org/web/packages/seastests/seastests.pdf>.
22. Orellana, J. (2012). *Modelación y pronóstico de la demanda de energía eléctrica de mediano plazo de El Salvador* (Tesis de maestría). Universidad de El Salvador. El Salvador.
23. Ortiz, M. (2015). *Pronóstico de la demanda eléctrica residencial basado en el modelo regresión adaptativa Multivariante Spline (MARS)* (Tesis de pregrado). Universidad Politécnica Salesiana. Ecuador, Quito.
24. Palma, A. (2015). *Pronóstico de demanda de energía y potencia eléctrica en el largo plazo para la red de Chilectra s.a. utilizando técnicas de minería de datos* (Tesis de maestría). Universidad de Chile. Chile.
25. Polanco, A. (2020). *Series de Tiempo: Conceptos básicos de análisis de series de tiempo en el dominio del tiempo*, Madrid, España: Daniel José Reyes Valero.
26. Pohlert, T. (abril 2020). *Package 'trend'* [Mensaje en un blog]. Recuperado de <https://cran.r-project.org/web/packages/trend/trend.pdf>
27. Rueda, V., Velásquez, J. y Franco, C. (2010). Avances recientes en la predicción de la demanda de electricidad usando modelos no lineales. *Dyna*, 78(167), 36-43.

28. Soret I. y Gimenez E. (2013). *Previsión de ventas y fijación de objetivos*, Madrid, España: ESIC editorial.
29. Tomás-Sábado, J. (2010). *Fundamentos de bioestadística y análisis de datos para enfermería*, Barcelona, España: Universidad Autónoma de Barcelona.
30. Velásquez, J., Dyner, I. y Castro R. (2007). ¿Por qué es tan difícil obtener buenos pronósticos de los precios de la electricidad en mercados competitivos? *Cuadernos de Administración*, 20 (34), 259 – 282.
31. Villavicencio, J. (octubre 2012). *Introducción a series de tiempo* [Mensaje en un blog]. Recuperado de: http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D#:~:text=Una%20serie%20tiempo%20es%20una,su%20análisis%20para%20hacer%20pronóstico.
32. Vinuesa, P. (2016). Tema 8 - Correlación: teoría y práctica. México: CCG-UNAM. Recuperado de https://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema8_correlacion.html#introduccion-el-concepto-de-correlacion
33. Wang, J., y Zhu, Q. (2015). Short-term Electricity Load Forecast Performance Comparison Based on Four Neural Network Models. *The 27th Chinese Control and Decision Conference (2015 CCDC)*, 2928–2932.

