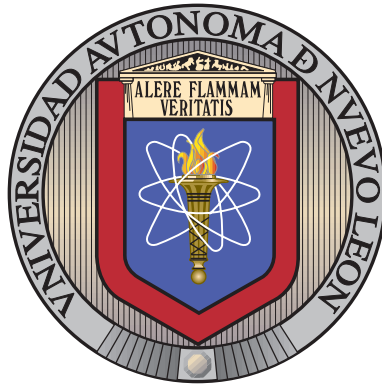


UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



TESIS
AJUSTE DE UN MODELO AUTORREGRESIVO PARA
EL PRONÓSTICO DEL PRECIO HORARIO DE LA
ENERGÍA EN EL MERCADO ELÉCTRICO
MAYORISTA MEXICANO

POR

CARLOS ANDRÉS RODRÍGUEZ AVENDAÑO

EN OPCIÓN AL GRADO DE
MAESTRÍA EN CIENCIAS
CON ORIENTACIÓN EN MATEMÁTICAS

SAN NICOLÁS DE LOS GARZA, NUEVO LEÓN, MÉXICO

OCTUBRE 2023

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

CENTRO DE INVESTIGACIÓN EN CIENCIAS FÍSICO MATEMÁTICAS



TESIS
AJUSTE DE UN MODELO AUTORREGRESIVO PARA
EL PRONÓSTICO DEL PRECIO HORARIO DE LA
ENERGÍA EN EL MERCADO ELÉCTRICO
MAYORISTA MEXICANO

POR

CARLOS ANDRÉS RODRÍGUEZ AVENDAÑO

EN OPCIÓN AL GRADO DE
MAESTRÍA EN CIENCIAS
CON ORIENTACIÓN EN MATEMÁTICAS

SAN NICOLÁS DE LOS GARZA, NUEVO LEÓN, MÉXICO

OCTUBRE 2023

Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas
Centro de Investigación en Ciencias Físico Matemáticas

Los miembros del Comité de Tesis recomendamos que la Tesis “Ajuste de un modelo autorregresivo para el pronóstico del precio horario de la energía en el Mercado Eléctrico Mayorista mexicano”, realizada por el alumno Carlos Andrés Rodríguez Avendaño, con número de matrícula 2085205, sea aceptada para su defensa como opción al grado de Maestría en Ciencias con Orientación en Matemáticas.

El Comité de Tesis

Dr. Francisco Javier Almaguer Martínez
Asesor

M. Rubi Isela Gutierrez López
Revisor

Dr. Francisco Hernández Cabrera
Revisor

Vo. Bo.

Dr. Omar Jorge Ibarra Rojas
Coordinador del Posgrado en Ciencias con Orientación en Matemáticas

AGRADECIMIENTOS

Me gustaría agradecer en estas líneas la ayuda que muchas personas y colegas me han prestado durante el proceso de investigación y redacción de este trabajo.

En primer lugar, a mi director de tesis, Dr. Javier Almaguer, y a mis coasesores y jurados, Lic. MS. Rubí Gutierrez y Francisco Cabrera, por haber orientado la presente investigación y haber proveído la información y el soporte requerido para culminar exitosamente el estudio.

A CONACyT por la beca otorgada, mediante CVU 1100585, y a la Universidad Autónoma de Nuevo León UANL, gracias a los cuales pude cursar la presente maestría, adquirir nuevos conocimientos y desarrollar consecuentemente esta tesis.

A Clemencia Gaitán y Misael González (QEPD) por su orientación y atención a mis consultas sobre metodología de la investigación, así como la cuidadosa revisión del presente documento de tesis y sus pertinentes sugerencias.

A mi familia, mis padres y hermana, quienes siempre me inculcaron la importancia de la educación en mi desarrollo personal y profesional, así como por su apoyo incondicional durante el curso de mi maestría.

Por último, un especial agradecimiento a mi esposa e hija, por su paciencia, comprensión y solidaridad con este proyecto, por el tiempo que me han concedido y que han dedicado para hacer de esta tesis una realidad; por eso, este trabajo es también el suyo.

ÍNDICE GENERAL

INTRODUCCIÓN	IX
1. PLANTEAMIENTO DEL PROBLEMA	1
1.1. TEMA	1
1.2. DESCRIPCIÓN DEL ÁREA PROBLEMÁTICA	1
1.3. PREGUNTA DE INVESTIGACIÓN	3
1.4. JUSTIFICACIÓN	3
1.5. OBJETIVOS	4
1.6. HIPÓTESIS	4
1.7. VARIABLES	5
1.8. ALCANCE Y LIMITACIONES	5
2. MARCOS DE REFERENCIA	6
2.1. REVISIÓN BIBLIOGRÁFICA	6
2.1.1. Estado del arte: Métodos para el pronóstico de los precios de la energía eléctrica	7
2.1.1.1. Modelos multiagente: Simulación multiagente, equilibrio, teoría de juegos	7
2.1.1.2. Métodos fundamentales: estructurales	7
2.1.1.3. Modelos de forma reducida	9
2.1.1.4. Enfoques estadísticos (Econometría, análisis técnico)	10

2.1.1.5.	Técnicas de Inteligencia computacional (Inteligencia artificial, estadística no lineal)	12
2.1.2.	Pronóstico de precios de la energía mediante modelos autorregresivos	12
2.2.	ANTECEDENTES	18
2.2.1.	Los inicios (1879-1936)	18
2.2.1.1.	Mexican Light and Power Company (MLPC)	19
2.2.1.2.	Primeros intentos de regulación	20
2.2.2.	Co-regulación (1937-1959)	21
2.2.2.1.	La CFE (Comisión Federal de Electricidad)	21
2.2.2.2.	La ley de la industria eléctrica de 1939 y sus efectos	22
2.2.3.	Integración vertical (1960-1991)	23
2.2.4.	Inicio de la apertura (1992-2012)	24
2.3.	MARCO TEÓRICO	27
2.3.1.	Sistema Eléctrico Nacional (SEN)	27
2.3.1.1.	Infraestructura del Sistema Eléctrico Nacional	27
2.3.1.2.	Demanda y Consumo de Energía Eléctrica	32
2.3.1.3.	Funcionamiento del Mercado Eléctrico Mayorista (MEM)	35
2.3.2.	Fundamentación Matemática	38
2.3.2.1.	Series de tiempo	38
2.3.2.2.	Coeficiente de Hurst	53
2.3.2.3.	Transformada de Fourier	58
3.	METODOLOGÍA	63
3.1.	Tipo de estudio	65
3.2.	Muestreo	65
3.3.	Estructura de los datos	65

3.3.1. Datos atípicos	66
3.3.2. Correlación	84
3.4. Técnicas de recolección de la información	93
3.5. Prueba piloto	95
3.5.1. Descomposición	96
3.5.2. Predicción	99
3.6. Técnicas de análisis de la información	101
3.6.1. Espectograma	101
3.6.2. Análisis polar	107
3.7. Predicción	108
3.8. Cronograma	119
4. PRESENTACIÓN Y ANÁLISIS DE RESULTADOS	121
4.1. Comparación entre una y múltiples periodicidades	121
4.2. Resultados de la prueba piloto para la selección del modelo autorregresivo .	122
4.3. Análisis de correlación de largo alcance (Exponente de Hurst)	123
4.4. Aplicación del modelo autorregresivo	126
4.5. Modelos SARIMA	127
4.5.1. Región Centro Año 2022	127
4.5.2. Región Centro Año 2019	129
4.5.3. Región Centro Año 2020	130
4.5.4. Resumen de los resultados	132
5. CONCLUSIONES	139
6. RECOMENDACIONES	143
Apéndice A. Solución del modelo autorregresivo AR de primer orden	145

ÍNDICE GENERAL	VIII
<hr/>	
Apéndice B. Frecuencia de Nyquist	150
Apéndice C. Pruebas Estadísticas	153
C.1. Box-Pierce	153
C.2. Ljung-Box	153
C.3. Dickey Fuller	154
BIBLIOGRAFÍA	162
Índice de figuras	162
Índice de tablas	167
Glosario	168
Siglas	170

INTRODUCCIÓN

El mercado eléctrico mexicano, al igual que otros mercados, funciona con base en la oferta y la demanda que se da a partir de interacciones de quienes generan y consumen. Esta interacción es precisamente la que determina los precios horarios de la energía y por ende, guía la estrategia de los principales consumidores mayoristas. En este orden de ideas, la predicción del precio horario de la energía eléctrica en diferentes mercados en el mundo ha sido sujeto de numerosas investigaciones y se considera de suma importancia en la competitividad de las empresas y los países.

El presente trabajo de investigación tiene como objetivo ajustar un modelo matemático que permita predecir el precio horario de la energía eléctrica en las diferentes regiones del mercado eléctrico mexicano y con esto, proveer a los jugadores del mercado, información relevante y confiable en lo que a este factor respecta.

El **Capítulo 1** presenta el planteamiento del problema y describe de manera general el funcionamiento del mercado eléctrico incluyendo el concepto de subasta inversa y la forma como el precio horario de la energía eléctrica es fijado gracias a las interacciones que ocurren entre los diferentes jugadores del mercado en lo que se denomina. Se plantea además la importancia que tiene dicho pronóstico para los diferentes participantes del mercado eléctrico mayorista soportando la justificación que sustenta este estudio. Cabe mencionar que este capítulo propone la pregunta de investigación que se resume en la búsqueda de un modelo autorregresivo capaz de predecir, de forma confiable, el precio horario de la energía en un horizonte de tiempo de una semana. Finalmente, se limita el alcance y se plantean posibles corrientes de estudio futuras.

El **Capítulo 2** presenta los marcos de referencia en los cuales se describen de forma detallada, las bases teóricas que sustentan la presente investigación. El Capítulo inicia presentando las diferentes técnicas que se pueden utilizar para el pronóstico del precio horario de la energía y prosigue con un compendio de diferentes trabajos de investigación que han utilizado modelos autorregresivos para este fin en mercados alrededor del mundo. Posteriormente, hace un recuento de los antecedentes del mercado eléctrico mexicano, su

progreso, reformas y hechos relevantes que llevaron a desarrollar un mercado como el que hoy en día existe. Finalmente se presenta la fundamentación matemática relevante.

El **Capítulo 3** describe la metodología utilizada incluyendo todo lo relacionado con el muestreo y los datos utilizados para el ajuste del modelo, los resultados de las pruebas piloto que se corrieron con dicho set de datos y el procesamiento de los mismos, para adaptarlos al formato y forma que requiere el modelo.

El **Capítulo 4** introduce los resultados obtenidos iniciando con una presentación de la tendencia histórica de los precios de la energía eléctrica, mediante el análisis de los datos de un periodo de 4 años en las diferentes zonas del país. Explora también los resultados de las pruebas piloto como un medio para determinar la forma más apropiada de abarcar la predicción de precios requerida y concluye con la aplicación del modelo autorregresivo SARIMA, que presenta el mejor desempeño al eliminar los datos atípicos de la serie.

Finalmente la sección de conclusiones, lista los hallazgos más relevantes del desarrollo y resultados del presente trabajo de investigación. La disertación finaliza con una serie de recomendaciones que pueden guiar futuros trabajos y extender el alcance de los mismos con el fin de mejorar el desempeño del modelo e incluir variables exógenas que aumentan el nivel de precisión del modelo aquí presentado.

CAPÍTULO 1

PLANTEAMIENTO DEL PROBLEMA

La electricidad es una de las principales formas de energía utilizadas hoy en día y se consolida como parte fundamental en el funcionamiento y desarrollo de las principales actividades económicas en el mundo. Es por esto, que el establecimiento de un mercado para su comercialización es de vital importancia para la competitividad de un país.

El presente capítulo presenta el tema, la descripción del área problemática, la pregunta de investigación que se pretende responder, la justificación que respalda la misma, los objetivos establecidos, las hipótesis planteadas y las variables.

1.1 TEMA

Modelo para la predicción de precios horarios de la energía eléctrica en el Mercado Eléctrico Mayorista (Mercado Eléctrico Mayorista (MEM)) de México.

1.2 DESCRIPCIÓN DEL ÁREA PROBLEMÁTICA

El mundo está sumergido en la dicotomía de continuar explotando materias primas hasta que se agoten o comenzar una transformación energética que prepare a los países para cuando estos recursos ya no estén disponibles o bien que, por situaciones ambientales, ya no sea posible explotar. México no es la excepción. Durante las últimas dos décadas ha venido dándose una serie de reformas para estimular la generación de electricidad por medio de fuentes renovables – modificando la matriz energética- que como se verá más adelante, privilegió la venta de energía proveniente de fuentes renovables en detrimento de los generadores tradicionales; la última gran reforma ocurrió en el 2014.

Esta reforma fue una profundización de pequeñas reformas que ya venían en discusión durante varios años anteriores. La reforma finalmente aprobada incluyó una nueva propuesta de reforma eléctrica en la que se tienen muchos actores involucrados que ac-

túan en el mercado eléctrico. La última reforma energética del 2014 permitió que nuevos actores entraran al mercado eléctrico, no solo en la parte de generación o la matriz energética del país sino también en el sector de la comercialización que permitió dinamizar las transacciones energéticas. Las compañías privadas, desde el periodo de liberalización han venido realizando inversiones para ampliar las capacidades de las plantas o desarrollando empresas de comercialización de energía sin poseer equipos de generación. Durante este periodo estas nuevas empresas han trabajado en paralelo con la única empresa estatal, que previo a la reforma en mención, era la única empresa encargada de fijar el precio de la energía, al ser una compañía integrada verticalmente.

La reforma, en la forma como está concebida, tiene algunos retos y desafíos debido a la gran volatilidad del precio de la energía, en gran parte por la complejidad del mercado que se desarrolló cuando estos actores se integraron, lo que hace difícil que el público en general pueda entender de forma clara su funcionamiento o inclusive las decisiones que se toman, y en qué medida los benefician o les afectan. La fijación de precios se realiza mediante un mecanismo que se conoce como subasta inversa.

La Subasta Inversa consiste en que cada generador compite por suministrar el Megavatio-hora más barato, con el fin de ser los primeros a los que les van a comprar toda la energía que producen, en este mismo proceso los compradores se encargan de enviar un tercero (el centro de control - Centro Nacional de Control de Energía (CENACE)) la cantidad de energía que necesitan para el día siguiente. El resultado de cruzar los precios es que se llega a un precio común, que será el precio en el que se pagará la demanda de energía, esto se conoce como el precio del Mercado de Día en Adelanto, suele abreviarse como Mercado de Día de Adelanto (MDA).

Si ningún evento anormal sucede estos precios suelen tener un comportamiento persistente, es decir si su tendencia es a crecer seguirán creciendo, por el contrario, si la tendencia es descendente se mantendrán de esa forma. Lo que nos indican una dependencia de largo tiempo sobre los precios del pasado. Y esta es la razón para emplear las series de tiempo, ya que los modelos autorregresivos nos indican la existencia de una relación con los datos pasados, dependiendo del ajuste de los parámetros propios del modelo; la pregunta aquí es que tan larga es esta dependencia, ¿los precios de hace un año tienen alguna incidencia en los de hoy? Esta pregunta será respondida conforme se vaya desarrollando el trabajo.

La red eléctrica mexicana está compuesta por diferentes tipos de tecnología de generación de energía eléctrica, tanto renovable como no renovable, que implican una compleja red de generación y distribución. Un aspecto relevante de su complejidad radica en la fijación de los precios de la energía eléctrica, que depende de distintos factores como la

capacidad máxima diaria, el clima, la demanda, los tipos de tecnología y la interacción entre nodos. El cálculo de este precio se realiza a través una subasta inversa, en la cual los generadores compiten por ofrecer la energía al menor precio, mientras los compradores proponen una cantidad de energía a comprar. Como resultado se llega a un precio común que es el que se determina como valor a utilizar. Esta forma de operación determina lo que es un mercado no regulado en el que los precios son determinados por las acciones entre los participantes del mercado. Esto, además, hace que factores que afecten la generación o el consumo, jueguen un papel relevante en la determinación de los precios y los hagan altamente volátiles, tales como el clima y la temperatura, los precios de los combustibles, etc.

La volatilidad de los precios y la forma en la que opera este tipo de mercados, hacen que la estimación y predicción de los precios de la energía eléctrica sea vital para para que productores y consumidores puedan planear su operación, su estrategia de subasta, negociar contratos y planear inversiones. Debido a la importancia económica que esto conlleva, numerosas investigaciones han modelado distintos mercados alrededor del mundo, utilizando diferentes tipos de herramientas y modelos. No obstante, el mercado mexicano parece no haber sido explorado más allá de lo que respecta a su estructura y funcionamiento por lo que carece de herramientas para el pronóstico de precios que pueda apoyar la toma decisiones de sus participantes.

1.3 PREGUNTA DE INVESTIGACIÓN

Con base en lo anterior, el presente trabajo de grado buscar responder a la siguiente pregunta de investigación:

¿Cómo ajustar un Modelo autorregresivo para predecir el precio de la energía hora a hora, en un horizonte de tiempo de una semana, para el mercado eléctrico mexicano?

1.4 JUSTIFICACIÓN

De acuerdo con lo anteriormente mencionado, y dada la importancia de las tarifas de la energía eléctrica para la planeación y desarrollo de la economía de un país, numerosas investigaciones han buscado estimar modelos matemáticos que permitan pronosticar el precio horario de la energía eléctrica con un nivel de exactitud aceptable. Dichas investigaciones se han enfocado en mercados maduros con disponibilidad de datos suficiente como son los mercados del norte de Europa, España, India, California y otros mercados

de Estados Unidos que permitan capturar en un modelo, la volatilidad que, normalmente, presentan los precios de la electricidad y representar los diferentes factores estacionales que los pueden afectar. No obstante, aunque existen trabajos descriptivos y de contexto en mercados latinoamericanos, la literatura disponible no reporta estudios donde se presente un modelo matemático de pronóstico de tarifas que le permita a los jugadores del mercado de esos países tomar mejores decisiones.

Así pues, la presente investigación busca llenar ese vacío proponiendo un modelo de predicción de precios horarios de electricidad de buen nivel de rendimiento y exactitud aplicado al mercado eléctrico mexicano; este desarrollo permitirá a los diferentes jugadores del mercado establecer sus estrategias de subasta y por ende, optimizar sus operaciones.

1.5 OBJETIVOS

General:

Ajustar un modelo autorregresivo que permita predecir el precio de la energía hora a hora, en un horizonte de tiempo semanal, para las diferentes zonas del mercado eléctrico mexicano, con el fin de apoyar la toma de decisiones de las empresas del sector.

Específicos:

- Analizar las tendencias históricas de los precios de la energía eléctrica y los factores que impactan la determinación del Precio Marginal Local (PML).
- Identificar comportamientos atípicos en la estructura de precios.
- Seleccionar el modelo autorregresivo que describa el comportamiento del mercado eléctrico mexicano.
- Establecer el grado del polinomio del modelo autorregresivo y establecer el valor de los coeficientes del mismo.
- Probar el modelo para determinar el desempeño de la predicción mediante la medición del error, Root Mean Square Error, RMSE.

1.6 HIPÓTESIS

En este estudio, se plantea la hipótesis de que el comportamiento del mercado eléctrico mexicano puede ser modelado efectivamente mediante un enfoque de modelo auto-

regresivo (S-ARIMA). Se busca evaluar la capacidad del modelo AR para capturar las relaciones y patrones en los precios de la electricidad, considerando las dinámicas de oferta y demanda, así como posibles influencias externas. Se pretende investigar si la naturaleza temporal y la dependencia de valores pasados en los precios eléctricos hacen que el modelo AR sea una herramienta adecuada para describir y pronosticar el comportamiento del mercado eléctrico en México.

1.7 VARIABLES

- **Variable dependiente:** Precios futuros de la energía eléctrica (predicción mediante modelo autorregresivo).
- **Variable independiente:** Precios históricos de la energía eléctrica y sus medidas estadísticas.

1.8 ALCANCE Y LIMITACIONES

El presente trabajo de investigación limita su alcance al pronóstico del precio horario de la energía del mercado eléctrico mexicano en sus diferentes zonas de operación mediante la aplicación de un modelo autorregresivo. Este modelo se construye con base en información histórica disponible de los precios y no toma en cuenta otros factores exógenos pudieran afectarlos. Se asume que el set de datos disponible agrupa información de un periodo suficientemente largo y por lo tanto describe correctamente el comportamiento volátil de los precios y la estacionalidad de los mismos.

Del mismo modo, el modelo está en capacidad de pronosticar los precios horarios en una ventana de tiempo de una semana, teniendo en cuenta que el proceso de subasta inversa se da diariamente y que por ende, los jugadores del mercado eléctrico tendrían información suficiente para la toma oportuna de decisiones; por otro lado, se ha identificado, por la revisión bibliográfica, que se trata de un horizonte de tiempo comúnmente utilizado por otros investigadores para este fin.

Futuras investigaciones incluyendo otro tipo de modelos así como factores exógenos podrían aumentar la precisión del modelo aquí presentado y con esto, mejorar su desempeño.

CAPÍTULO 2

MARCOS DE REFERENCIA

Este capítulo de marcos de referencia pretende recopilar los estudios, investigaciones e información más relevante en lo que respecta a la modelación para el pronóstico de precios horarios de energía y al mercado eléctrico mexicano.

Por un lado, explora los diferentes métodos que existen para el pronóstico de precios de energía resaltando la relevancia de este tipo de investigaciones para los mercados mundiales y la carencia de estudios similares aplicados al mercado eléctrico nacional. En este punto, se hace especial énfasis en los modelos autorregresivos; Posteriormente, se hace una explicación del mercado eléctrico mexicano, desde su historia hasta los detalles de su distribución y funcionamiento; esto con el fin de localizar el área de trabajo y delimitar el alcance de la presente investigación. Finalmente, se presenta una sección de fundamentación matemática que explora información básica de teoría de series de tiempo.

2.1 REVISIÓN BIBLIOGRÁFICA

En esta sección se presenta una compilación de algunas de las más importantes investigaciones y publicaciones que han ajustado modelos matemáticos para el pronóstico del precio horario de la energía eléctrica, dividida en dos partes: Una primera parte expone el estado del arte de los diferentes métodos que se han ido desarrollando para el pronóstico de precios de energía y que la literatura ha distribuido en 5 grandes categorías o grupos de modelos; por otro lado, la segunda parte se enfoca principalmente en modelos autorregresivos que han sido utilizados para pronosticar precios de la energía en diferentes mercados a nivel global.

El resultado de esta revisión bibliográfica revela que que la mayoría de investigaciones y estudios relacionados, se han enfocado principalmente en mercados maduros con disponibilidad de datos suficiente como para capturar en la modelación, la volatilidad propia de los precios horarios de la energía.

2.1.1 ESTADO DEL ARTE: MÉTODOS PARA EL PRONÓSTICO DE LOS PRECIOS DE LA ENERGÍA ELÉCTRICA

A continuación se presenta una compilación de los diferentes métodos utilizados para el pronóstico del precio de la energía eléctrica en 5 categorías de acuerdo con [65]: Modelos multiagente, modelos fundamentales, modelos forma reducida, enfoques estadísticos y técnicas de inteligencia computacional.

2.1.1.1 Modelos multiagente: Simulación multiagente, equilibrio, teoría de juegos

2.1.1.1.1 Definición

- Simulan la operación de un sistema de agentes heterogéneos (unidades de generación, compañías, etc.) interactuando entre ellos, y construye un proceso de precios coincidente entre la demanda y la oferta en el mercado.
- Se consideran modelos muy flexibles por lo que requieren una serie de suposiciones que deben ser soportadas teórica o empíricamente.
- Generalmente se enfocan en aspectos cualitativos y pueden entregar una idea de si los precios estarán o no por encima de los costos marginales y como estos pueden influenciar a los jugadores del mercado.

2.1.1.1.2 Tipos

La tabla 2.1 resume los tipos de modelos que se consideran dentro de la categoría multiagente y sus respectivas consideraciones.

2.1.1.2 Métodos fundamentales: estructurales

2.1.1.2.1 Definición

- Describen la dinámica de precios modelando los impactos de factores físicos y económicos relevantes en el precio de la electricidad..
- No se consideran muy precisos en el pronóstico de precios horarios y por ende, se usan más en análisis diarios.

Tabla 2.1: Modelos multiagente

Tipo	Consideraciones y resultados
Nash-Cournot	La electricidad es tratada como un bien homogéneo y el equilibrio del mercado es determinado por la capacidad del suministro.
	Tienden a proveer precios más altos que los que suceden en la realidad.
	Propone un modelo estocástico basado en subastas para predecir el precio horario y promedio de la electricidad en un periodo de tiempo.
Equilibrio función suministro	Modela el precio como el equilibrio entre las compañías que subastan con curvas de suministro en el mercado eléctrico mayorista.
	Requiere una serie de ecuaciones que manejo más complejo.
	Mayormente utilizado en análisis de estrategias de subasta, diseño de mercados y manejo de la congestión pero poco en pronóstico de precios.
Costo de producción estratégico	Asume que cada agente trata de maximizar sus beneficios basado en sus propios costos y las posibles reacciones de sus competidores. Utiliza un parámetro estratégico que representa la pendiente de la función de la demanda residual para cada nivel de producción del generador..
	Al no hacer iteraciones, no tiene en cuenta las reacciones posteriores de los competidores pero es compila rápidamente debido a su mayor simplicidad.
Basado en agente	Es una clase de estructuras computacionales para simular las acciones e interacciones entre agentes con el objetivo de evaluar sus efectos en el sistema como un todo.

- Entregan un panorama razonable de la dinámica de los precios de la energía eléctrica y se comporta bien ante escenarios de alta volatilidad y pronóstico de precios pico.

2.1.1.2.2 Tipos

La tabla 2.2 resume los tipos de modelos que se consideran dentro de la categoría de métodos fundamentales y sus respectivas consideraciones.

Tabla 2.2: Métodos fundamentales

Tipo	Consideraciones y resultados
Fundamental rico en parámetros	Introduce variables para incluir factores como temperatura, clima, combustibles, paros)
Estructural Parsimonioso	Introduce también factores como costos de combustibles, demanda y capacidad de generación.
	Utiliza únicamente las variables y características esenciales que tienen un impacto significativo en la variable objetivo que se está tratando de predecir.

2.1.1.3 Modelos de forma reducida

2.1.1.3.1 Definición

- Caracterizan las propiedades estadísticas de los precios de la electricidad en el tiempo con el objetivo de evaluar derivados y manejo de riesgos.
- No se consideran muy precisos en el pronóstico de precios horarios y por ende, se usan más en análisis diarios. .
- Entregan un panorama razonable de la dinámica de los precios de la energía eléctrica y se comporta bien ante escenarios de alta volatilidad y pronóstico de precios pico.

2.1.1.3.2 Tipos

La tabla 2.3 resume los tipos de modelos que se consideran dentro de la categoría de modelos de forma reducida y sus respectivas consideraciones.

Tabla 2.3: Modelos de forma reducida

Tipo	Consideraciones y resultados
Difusión de saltos	Basado en el movimiento browniano.
	Ignora el hecho de que los precios de la energía son un proceso que revierte a la media del precio "normal" del régimen. Así, si un pico ocurre, el modelo asume que éste pico establece un nuevo nivel de precio y hace el proceso de difusión a partir de éste..
Cadena de Markov	Mejora el manejo de precios pico incluso en horizontes de tiempo horarios. Retorna los precios a su régimen "normal", en la medida en la que acepta cambios temporales en la dinámica del modelo.

2.1.1.4 Enfoques estadísticos (Econometría, análisis técnico)

2.1.1.4.1 Definición

- Son aplicaciones directas de las técnicas estadísticas para el pronóstico de carga o para la implementación de modelos econométricos en el mercado de energía.
- Utiliza una combinación matemática de precios previos y/o factores exógenos previos o actuales. Sin embargo, la precisión de los modelos depende no solo de la eficiencia de los algoritmos empleados sino también de la calidad de la data analizada y de la capacidad de incorporar factores fundamentales como demanda histórica, pronósticos de demanda y consumo, pronóstico del clima o precios de combustibles.
- Sin embargo, en presencia de picos, los métodos estadísticos pueden perder precisión en modelos que solo tienen en cuenta precios.

2.1.1.4.2 Tipos

La tabla 2.4 resume los tipos de modelos que se consideran dentro de la categoría de enfoques estadísticos y sus respectivas consideraciones.

Tabla 2.4: Modelos de forma reducida

Tipo	Consideraciones y Resultados
Suavizado exponencial (día similar)	Basado en la búsqueda de data histórica para un día con condiciones similares al día a ser pronosticado. Dichas condiciones similares pueden ser: el día de la semana o del año, clima o datos de consumo.
	El pronóstico puede ser una combinación lineal o un regresión que incluya varios días similares.
	El suavizado exponencial hace referencia al uso del promedio ponderado de observaciones pasadas.
Regresión	Su objetivo es entender las relaciones entre variables independientes y dependientes.
AR, ARX	ARMA: Media Móvil. Asume que la serie de tiempo bajo estudio es débilmente estacionaria.
	ARIMA: Media Móvil Integrada; SARIMA: Periódica.
	ARX: Incluye además variables exógenas como carga, temperatura, viento, etc.
TAR (AR "Threshold")	El régimen puede ser determinado por una variable observable y consecuentemente los regímenes que han ocurrido en el pasado y presente son conocidos con certeza.
	Se pueden incluir igualmente variables exógenas.
	Utiliza una variable límite (threshold variable) que debe ser seleccionada y de la cual depende la precisión del modelo.
GARCH (Heterocedasticidad)	Usados para comportamientos con varianza y covarianza NO constantes.
	Los precios de la energía eléctrica reportan dinámicas no lineales (fuerte dependencia de la variabilidad de su propio pasado).
	La varianza condicional depende no solo de los valores pasados de la serie de tiempo, sino también de la media móvil de las varianzas condicionales pasadas.

2.1.1.5 Técnicas de Inteligencia computacional (Inteligencia artificial, estadística no lineal)

2.1.1.5.1 Definición

- Combina elementos de aprendizaje y evaluación para crear enfoques que sean capaces de adaptarse a sistemas dinámicos complejos.
- Apropiado para manejar complejidad y no-linearidad aunque puede resultar en modelos no tan precisos.

2.1.1.5.2 Tipos

La tabla 2.5 resume los tipos de modelos que se consideran dentro de la categoría de inteligencia computacional y sus respectivas consideraciones.

Tabla 2.5: Técnicas de inteligencia computacional (Inteligencia artificial y estadística no lineal)

Tipo	Consideraciones y resultados
Redes neuronales prealimentadas	La más simple red neuronal equivalente a una regresión lineal
	Los pronósticos son obtenidos mediante una combinación lineal de las entradas del modelo.
	Los coeficientes de la de la regresión son seleccionados utilizando un algoritmo de aprendizaje que minimiza la función costo.
Redes neuronales recurrentes	Son sistemas dinámicos: Su respuesta a una entrada tiene en cuenta estados previos de la red.
Redes neuronales difusas	Generalización de la lógica Booleana. Tiene ciertos rangos cualitativos asociados (ej: Temperatura puede ser baja, media o alta).
Máquinas de soporte vectorial	Ejecuta un mapeo no lineal de data en un espacio dimensional: usa funciones lineales simples para crear límites lineales en el nuevo espacio.
	Entrega una única solución que es caracterizada por un mínimo global de la función óptima.

2.1.2 PRONÓSTICO DE PRECIOS DE LA ENERGÍA MEDIANTE MODELOS AUTORREGRESIVOS

Tabla 2.6: Revisión bibliográfica: Pronóstico de precios de la energía mediante modelos autorregresivos

No.	Título	Año	Mercado Eléctrico	Modelo	Test	Fuente
1	Forecasting electricity price volatility with the Markov-switching GARCH model: Evidence from the Nordic electric power market	2013	Nordic Electric Power Market (Países Nórdicos)	MS-GARCH; GARCH; GJR.	RMSE	[12]
2	Hourly electricity prices in day-ahead markets	2007	APX (Holanda), EEX (Alemania), PPX (Francia)	Panel	Fisher	[26]
3	Is Indian Spot Electricity Price Stationary?	2017	India	Solo tests	ADF; Phillips Peron; KPSS; Narayan-Popp	[64]
4	Long-term memory in electricity prices: Czech market evidence	2013	República Checa	DFA, Coef de Hurst	ADF; KPSS	[31]
5	On the impact of outlier filtering on the electricity price forecasting accuracy	2019	ATS Europe Ural (EU); ATS Siberia (SI); Nord Pool; PJM	ARX; SCARX + filtros	WMAE	[2]
6	Forecasting electricity prices with expert, linear, and nonlinear models	2022	Italia	58 modelos AR	RMSE, Diebold-Mariano	[6]
7	A seasonal ARIMA model with exogenous variables for elspot Electricity prices in Sweden	2013	Suecia (Elspot)	SARIMAX	MAPE	[68]

Continúa en la siguiente página

Tabla 2.6 – continua de la pagina anterior

No.	Título	Año	Mercado Eléctrico	Modelo	Test	Fuente
8	Forecasting hourly electricity prices using ARMAX-GARCH models: An application to MISO hubs	2012	MISO	GARCH, EGARCH, APARCH, CGARCH	MAE, RMSE, DFA, Philips Peron, KPSS	[25]
9	Price forecasting in the day-ahead Iberian electricity market using a conjctural variations ARIMA model	2012	Portugal y España MIBEL	ARIMA	MAPE, RMSE	[32]
10	ARIMA model to predict Next-Day electricity prices	2003	España y California	ARIMA	MWE, FMSE	[14]
11	Spot electricity price forecasting in Indian electricity market using autoregressive-GARCH models	2016	India	ARIMA-GARCH, ARIMA-EGARCH, ARIMA-PARCH	RMSE, MAE, MAPE	[22]
12	A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices	2005	España y California	GARCH	MWE, FMSE	[20]
13	Comparison of ARIMA and ANN Models Used in Electricity Price Forecasting for Power Market	2017	Reino Unido	ARIMA; ANN	RMSE	[19]
Continua en la siguiente pagina						

Tabla 2.6 – continua de la pagina anterior

No.	Título	Año	Mercado Eléctrico	Modelo	Test	Fuente
14	Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models	2010	España y PJM	GARCH y ARIMA	MAPE	[70]
15	Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models	2005	España	ARIMA	Error Semanal	[13]
16	Statistical Model Approach of Electricity Price Forecasting for Indian Electricity Market	2021	India	SARIMA, GARCH	RMSE; MAPE	[51]
17	Hourly electricity price forecasting with NARMAX	2022	ISEM (Irlanda)	NARMAX	RMSE	[38]
18	An empirical comparison of alternate regime-switching models for electricity spot prices	2010	EEX (Europa), PJM (US), NEPPOL (US)	Markov-Regime	N/A	[29]
19	An empirical comparison of alternative schemes for combining electricity spot price forecasts	2014	EEX (Europa), PJM (US), Nord Pool (Escandinavia)	6 modelos AR	RSME; Diebold-Mariano	[45]
20	Component estimation for electricity market data: Deterministic or stochastic?	2014	Italia, Nord Pool, Reino Unido, PJM	Determinístico Vs Estocástico	RMSE	[36]

Continúa en la siguiente página

Tabla 2.6 – continua de la pagina anterior

No.	Título	Año	Mercado Eléctrico	Modelo	Test	Fuente
21	Component estimation for electricity prices: Procedures and comparisons	2014	Reino Unido, PJM y Nord Pool	Estimación de componentes	MSE y MAE	[35]
22	Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark	2021	Escandinavia (NP), US (PJM), Bélgica (EPEX BE), Francia (EPEX FR), Alemania (EPEX DE)	LEAR, ARX y DNN	rMAE, MAE, MAPE, sMAPE, RMSE	[33]
23	Forecasting Next-Day Electricity Prices by Time Series Models	2002	España y California	Regresión dinámica	MEW, MSE	[41]
24	Global sensitivity analysis for a real-time electricity market forecast by a machine learning approach: A case study of Mexico	2022	Yucatán, México	Inteligencia Artificial	RMSE, MAPE	[15]
25	Modeling energy price dynamics: GARCH versus stochastic volatility	2016	US - Europe (Crudo), US (Productos del petróleo), US (Gas Natural)	GARCH y volatilidad estocástica	N/A	[11]
26	On the importance of the long-term seasonal component in day-ahead electricity price forecasting	2016	Escandinavia (NP), GEFCom2014	SCARX y mSCARX	WMAE y Diebold-Mariano	[43]

Continúa en la siguiente pagina

Tabla 2.6 – continua de la pagina anterior

No.	Título	Año	Mercado Eléctrico	Modelo	Test	Fuente
27	Recent advances in electricity price forecasting: A review of probabilistic forecasting	2018	GEFCom2014	AR, ARIMA, redes neuronales	N/A	[44]
28	Robust estimation and forecasting of the long-term seasonal component of electricity spot prices	2013	Australia, Alemania, Noruega, US	304 modelos AR	MAE, MSE, MAPE	[42]
29	Short- to Mid-term Day-Ahead Electricity Price Forecasting Using Futures	2019	Alemania y Austria	AR	MAE, MMAE	[63]

2.2 ANTECEDENTES

La presente sección pretende introducir la historia del sector eléctrico mexicano, así como las transformaciones y capacidades del sistema desde su creación hasta la actualidad. Para ello se va a introducir de forma cronológica su evolución y se analizará el contexto en que estas acciones de cambio ocurrieron.

2.2.1 LOS INICIOS (1879-1936)

A finales del siglo XIX, el mundo se encontraba en constante evolución y transformación. Los desarrollos logrados por Thomas Alba Edison y Nikola Tesla lo demuestran, ya que permitieron la transmisión de la electricidad a través de largas distancias, y con esto marcaron un hito histórico que introdujo la forma en la que la energía eléctrica es distribuida y usada alrededor del mundo hoy en día. Una muestra de esta evolución puede verse con la fundación de la Edison Illuminating Company en 1879 y la inauguración de la primera central para venta de electricidad en 1882. Es importante resaltar que el uso de electricidad se puede datar desde mucho antes, pero solamente para usos puntuales en los cuales la electricidad se generaba para aplicaciones específicas. No obstante, los desarrollos aquí mencionados se refieren a la generación y distribución de electricidad para su venta y con esto la creación del concepto de mercado eléctrico.

En el caso de México, [8] mencionan que el primer uso de la electricidad en el país se encuentra en el año de 1879, cuando se usó en una fábrica textil de Hayser y Portillo en la ciudad de León, Guanajuato. Al poco tiempo, las principales ciudades registraban uso de la electricidad, principalmente para la iluminación, aunque no se limitaba solo ello; la introducción de este servicio se vio en dichas ciudades durante la década de los 80s, así: Ciudad de México en 1880, Guadalajara en 1884, Monterrey en 1888 y Veracruz y Mérida en 1889. Durante este periodo, el gobierno de Porfirio Díaz impulsó una política de atracción de inversiones, principalmente provenientes de Estados Unidos y Europa, que buscaba la modernización del país, y con esto del sector eléctrico nacional. Esto ayudó a que el uso de la electricidad se extendiera a la minería, la manufactura y el transporte.

Este periodo que culminó en 1910, se caracterizó por un incremento en la demanda de la electricidad por parte de talleres e industrias que no tenían la capacidad para generar su propia electricidad, y se extendió a la atención del sector servicios (iluminación, tranvías, transporte agua). Esto llevó a la aparición de una basta cantidad de empresas generadoras. De acuerdo con [53], para 1899 existían 174 plantas de vapor y tres hidroeléctricas que atendían los sectores mencionados anteriormente.

Para cumplir con estas demandas de electricidad se empezaron a fundar empresas como la Mexican Gas and Electric Company, la American and Foreign Power Company o la Compañía Explotadora de las Fuerzas Hidroeléctricas de San Ildefonso . Sin embargo, la más importante de ellas fue la Mexican Light and Power Company (MLPC) fundada en 1903 y que merece mención especial en este apartado.

2.2.1.1 Mexican Light and Power Company (MLPC)

La compañía se funda en 1903 durante el gobierno de Porfirio Díaz en la ciudad de Toronto, Canada, teniendo como socio mayoritario al ingeniero Frederick Stark Pearson, para desarrollar un proyecto hidroeléctrico en la región de Necaxa, Puebla y su energía ser usada en la Ciudad de México. En el contrato se estipulaba que la compañía debía generar 5 MW durante sus primeros 10 años, y después de este tiempo debía duplicar su capacidad. En 1905 se enciende la primera unidad de 5 MW, hecho sin precedentes pues anteriormente una concesión francesa había intentado hacer lo mismo durante cinco años, sin resultados positivos; en 1906 había triplicado su capacidad de generación y en 1907 podía generar cerca de 30 MW cumpliendo ampliamente con lo estipulado en el contrato y brindando electricidad a la Ciudad de México, y los Estados de Puebla, México y Michoacan.

Durante este quinquenio la compañía ganó la concesión para el aprovechamiento de más ríos, y por tanto aumentar la potencia de generación a través de plantas hidroeléctricas. En 1911, la Secretaría de Estado y del Despacho de Fomento, Colonización e Industria le concedió una extensión de la concesión por 99 años ([53]). Con el incremento en la capacidad de generación y, de acuerdo con [53], la compañía pudo reducir los costos de operación en un 60 % y adquirir a la competencia.

Entre los datos curiosos, [53] menciona que, en 1909, el complejo hidroeléctrico de Necaxa estuvo considerado como uno de los ocho mas importantes del mundo; registró el récord de mayor distancia de transmisión de energía eléctrica, siendo éste de 278 km, divididos en 156 km a la Ciudad de México más otro tramo de 122 km a la región minera de El Oro (Estado de México), superando los 200 km que existen entre Niágara y Toronto.

No obstante, con la irrupción de la Revolución de 1910, las empresas fueron objeto de préstamos obligatorios y un aumento de los impuestos para financiar los ejércitos en pugna. Esto, sumado a la I Guerra Mundial, ocasionó una precarización de la condiciones laborales que, en 1914, conllevó a la fundación del Sindicato Mexicano de Electricistas - SME, al que también se afiliaron trabajadores de los telégrafos de México. En 1915,

Pearson fallece a causa del hundimiento del *RMS Lusitana* cuando se encontraba rumbo al Reino Unido durante la primera guerra mundial.

2.2.1.2 Primeros intentos de regulación

Es importante señalar, que durante estos primeros años no existía una institución que guiara la política respecto a la expansión de la red eléctrica nacional; por el contrario, cada empresa decidía donde quería aumentar su cobertura y a qué precio. Esto propició que los sectores lejanos a los grandes poblados urbanos o que no contaran con presencia de industrias no tuvieran acceso a la electricidad.

Con la introducción de la nueva constitución, y su artículo 27, ([47]), se comienza a gestar lo que sería la base para una legislación en materia eléctrica; en términos generales, la constitución considera al agua como propiedad y de manejo directo de la nación que puede ser concesionado a particulares para su aprovechamiento.

Como un primer intento de regulación, en 1923, se crea la *Comisión de Fomento y Control de la industria de Generación de Fuerza*¹ que, según [47], tenía entre sus funciones:

1. La conservación y mejor aprovechamiento de los recursos nacionales en materia de fuerza hidráulica.
2. La protección de nuevas empresas para la explotación de la fuerza motriz.
3. Establecer restricciones que convengan.
4. Establecer tarifas.

En este punto es importante resaltar, que este fue un periodo de consolidación de monopolios y en el que, a nivel gobierno, no existía un órgano centralizado que dirigiera la política energética; cada empresa decidía qué tarifas y qué política de inversiones desarrollaba. Debido a que no contaba con atribuciones regulatorias, no hubo cambios relevantes en la situación que hasta entonces se tenía.

Posteriormente en 1926, como resultado de la reestructuración de la entidad creada en 1923, se crea la *Comisión Nacional de Fuerza Motriz*, a la cual se le asigna principalmente la tarea de regular las tarifas de las empresas eléctricas. Además, para este

¹Decreto Presidencial, DOF, 2 de junio de 1923

mismo año, se publica el Código Nacional Eléctrico que da las definiciones técnicas para la generación y el uso de la energía eléctrica.

Como contexto general, a finales de los años 20, se da la caída de la Bolsa de Nueva York, que trajo consigo la Gran Depresión de los años 30; en México afectó al sector eléctrico en particular ([8]), llevando a la quiebra a muchas pequeñas generadoras eléctricas, y afectando a los consumidores que vieron como las tarifas se elevaban.

Esta es la principal causa del proceso de monopolización por parte de los grandes generadores, que además, se vio estimulado por un gobierno que, en un principio, no se interesó por regular el sector y que, por lo tanto, otorgó poder para que dichas empresas decidieran libremente las tarifas y las zonas de expansión de la red eléctrica. Esta situación trajo consigo muchas presiones para el gobierno, por parte de los industriales y el público en general, provocando que, a inicios de 1934 se le otorgaran facultades al Congreso de la Unión para legislar en temas eléctricos incluyendo temas de impuestos y la facultad del gobierno federal para legislar en materia eléctrica.

2.2.2 CO-REGULACIÓN (1937-1959)

Este periodo inicia con aires nacionalistas, tendencia que, para la época, se vivía en todo el mundo, incluyendo a México con el gobierno de Lázaro Cárdenas. De acuerdo con [47], las tensiones entre comercializadores y consumidores de la energía eléctrica hizo que estos últimos, desde finales de 1932, se agruparan en la Confederación Nacional de los Servicios Públicos (CNDSP).

Como lo expone [47], se buscaba *“un sistema nacional de generación, transmisión y distribución de energía eléctrica, integrado por empresas semi-oficiales y cooperativas de consumidores”*. Igualmente la respuesta del gobierno a estas presiones fue la creación de un decreto², en el que se limitaban las nuevas concesiones a extranjeros, incluyendo normas como aquellas sociedades mixtas, donde el socio mexicano debería poseer la mayor parte y el extranjero renunciar a su nacionalidad.

2.2.2.1 La CFE (Comisión Federal de Electricidad)

La Comisión Federal de Electricidad (CFE) se funda el 24 de agosto de 1937, como una dependencia oficial, sin ánimo de lucro. De acuerdo con [47], la CFE estaría presidida por el secretario de economía y dos vocales, y tendría dentro de sus facultades las siguientes responsabilidades:

²Decreto Presidencial, DOF, 18 de enero de 1937

- Planeación del sistema nacional de electrificación.
- Financiamiento de los proyectos que resultaran de la planeación. Incluyendo la adquisición de bienes muebles e inmuebles.
- Organizar sociedades con el objetivo de generar, transmitir y distribuir energía eléctrica.
- Fabricar equipos y materiales que se requirieran para cumplir sus objetivos.
- Prioridad sobre los particulares para la expropiación de los bienes que se consideraran de utilidad pública.

Dos años después de su fundación, entró en rigor la ley del Impuesto sobre consumo de energía eléctrica , cuya finalidad última consistía en hacer el recaudo necesario para que la CFE tuviera el músculo financiero para promover el mejoramiento y la construcción del sistema de electrificación.

De este modo, para finales de la década de los 30s existían 3 monopolios dominantes ([30]) :

- Mexican Light and Power Company-MPLC.
- American Foreign Power Company.
- Compañía Hidroeléctrica e Irrigadora de Chapala.

En 1940, esta última se declara en bancarrota y es comprada por el Estado a través de la CFE.

2.2.2.2 La ley de la industria eléctrica de 1939 y sus efectos

Esta ley es promulgada el 11 de febrero de 1939 de acuerdo con [47]. Entre sus aspectos más relevantes se dictan facultades para modificar los términos de las concesiones y establecer las bases de la fijación de tarifas. Esta ley estuvo vigente hasta la nacionalización del sector en 1960.

Dentro de las particularidades de esta ley, se encuentra la forma como se consideraba la estructura de la industria eléctrica, donde se encuentra que la generación, el transporte, la distribución, la exportación, la importación, la compra-venta, la utilización y el consumo de energía eléctrica hacen parte del sistema. En el ámbito de las tarifas, éstas quedaron a

cargo de la Secretaria de Energía (SE), y por consiguiente, la venta de energía solamente se podía realizar de conformidad a las tarifas autorizadas por SE.

Por último, esta ley obligaba a los generadores de electricidad que utilizaban recursos naturales (i.e hidroeléctrica) a solicitar la concesión para el aprovechamiento de dichos recursos.

2.2.3 INTEGRACIÓN VERTICAL (1960-1991)

Este periodo se caracteriza por el impulso que recibió la política de nacionalización de las pequeñas compañías privadas que finalizó en 1972 con la Compañía de Servicios Públicos de Nogales tal como lo comenta [30]. De acuerdo con [8] para 1960 CFE tenía el 54 % de participación en la capacidad instalada en el país (2308 MW), mientras que la Compañía Mexicana de Luz y Fuerza del Centro tenía el 25 %, The American & Foreign Power Company el 12 % y el resto de compañías el 9 %. El proceso es completado con la declaración de la *Ley de Servicios Públicos de Energía Eléctrica* en 1975. De acuerdo con [30] la ley estableció y designó a la CFE y a la Compañía de Luz y Fuerza del Centro [LyFC], que operó en el centro de México, como los proveedores de la energía eléctrica. Estas dos compañías eran dueñas de las plantas generadoras y de las líneas de transmisión y distribución.

Esto sirvió para la unificación de tarifas, pues se crea el sistema interconectado nacional que cubre todo el territorio nacional a excepción de la península de Baja California. Su hito mas importante es la unificación de las frecuencias para el servicio público en los 60 Hz en el año 1976. De acuerdo con [8], para dicha época, se expande la cobertura del sistema eléctrico mediante la diversificación de las fuentes generadoras y se empieza a tener una creciente preocupación por los temas ambientales.

Dentro de este periodo, el acceso a la energía eléctrica pasó de un 44 % a un 88 % ([8]), debido al lanzamiento de programas de electrificación nacional, especialmente en las zonas rurales; igualmente, durante este periodo se crea una política social enfocada en subsidiar el precio de la energía, tanto para los hogares como para los agricultores. Esto se logró principalmente porque, anterior a la nacionalización, la expansión del servicio estaba condicionada a que la demanda en la zona a intervenir fuera rentable para la empresa generadora, promoviendo la aparición de varios sistemas eléctricos independientes.

Como ejemplo de lo anterior, para 1962, existían 168 diferentes tarifas eléctricas, que fueron unificadas a 11 tarifas que se mantuvieron estables con una fuerte carga de subsidios ([30]). Para 1973, se crearon dos nuevas tarifas, mientras que las que, desde 1962 habían permanecido estables, empezaron a ser ajustadas.

El gobierno también trabajó fuertemente en el desarrollo de tecnología para el sector eléctrico, y en 1975 crea el Instituto de Investigaciones Eléctricas; durante estos años la creación de nuevas plantas se enfocó en las hidroeléctricas y las termoeléctricas dependientes del petróleo. La década de los 70's de acuerdo con [52] fue caracterizada por la puesta en marcha de gran cantidad de proyectos hidroeléctricos.

A pesar de que la matriz energética dependía fuertemente del petróleo, CFE empezó a buscar nuevas fuentes de energía, y creó la primera planta de carbón de 37.5[MW] en el estado de Coahuila que operó hasta 1978. También, se exploró la opción de plantas geotérmicas a inicios de 1960 con una planta de 3.5[MW]. Finalmente, para 1971 se tomó la decisión de construir la primera central nuclear que debido a varios inconvenientes, tanto internos como externos su construcción y operación se retrasó hasta 1990. Debido a la crisis energética de 1982, se buscó explorar la generación de electricidad mediante plantas a turbogas.

Si tenemos que definir a este periodo se puede decir sin lugar a equivocaciones, que fue un periodo de total control estatal en la generación, transmisión y distribución, un modelo de gobernanza jerárquica vertical. Es importante resaltar que durante este periodo el sector eléctrico mexicano (CFE y LyFC) y la expansión consecuente fue financiado con recursos federales, préstamos bilaterales, préstamos del Banco Mundial, del Banco Interamericano de Desarrollo así como créditos de proveedores. Debido a los aprietos económicos y la recesión de la década de los 80, se generó una crisis en CFE y LyFC que las obligó a buscar nuevas fuentes de financiamiento, tanto para modernizar equipos como para nuevos proyectos de generación, a inicios de 1990 las tarifas de energía fueron redefinidas e incrementadas y se empezó un desmonte gradual de los subsidios de energía que había caracterizado estos últimos 30 años.

2.2.4 INICIO DE LA APERTURA (1992-2012)

Este nuevo periodo inicia con la liberalización del sector eléctrico durante la administración de Carlos Salinas de Gortari (1988-1994). En ella se buscaba maximizar la eficiencia y reducir los gastos; dentro de los acuerdos de NAFTA y GATT (Acuerdo General sobre Aranceles Aduaneros y Comercio) los subsidios continuaron a ser gradualmente desmontados. Esta liberalización permitió a los Generadores Independientes Privados (IPP) participar en la generación de electricidad bajo 6 nuevas modalidades; mientras que CFE y LyFC seguían siendo las únicas autorizadas para transmitir y distribuir energía. Básicamente lo que se hizo fue permitir a los IPP producir y vender la energía a CFE bajo Acuerdos de Compra de Potencia (PPA), de acuerdo con ([8]) se les dio el nombre de Generadores Pequeños de Potencia (SPP), aquellos que estuvieran ubicados en comunidades

rurales con menos de 1[MW] de generación y que se usara para auto consumo, plantas generadoras con menos de 30[MW] y que vendieran a CFE y a los que exportaran hasta [30MWh]; igualmente la exportación e importación de energía fue permitida solamente para cubrir necesidades específicas. En este periodo (1993) se crea la Comisión Reguladora de Energía (CRE), debido al alto potencial de nuevos participantes en el sector eléctrico.

De acuerdo con [5] en noviembre de 1994, no pasar por alto que en ese mismo año México paso a ser parte de la Organización para la Cooperación y Desarrollo Económico (OECD), que también se conoce como el club de las buenas prácticas. En sintonía con esta adhesión se establecieron nuevas regulaciones ambientales que se hicieron efectivas en enero de 1998, en ellas se designaron nuevos límites para la emisión de NO_x y SO_x . Como dato interesante de acuerdo con [5] en el periodo de 1970 a 1995 las emisiones de carbono se incrementaron en un 221 %. Para 1995 se calculó que México producía 95 millones de toneladas métricas, convirtiéndolo en en undécimo país con mas emisiones en el mundo. El principal afectado con estas nuevas leyes fue la CFE, a la cual se le impidió seguir generando electricidad a base de petróleo con alto contenido de sulfuros. Esto puso gran presión económica a la CFE para modernizar las plantas, así como una reconversión hacia el gas natural.

Para mayo de 1995 el Congreso Mexicano aprobó una enmienda a la Ley Constitucional, artículo 27 sobre el petróleo. En ella se permitió que operadores privados ya fueran nacionales o extranjeros a importar, distribuir y almacenar gas natural. Todo esto con el fin de asegurar el suministro de energía.

Como un primer intento privatizador del sector, en 1999 se presentó al senado ([30]) una propuesta para privatizar la generación y la distribución, manteniendo la transmisión y la operación del sistema bajo el control gubernamental propiamente la CRE que regularía la transmisión y distribución, garantizando la competencia efectiva en el mercado eléctrico. Con el proyecto se proponía vender todas las plantas de generación a excepción de las nucleares y algunas hídricas. Esta reforma no logró llegar a buen término.

El cambio de milenio llegó trajo consigo el fin del Partido Revolucionario Institucional (PRI) como partido de gobierno. El turno para nuevas reformas fue de Vicente Fox, que propuso una reforma similar a la de 1999. En ella se buscaba que la CRE tuviera un rol mas importante, por ejemplo, establecer precios, y redefinir los subsidios. Esta propuesta tuvo el mismo fin de la anterior y no fue aprobada. Sin embargo en 2002, las tarifas eléctricas fueron revisadas a la baja y eliminada para los consumidores de altos ingresos ([30]). En el 2008 durante el gobierno de Felipe Calderon (2006-2012), se intentó una nueva reforma, en esta ocasión con igual resultado, aunque algunos cambios mínimos

fueron aprobados, principalmente en el sector petrolífero, y especial dos leyes que fueron transversales a todo el sector energético.

- Ley para el uso sustentable de la energía.
- Ley para el aprovechamiento de energías renovables y el financiamiento de la transición energética.

En términos de eficiencia energética, el uso de electricidad por unidad de PIB [GWh/PIB] se incrementó después de la crisis del peso mexicano, hasta 1995, fecha en la que comenzó una tendencia decreciente (1995-2012)([8]). Durante este periodo se disolvió la empresa estatal LyFC, debido a su ineficiencia administrativa y operacional ([30]), transfiriendo sus responsabilidades a CFE.

En el periodo 2002 al 2009 los generadores independientes (IPP) representaban el 31 % de la capacidad instalada de CFE y el 45 % de la generación, autorizándose 24 permisos. Durante este periodo, a los Auto-generadores (PSS) también adquirieron permisos siendo en total 25 para generar electricidad, los principales sectores fueron, petrolero, cementero, acerías, azucarero y minería, llegando a ser 48 % de la capacidad instalada y 47 % de la generación. En el caso de los Co-generadores (CPP), donde se aprobaron 22 permisos ([52]) siendo la industria petrolera la mayor beneficiada ([30]). Durante este periodo se expidieron 36 permisos de exportación/importación de energía eléctrica la mayoría localizados en los estados de Sonora y Baja California por su cercanía con USA ([52]).

Respecto a la tecnología, se inicio una transición hacia el gas natural, por considerarse más limpia; igualmente se iniciaron proyectos eólicos y fotovoltaicos. Se instalaron proyectos eólicos a corte 2012 con una capacidad de 598 MW, siendo 511 MW propiedad de IPPs y el primer proyecto fotovoltaico de 5 MW (propiedad de CFE) inició operación. Es importante aclarar que los otros métodos de generación se mantuvieron estables.

Por otra parte, el objetivo de los gobiernos de esta época fueron principalmente dos, diversificar la matriz energética, aunque como se menciona anteriormente el gas natural pasó de representar el 15 % de la demanda en las termoeléctricas en 1993 a ser el 47 % en 2012, buscando garantizar el suministro eléctrico por medio de la inversión de privados en la generación y/o la importación de electricidad desde USA. El segundo gran objetivo fue incrementar el acceso a la electricidad en las zonas rurales. Esta última alcanzó en el 2012 un 99 % de cobertura.

Finalmente, en 2012 se promulga la Ley General de Cambio Climático (LGCC) la cual buscaba adaptar la legislación del país a los compromisos adquiridos mediante

convenios internacionales. En un ambiente en donde la producción de petróleo disminuía, y los descubrimientos de reservas de gas natural y petróleo no convencional, el gobierno de Enrique Peña Nieto (2012-2018) propuso en el 2013 una reforma energética que fue aprobada por el congreso. Esta reforma se puede definir en los siguientes puntos,

- Permitía a inversores privados participar en la exploración, producción, refinamiento, transporte, distribución y almacenamiento de hidrocarburos y sus derivados.
- El estado mantenía el control de PEMEX y CFE.
- En el sector eléctrico se permitía a inversores privados en la generación y comercialización de la misma. La transmisión y distribución seguía bajo monopolio estatal.
- CENACE (Centro Nacional de Control de Energía) se convirtió en el operador del sistema energético.
- Por último, la reforma buscó promover el uso de la energía renovable.

2.3 MARCO TEÓRICO

2.3.1 SISTEMA ELÉCTRICO NACIONAL (SEN)

El Sistema Eléctrico Nacional (SEN) es un sistema integrado que da servicio a 128 millones de personas que habitan dentro del territorio mexicano y que busca proveer la energía eléctrica necesaria para la realización de las actividades cotidianas durante las 24 horas del día. Para esto, el SEN debe garantizar confiabilidad, calidad y eficiencia en el suministro del servicio, ya que es un actor vital en el bienestar económico y social del país. ([59, 10])

2.3.1.1 Infraestructura del Sistema Eléctrico Nacional

2.3.1.1.1 Conformación Actual

La Secretaría de Energía (SENER) presenta una descripción detallada de la estructura del SEN en su Programa de Desarrollo del Sistema Eléctrico Nacional 2020-2034 ([59]). Dicho documento presenta al SEN como un sistema conformado por nueve (9) regiones de control y un pequeño sistema eléctrico aislado. (Ver Figura 2.1)



Figura 2.1: Regiones del Sistema Eléctrico Nacional (SEN) ([59])

Estas regiones son controladas por sus respectivos Centros de Control Regional ubicados en las Ciudades de México, Puebla, Guadalajara, Mérida, Hermosillo, Gómez Palacio, Monterrey, Mexicali, La Paz y Santa Rosalía Baja California Sur. Además, existen dos Centros Nacionales ubicados en la Ciudades de México y Puebla quienes se encargan de coordinar el Mercado Eléctrico Mayorista (MEM) y la operación del SEN.

De modo similar, el Sistema Interconectado Nacional (SIN), como componente fundamental del SEN, divide el país en siete (7) regiones que comparten entre ellas recursos y reservas de capacidad que permiten suplir la demanda de energía a nivel nacional de la forma más confiable y económica.

Por otro lado, el sistema de Baja California opera interconectado a la región oeste de los Estados Unidos, mientras que los sistemas de Baja California Sur y Mulegé están eléctricamente aislados.

2.3.1.1.2 Capacidad de Transmisión del SEN

Desarrollado por la Comisión Nacional de Electricidad (CFE), el SEN distribuye energía en diferentes niveles de tensión, ver 2.2:

- **Red Nacional de Transmisión (RNT):** Distribuye a las Redes de Distribución General (RDG) y a otros usuarios manejando tensiones iguales o mayores a 69 kV.

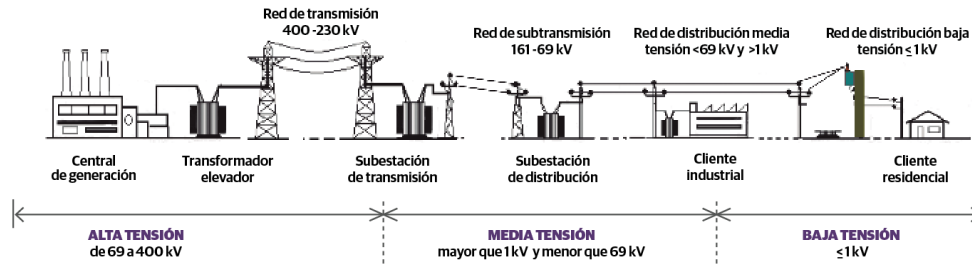


Figura 2.2: Sistema de suministro eléctrico - Elaborado por SENER con información de CENACE ([59])

- **Red de Distribución General (RDG):** Distribuye energía eléctrica al público en general a tensiones menores o iguales a 35 kV.

El mismo reporte de la Secretaría de Energía hace mención que, para diciembre de 2019, la RNT tenía 108,908 km de líneas que se han venido expandiendo y desarrollando consistentemente con la creciente demanda. Lo mismo se ha evidenciado con la expansión de capacidad en Subestaciones Eléctricas e infraestructura de distribución en general.

2.3.1.1.3 Enlaces Internacionales

En el mapa de la Figura 2.3, elaborado por la SENER con información del CENACE, se observan los principales enlaces internacionales y sus respectivas capacidades.

2.3.1.1.4 Centrales Eléctricas del Mercado Eléctrico Mayorista (MEM)

La tabla 2.7 presenta la capacidad instalada de la CFE y otros permisionarios por tipo de tecnología en MW a octubre de 2020.

Del mismo modo, las Figuras 2.4 y 2.5 muestran la distribución geográfica de las principales centrales eléctricas del Mercado Eléctrico Mayorista (MEM), por tipo de tecnología, divididas entre aquellas de la CFE y los Productores Independientes de Energía Eléctrica (PIE) y, aquellas que son propiedad de privados.

A la fecha en mención, se puede observar que las tecnologías de generación de energía eléctrica de la CFE y de los PIE se basan en un matriz principalmente térmica; mientras que la inversión privada se ha enfocado en el desarrollo de tecnologías limpias renovables que han dinamizado la matriz energética del país.

De hecho, se observa que para 2020 un 75 % de la generación total de energía eléctrica en México provenía de fuentes fósiles haciendo que el país sea altamente dependiente de



Figura 2.3: Capacidad de las interconexiones internacionales en 2019 del SEN - Elaborado por SENER con información de CENACE ([59])

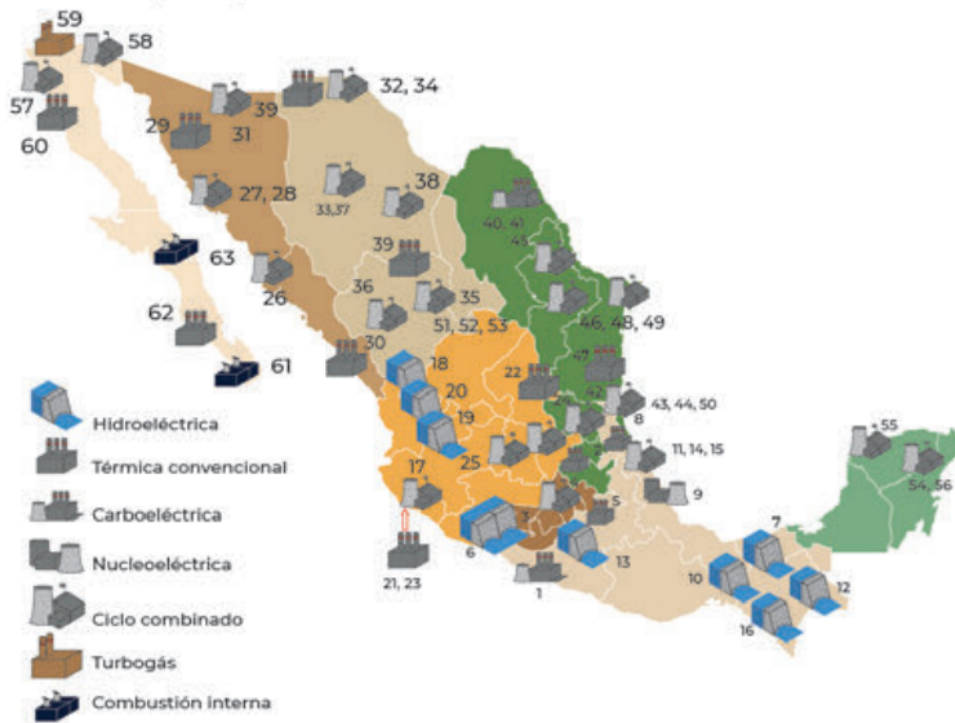


Figura 2.4: Principales Centrales Eléctricas de la CFE y PIE al 31 de diciembre de 2019 ([59])

Capacidad Instalada - Limpia Renovable	
Tecnología	2020 (MW)
Hidroeléctrica	12,612
Geotermoeléctrica	951
Eoloeléctrica	6,977
Fotovoltaica	5,795
Bioenergía	408
Limpia Renovable	26,743

Capacidad Instalada - Limpia No Renovable	
Tecnología	2020 (MW)
Nucleoeléctrica	1,608
Cogeneración Eficiente	1,906
Frenos Regenerativos	0
Limpia No Renovable	3,514

Capacidad Instalada - Origen Fósil	
Tecnología	2020 (MW)
Ciclo Combinado	33,746
Térmica Convencional	11,831
Turbogás	3,793
Combustión Interna	943
Carboeléctrica	5,464
Origen Fósil	55,776

TOTAL: 86,034 MW

Tabla 2.7: Capacidad instalada de la CFE y del resto de permisionarios (MW) - Elaboración de la SENER con información de CENACE ([59])

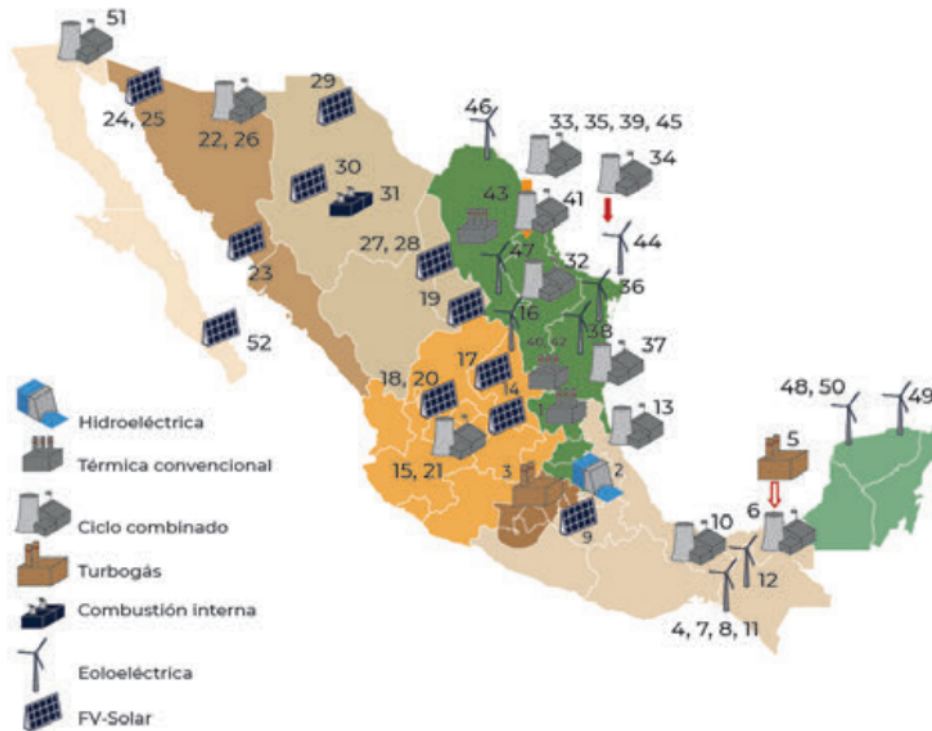


Figura 2.5: Principales Centrales Eléctricas privadas al 31 de diciembre de 2019 ([59])

estas fuentes; sin embargo, se observa que el uso de carbón para estos fines ha caído en un 60 % en comparación con 2015, lo cual ha permitido la entrada de fuentes limpias y renovables y un aporte del 10 % de la energía solar y eólica. No obstante, algunos analistas mencionan que los progresos recientes en la introducción de energías limpias renovables se pueden ver en peligro por un posible cambio en la política energética promovida por el gobierno actual ([18].)

2.3.1.2 Demanda y Consumo de Energía Eléctrica

La SENER presenta en su Programa de Desarrollo del Sistema Eléctrico Nacional ([59]) el pronóstico de la demanda y consumo de energía eléctrica de 2020 a 2034, que busca apoyar la planeación y guiar la toma de decisiones en lo que respecta al crecimiento, modernización y desarrollo de los diferentes componentes del SEN.

Dicho pronóstico de demanda depende de diferentes factores:

- Crecimiento económico
- Crecimiento poblacional
- Estacionalidad

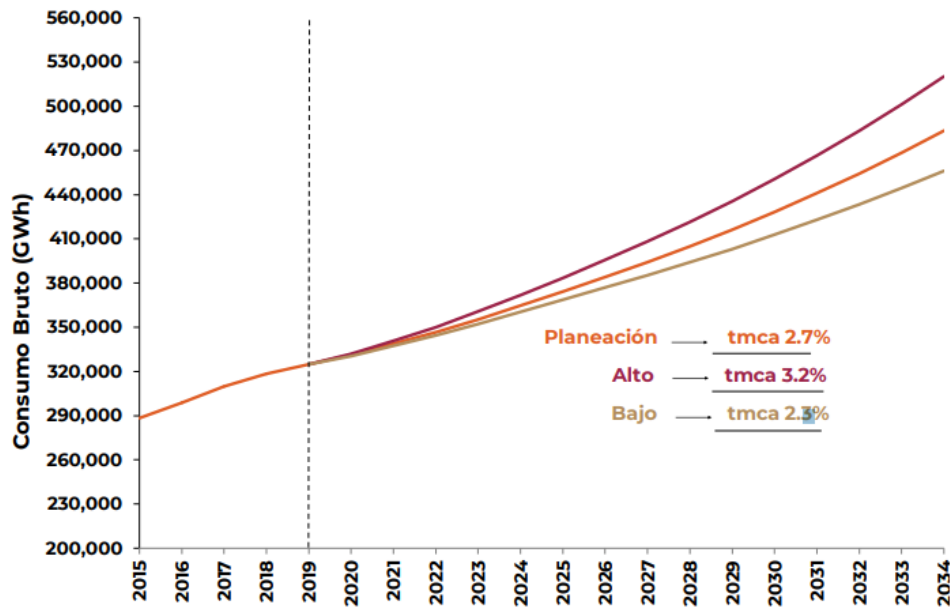


Figura 2.6: Pronóstico del Consumo Bruto del SEN 2020 - 2034 en los escenarios de Planeación, Alto y Bajo - Elaborado por SENER con información de CENACE ([59])

- Precios de combustibles
- Precio de la energía eléctrica
- Pérdidas de energía eléctrica
- Eficiencia energética
- Generación distribuida
- Electromovilidad
- Estructura de consumo final eléctrico

Y se diseña para tres diferentes escenarios de consumo que difieren en la Tasa media de crecimiento anual (tmca) que consideran: 1) Planeación (tmca: 2.7%), 2) Alto (tmca: 3.2%) y 3) Bajo (tmca: 2.3%).

Con base en la anterior, la SENER presenta tres posibles pronósticos de consumo bruto del SEN ente 2020 y 2034 (ver Figura 2.6)

Sin embargo, la misma fuente ([59]) hace un análisis de la afectación de la contingencia sanitaria debido al COVID 19 que afectó los tmca en los tres escenarios considerados y por ende los consumos brutos de energía que se estimaban inicialmente. La Figura 2.7 muestra el efecto que tuvo la contingencia sanitaria en el consumo bruto de energía en

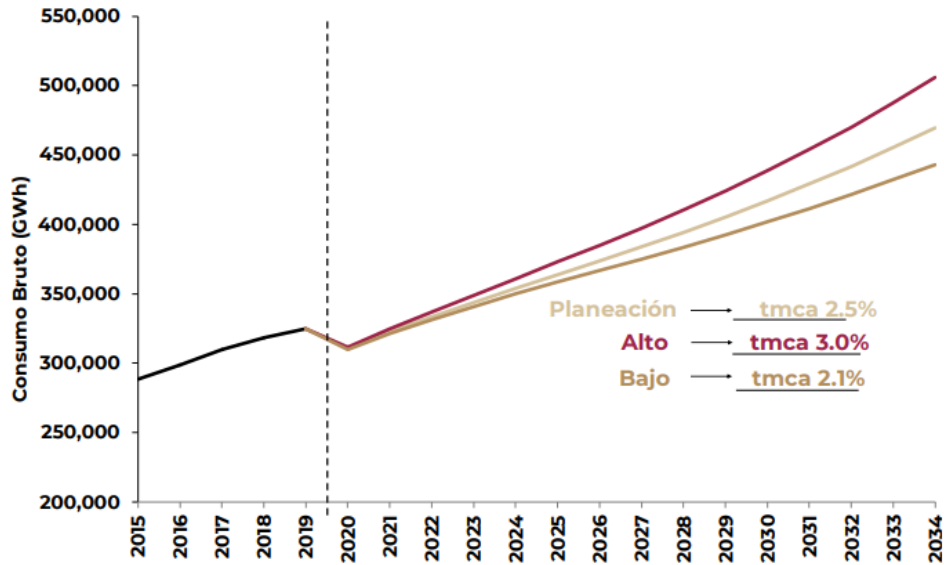


Figura 2.7: Pronóstico del Consumo Bruto del SEN 2020 - 2034, contingencia sanitaria, en los escenarios de Planeación, Alto y Bajo - Elaborado por SENER con información de CENACE ([59])

2020, en los tcma y por ende, en el pronóstico de consumo bruto esperado para los años posteriores hasta 2034.

Para explicar lo anterior, es importante tener en cuenta que la contingencia sanitaria ocasionada por el COVID 19, trajo consigo una serie de medidas que impactaron de forma sustancial al sector energético debido a la disminución de las actividades comerciales e industriales que llevaron consigo un decremento relevante en la demanda y consumo de energía, no solo en México sino también en la mayoría de países del mundo. Esto, sumado a la ralentización de las cadenas productivas a nivel global, la gran incertidumbre generada y la volatilidad de los combustibles primarios implicó un menor crecimiento económico durante 2020 y consecuentemente, un pronóstico menos optimista.

Ahora bien, un ejercicio similar se llevó a cabo para pronosticar la demanda máxima en el mismo periodo de tiempo; para este caso en particular, la tmca se determinó para los mismos tres escenarios así: 1) Planeación (tmca: 2.8%), 2) Alto (tmca: 3.3%) y 3) Bajo (tmca: 2.4%). Arrojando los resultados presentados en la Figura 2.8

Del mismo modo, [59] establece un pronóstico de la demanda máxima 2020 - 2034 teniendo en cuenta la contingencia sanitaria y bajo supuestos de tema consecuentes con la situación descrita.

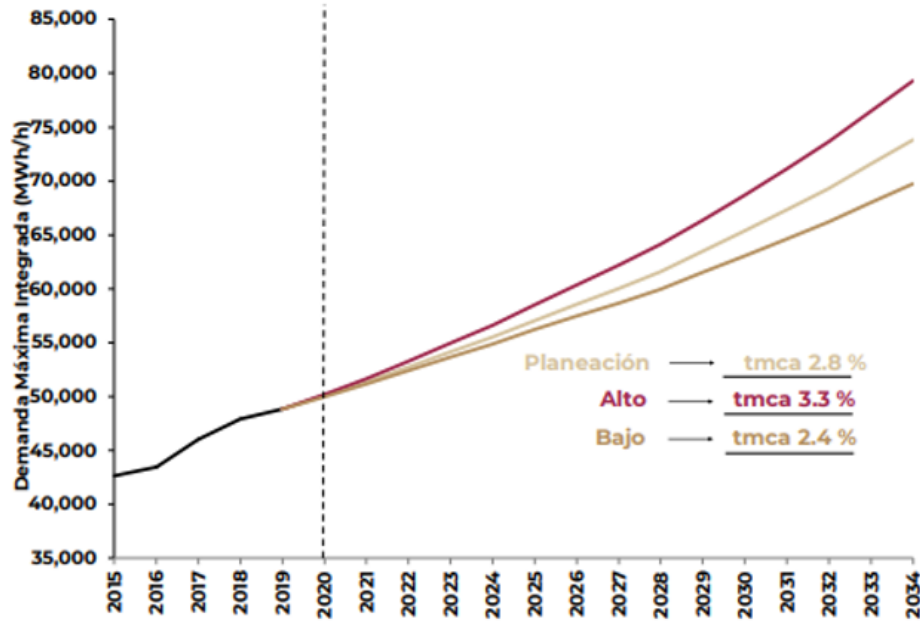


Figura 2.8: Pronóstico de la Demanda Máxima Integrada del SEN 2020 - 2034, escenarios de Planeación, Alto y Bajo - Elaborado por SENER con información de CENACE ([59])

2.3.1.3 Funcionamiento del Mercado Eléctrico Mayorista (MEM)

El Mercado Eléctrico Mayorista (MEM) de México se creó con la reforma energética de 2013 y el cambio estructural del SEN; esta reforma buscaba migrar de un esquema centralizado y controlado por la CFE a una estructura de mercado competitivo, en el que fuera posible la interacción de distintos jugadores que ofertan y/o demandan. Este nuevo esquema estaría regulado por la Comisión Reguladora de Energía (CRE) y el Centro Nacional de Control de Energía (CENACE). [54]

Además del mercado de comercialización de energía eléctrica per se, el MEM incluye otros cuatro mercados entre los cuales se encuentran el de Certificados de Energía Limpia y Potencia. En la figura 2.9 se presentan dichos mercados y posteriormente se da una breve explicación de cada uno de ellos de acuerdo con [54].

- **Mercado de energía de corto plazo:** Los participantes de este mercado realizan ofertas de compra y venta de energía buscando ofrecer la cantidad de energía suficiente a un costo competitivo. Los Usuarios Calificados ³ y sus Suministradores ⁴, determinan la demanda, mientras que la oferta vendría determinada por los

³**Usuario Calificado:** Usuario final con una demanda de al menos 1 MW y que se ha registrado como tal ante la CRE con el fin de obtener energía eléctrica por parte de suministrador calificado

⁴**Suministrador de Servicio Calificado** Se trata de un proveedor y comercializador de energía eléctrica, que compra electricidad en el MEM para brindar sus servicios a los Usuarios Calificados

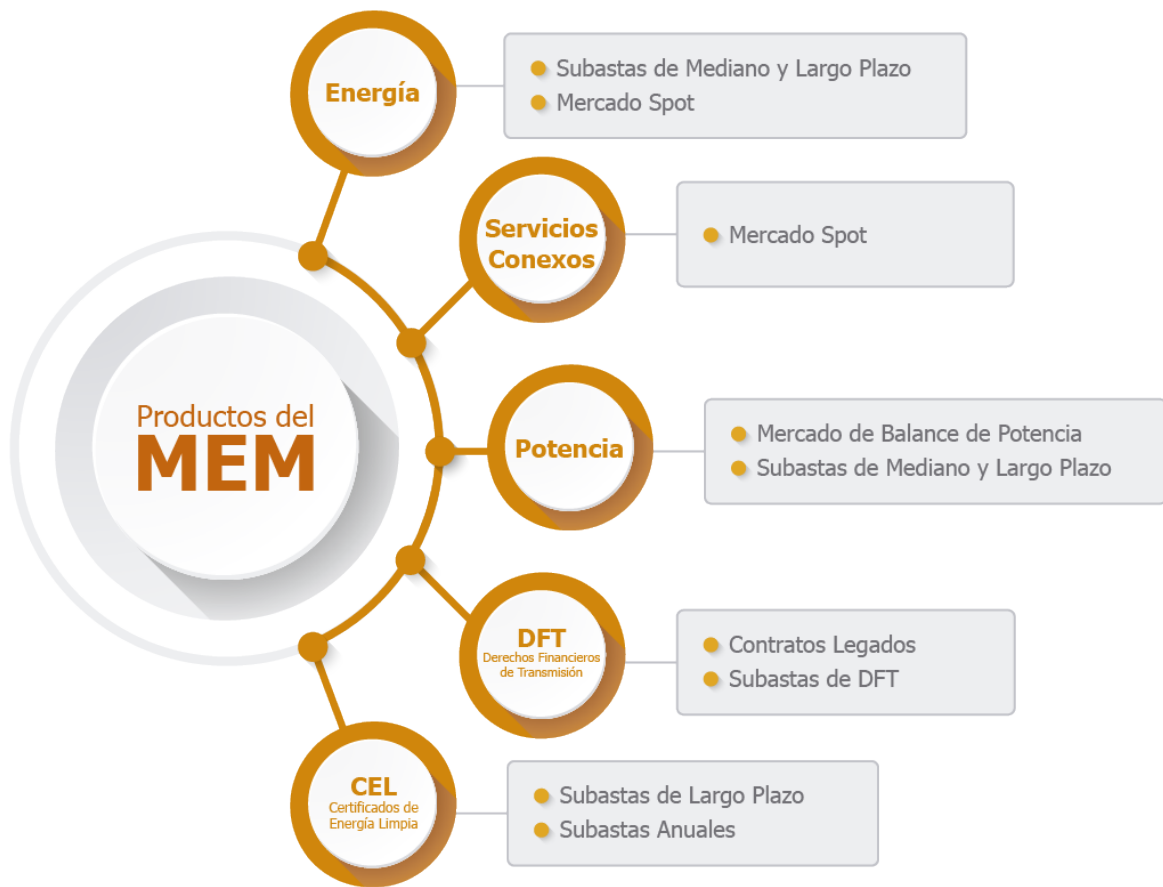


Figura 2.9: Mercados que componen el MEM ([9])

costos variables de cada Generador de Electricidad ⁵. En este orden de ideas, aquellas centrales eléctricas con los costos más bajos despacharían prioritariamente la energía.

En este mercado de energía a corto plazo se tienen tres variaciones que dependen del horizonte temporal de la transacción. Por ejemplo, el *Mercado de Día de Adelanto* estipula que la energía se comercializa un día antes de la transacción, mientras que si se comercializa el mismo día en la que es generada, se estaría hablando del *Mercado en Tiempo Real*; sin embargo, si en otro escenario, la energía se comercializa una hora antes de su generación, esta transacción se estaría llevando a cabo en el *Mercado de Hora de Adelanto*.

El presente trabajo de investigación estará enfocado en el Mercado de Día de Adelanto cuyo mecanismo de funcionamiento se presenta en la figura 2.10:

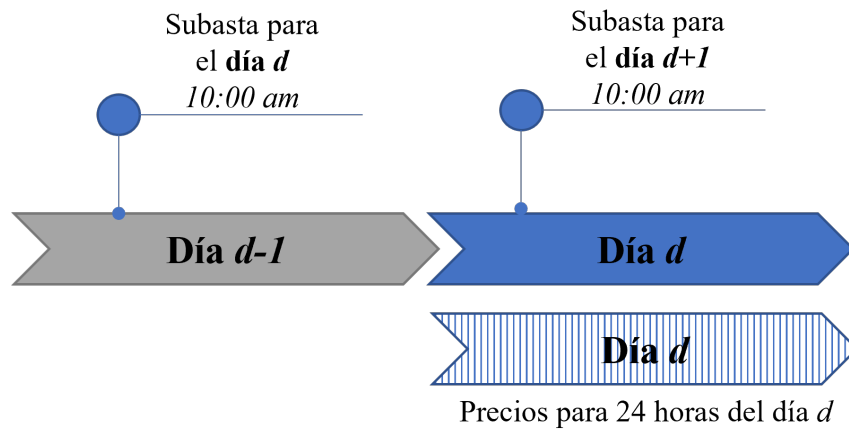


Figura 2.10: Mecanismo de subasta del Día de Adelanto: los participantes del MEM (compradores y vendedores) entregan sus necesidades de energía y sus ofertas antes del cierre en el día $d-1$ para la entrega de electricidad en el día d ([33])

- **Mercado de Balance de Potencia:** La potencia se define como la obligación que tiene el Generador de Energía de satisfacer la demanda de energía eléctrica y ofertarla en el Mercado de Energía a Corto Plazo. La idea es intercambiar la potencia para cubrir los desbalances entre los requisitos de Potencia de los Generadores y las cantidades registradas en sus respectivas Transacciones bilaterales. Este mercado opera con una periodicidad anual y los participantes del mercado deben ofertar dos meses antes de la finalización de cada año.
- **Mercado de Certificados de Energía Limpia (CEL):** Un CEL es un certificado que acredita la producción de energía eléctrica limpia y que se otorga con base en los

⁵**Generador:** Hace referencia a aquel participante del MEM que cuenta con centrales eléctricas que generan más de 0.5 MW y que tienen la capacidad de vender su electricidad. [16]

MWh generadores proveniente de estas fuentes. La CRE requiere a los generadores cumplir con un porcentaje mínimo de generación de energía eléctrica proveniente de estas fuentes. Así pues, la demanda de este mercado esta determinada por aquellos Generadores y Suministradores Calificados que no alcanzaron a cumplir dicha cuota mínima; mientras que la oferta estaría determinada por aquellos que alcanzaron y superaron el requerimiento mínimo establecido. Estos certificados tienen una vigencia de 20 años y las transacciones se efectúan al menos una vez al año.

- **Subasta de Derechos Financieros de Transmisión (DTF):** Se trata de coberturas que ayudan a cubrir la volatilidad del precio en los diferentes nodos del sistema. La transacción se hace entre titulares de DTF que pagan o cobran las diferencias de precios entre el nodo de origen y de destino. Dichas subastas se dan de forma periódica y son lideradas por el CENACE.
- **Subastas de Mediano y Largo Plazo:** Este tipo de subastas busca cubrir las necesidades de los participantes del mercado en contratos de cobertura eléctrica con los generadores. Las subastas denominadas de mediano plazo consideran transacciones de energía y potencia durante tres años, mientras que las de largo plazo, incluyen también intercambios de CELs, y por ende, pueden llegar a tener vigencias de entre 15 y 20 años.

2.3.2 FUNDAMENTACIÓN MATEMÁTICA

La presente sección explora los principios matemáticos que se emplean en el análisis de los datos disponibles y en el desarrollo del modelo de predicción de precios de la energía eléctrica. Para esto, se explorará teoría relevante con respecto a series de tiempo, el Coeficiente de Hurst y la Transformada de Fourier como herramientas matemáticas necesarias para el análisis y preparación de los datos utilizados para el pronóstico.

2.3.2.1 Series de tiempo

Una Serie de tiempo es una secuencia de datos ordenados, generalmente este ordenamiento se da en el tiempo. Y ¿para que se hace?, la respuesta es simple para entender el pasado y predecir el futuro. Por ejemplo, El Protocolo de Kyoto, uno de los anexos del Marco Regulatorio en la Convención de las Naciones Unidas para el Cambio Climático, tiene como argumento principal para la reducción de los gases de efecto invernadero los análisis económicos y de series de tiempo, ya que las decisiones que tomemos hoy afectarán el futuro. [48].

Cuando las variables son medidas de forma secuencial en el tiempo o en un periodo fijo, se conoce como intervalo de muestreo, y el conjunto de estos intervalos de muestreo dan como resultado una serie de tiempo. Se puede considerar cada secuencia como una variable aleatoria ó un proceso estocástico en tiempo discreto. Las características principales de muchas series de tiempo es la tendencia y periodicidad que se pueden modelar mediante funciones matemáticas. Es importante resaltar que en caso de que las observaciones tengan un período de muestreo cercano, la series tienden a estar correlacionadas (*Serie dependiente*).

El propósito final de la metodología con series de tiempo es explicar la existencia de esta correlación y las características de los datos que la conforman. Una vez que se ha encontrado un modelo que se ajuste a la información, se puede proceder a predecir el comportamiento futuro, inclusive generan escenarios para la adecuada toma de decisiones. El intervalo de muestreo puede variar de acuerdo con la información que se analice,

Agregada: el número de huéspedes de un hotel por día.

Muestreado: El precio diario de cierre de una acción.

Finalmente, en caso de que la información muestreada sea pequeña, se debe garantizar que cuando sea interpolada, estime de forma correcta los datos reales.

Entre las características de una serie se tiene:

- **Tendencia:** Cambio sistemático que no parece ser periódico. El modelo mas simple para la tendencia es un incremento/decremento lineal. Ejemplo, el precio de un alimento.
- **Variación periódica:** Patrón que se repite en cualquier periodo fijo de tiempo. Ejemplo, las reservas de un restaurante durante diferentes días de la semana.
- **Ciclos:** Cambios que no corresponden periodo natural fijo. Ejemplos, oscilaciones climáticas como El Niño.

2.3.2.1.1 Características

Notación

Una series de tiempo de longitud n , se puede representar como,

$$\{x_t : t = 1, \dots, n\} = x_1, x_2, \dots, x_n$$

Donde los n valores son muestreados de forma discreta $1, 2, \dots, n$. Esta notación puede ser abreviada a la forma $\{x_t\}$ si no se necesita especificar la longitud de n . Es importante tener presente que el modelo de series de tiempo es una secuencia de variables aleatorias, y los valores observados de la serie se consideran una realización del modelo. Para indicar una predicción utilizamos el símbolo \hat{x} . Por ejemplo la serie $\{x_t : t = 1, \dots, n\}$, $\hat{x}_{t+k|t}$ es la predicción hecha en el tiempo t para el valor futuro $t + k$, a la variable k se le conoce como tiempo de adelanto.

Descomposición

Las series de tiempo pueden ser descritas por sus componentes, como un sumatorio de la forma,

$$X_t = M_t + S_t + R_t$$

donde al tiempo t ,

- X_t es la serie observada.
- M_t es la tendencia ó la dirección de la serie en un intervalo largo de tiempo.
- S_t es el efecto periódico debido a un patrón observado en intervalos regulares de tiempo. Puede ser mensual, semanal, etc.
- R_t es el componente irregular, que consiste en las fluctuaciones de la serie después de remover los componentes previos.

Para descomponer una serie de tiempo, se debe primero calcular la tendencia por ciclo. Para ello se utiliza el suavizado por medias móviles. En donde una media móvil de orden m que se puede escribir como, cuando m es impar,

$$\hat{M}_t = \frac{1}{m} \sum_{j=-k}^k x_{t+j},$$

donde $k = (m - 1)/2$.

En caso que m fuese par, la ecuación se debe modificar introduciendo una ponderación α ,

$$\hat{M}_t = \sum_{j=-k}^k \alpha_j x_{t+j},$$

en donde todos los pesos $[\alpha_{-k}, \dots, \alpha_k]$ deben sumar uno y deben cumplir que $\alpha_k = \alpha_{-k}$. Esto quiere decir que para estimar la tendencia al tiempo t , evaluando la serie de tiempo en

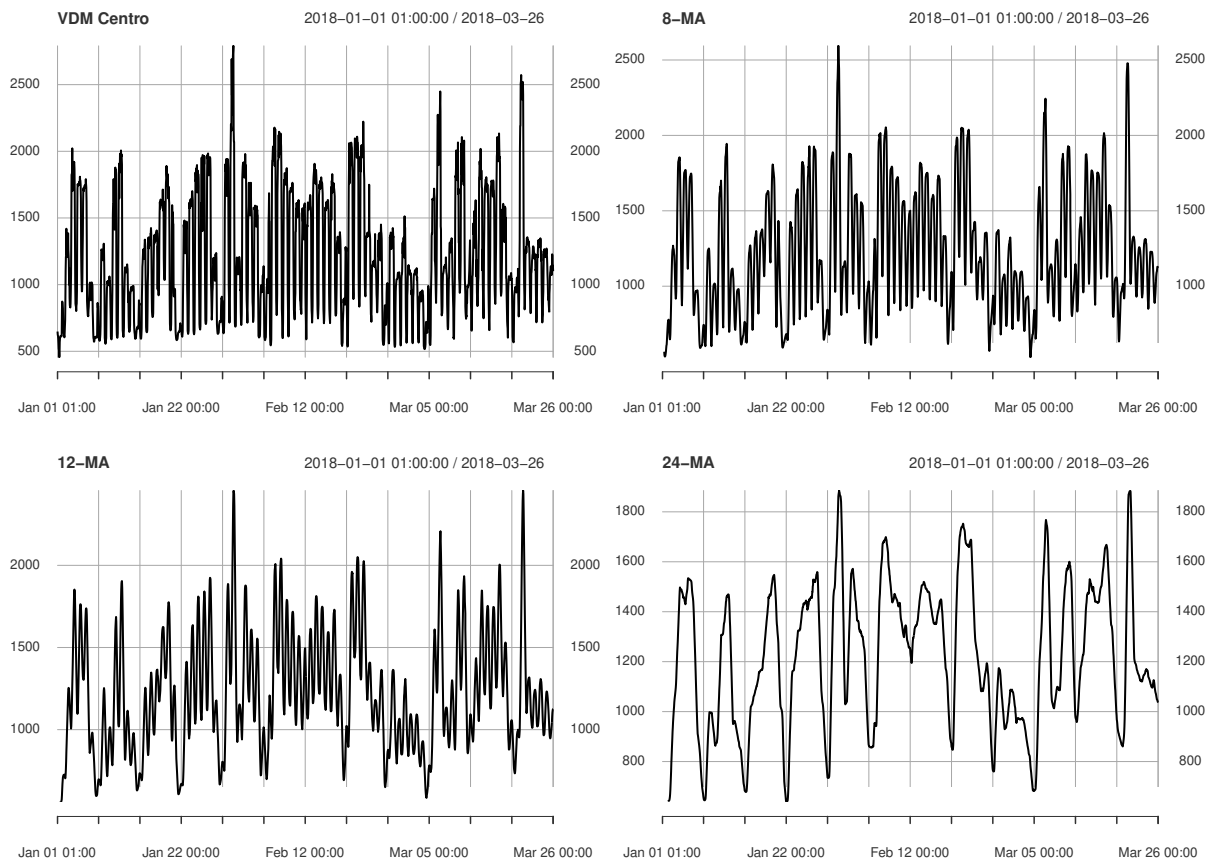


Figura 2.11: Diferentes medias móviles aplicadas a los precios de le energía en la región Centro para el primer trimestre de 2018

k periodos. Este procedimiento elimina los valores atípicos, dejando una curva suavizada. El resultado se conoce como m -MA, o una media móvil de orden m .

Una característica importante consiste en que la curva suavizada contiene el mismo perfil de los datos originales sin las pequeñas fluctuaciones. El orden de la media móvil la suavidad de la curva de tendencia. En general, a mayor orden una curva suave, como se muestra en la figura 2.11.

Para realizar la descomposición de una serie, se deben seguir los siguientes pasos

- Calcular el componente de tendencia \hat{M}_t utilizando m -MA.
- Calcular la serie sin tendencia $x_t - \hat{M}_t$.
- Calcular el componente periódico. Se debe calcular el valor promedio de la serie sin tendencia para cada uno de los periodos. Estos valores son ajustados para que la serie sea centrada en cero \hat{S}_t .

- El componente restante se calcula con la resta de los demás componentes. $\hat{R}_t = x_t - \hat{M}_t - \hat{S}_t$.

Por ejemplo, el suavizado a 168 horas, en este contexto, se refiere al cálculo del promedio móvil con una ventana de 168 observaciones. Esto implica tomar un conjunto de datos y calcular el promedio de los valores en intervalos de 168 observaciones. Esto puede ayudar a suavizar las fluctuaciones de corto plazo y resaltar tendencias más largas.

La fórmula para calcular un promedio móvil de 168 horas se vería de la siguiente manera:

$$\text{Promedio Móvil}_{168}(t) = \frac{x_{t-167} + x_{t-166} + \dots + x_{t-1} + x_t + \dots + x_{t+165} + x_{t+166} + x_{t+167}}{335}$$

Donde:

- x_t es el valor de la serie en el tiempo t .
- Los subíndices t_{-167} hasta t_{+167} representan las observaciones en torno a x_t en la ventana de 168 horas.
- 335 es el número total de observaciones en la ventana (167 antes + 1 actual + 167 después).

Es importante tener en cuenta que, como se mencionó anteriormente, el suavizado a 168 horas se basa en el cálculo de promedios móviles y no necesariamente en la construcción de un modelo matemático complejo. Los resultados de este cálculo proporcionarán una versión suavizada de la serie de tiempo original, lo que puede ayudar a visualizar tendencias de manera más clara al reducir el ruido de corto plazo.

Correlación

La correlación se mide de forma adimensional y representa la relación lineal entre un par de variables (x, y) y se obtiene estandarizando la covarianza dividiéndola entre la desviación estándar de las variables. La correlación toma valores entre -1 y 1 , siendo 0 el indicativo de ninguna relación lineal. La correlación de la población ρ entre el par de variables (x, y) se define como,

$$\rho(x, y) = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y} = \frac{\gamma(x, y)}{\sigma_x \sigma_y}$$

Podemos calcular la correlación de la muestra, Cor , como una estimación de ρ tal como lo propone [48],

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Donde,

- n es el número de observaciones en los conjuntos de datos x e y .
- x_i y y_i son las observaciones individuales en los conjuntos de datos x e y , respectivamente.
- \bar{x} y \bar{y} son las medias de los conjuntos de datos x e y , respectivamente.

La fórmula calcula la covarianza entre x e y , y luego la normaliza dividiendo por los desvíos estándar de x e y .

Autocorrelación

Así como la media y la varianza juegan un importante rol en el estudio de las distribuciones, porque resumen propiedades como la localización central y la dispersión, media, varianza y la correlación juegan un papel importante en los modelos de series de tiempo, de acuerdo con [48] se le conocen como propiedades de segundo orden.

Si la correlación entre las variables depende solamente del número de escalas de tiempo que las separan, y además es en la media y en la varianza estacionaria se conoce como estacionario de segundo orden. El número de escalas de tiempo se conoce como desfase. La correlación de una variable consigo misma en diferente tiempo se conoce como *autocorrelación* o *correlación serial*, por tanto si la series es estacionaria de segundo orden, la función de autocovarianza,

$$\gamma_k = E[(x_t - \mu)(x_{t+k} - \mu)]$$

Es de notar que la función γ_k no depende de t , ya que su valor esperado es el mismo para todos los t . La función de autocorrelación para el desfase (*lag*) k , se define como,

$$\rho_k = \frac{\gamma_k}{\sigma^2}$$

De aquí se concluye que $\rho_0 = 1$.

Correlograma

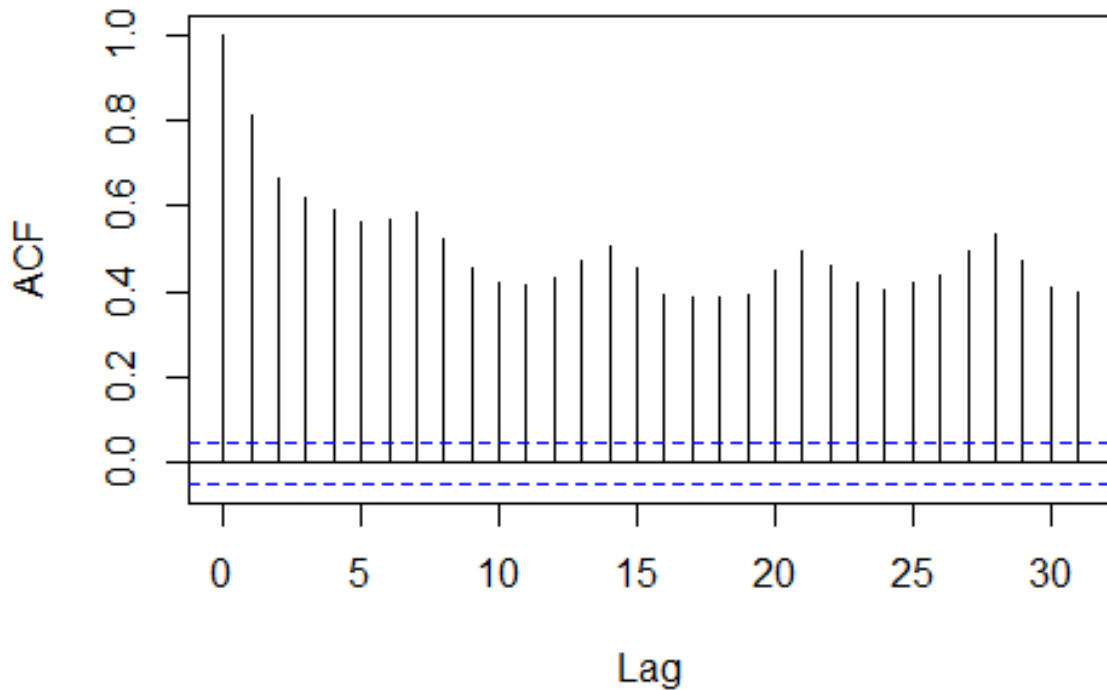


Figura 2.12: Correlograma de los precios de la energía diarios a las 7 AM en el periodo 2018 al 2022.

En la figura 2.12, se presenta un ejemplo de un correlograma. Para este caso en particular, se toma una muestra a 31 días para evidenciar la existencia de una periodicidad. Además, se puede observar que la caída gradual es típica de las series de tiempo con tendencia y que los picos en 7, 14, 21 y 28 indican una variación periódica.

La forma en que se puede explicar es la siguiente:

- El eje de las x corresponde al desfase, $lag(k)$ y el eje de las y nos indica la autocorrelación ρ_k . Las unidades en x corresponden al intervalo de muestreo, y la correlación al ser adimensional, no tiene unidades.
- Si $\rho_k = 0$, la distribución de la muestra se puede considerar normal con media $\frac{-1}{n}$ y varianza $\frac{1}{n}$, las líneas punteadas azules se dibujan en

$$\frac{-1}{n} \pm \frac{2}{\sqrt{n}}$$

Si el ρ_k esta por fuera de esas líneas, se tiene evidencia contra la hipótesis nula que $\rho_k = 0$ a un nivel del 5%. Cuando ρ_k cae fuera de estas líneas es mas probable que sean estadísticamente significantes.

- La autocorrelación cuando el desfase es 0 $lag(0)$ siempre será 1, como se ve en la figura 2.12, esto ayuda a comparar los valores de las otras autocorrelaciones, relativas al máximo teórico, 1. Por ejemplo si el $lag(1) = 0.1$ implica una dependencia lineal de x_t sobre x_{t-1} que solamente explica el 10 % de la variabilidad de x_t .

Entonces, el principal uso del correlograma es para detectar autocorrelaciones en las series de tiempo después que se ha removido las variaciones debidas a la tendencia y la estacionalidad.

Covarianza de la suma de variables aleatorias Sea x_1, x_2, \dots, x y y_1, y_2, \dots, y_m variables aleatorias, entonces,

$$Cov\left(\sum_{i=1}^n x_i, \sum_{j=1}^m y_j\right) = \sum_{i=1}^n \sum_{j=1}^m Cov(x_i, y_j)$$

De donde $Cov(x, y)$ es la covarianza entre el par de variables aleatorias x y y . Este resultado lo que dice es que la covarianza de la suma de dos variables es la suma de todas las posibles covarianzas del par de variables.

Correlación Cruzada

Se debe suponer que se tienen modelos de series de tiempo para las variables x y y , que son estacionarios en la media y la varianza. Las variables pueden estar correlacionadas entre ellas en diferentes desfases de t . Este modelo sera *estacionario de segundo orden* si todas las correlaciones dependen solo del desfase, y es posible definir la función de *covarianza cruzada* como función del desfase k ,

$$\gamma_k(x, y) = E[(x_{t+k} - \mu_x)(y_t - \mu_y)]$$

Lo que quiere decir que la variable x se está desfazando de la variable y en k . Por ejemplo, en un sistema físico, donde x es la entrada y y la respuesta, la causa precede el efecto (relación causa-efecto) se dice que y se desfazara de x . En este caso la función de covarianza cruzada sera 0 para valores positivos de k y para los valores negativos de k tendrá picos. Esta función se define como,

$$\gamma_k(x, y) = \gamma_{-k}(y, x)$$

El desfase k de la función de correlación cruzada (*ccf*), se define como $\rho_k(x, y)$,

$$\rho_k(x, y) = \frac{\gamma_k(x, y)}{\sigma_x \sigma_y}$$

2.3.2.1.2 Ruido Blanco

El error residual se define como la diferencia entre el valor observado y el valor estimado por el modelo en el tiempo t . Si el modelo es definido por la variable y_t y el valor estimado como \hat{y}_t , entonces el error residual x_t es,

$$e_t = y_t - \hat{y}_t$$

Ya que el error se tiene en cada espacio de tiempo, ello forma una series de tiempo (x_1, x_2, \dots, x_n) , por lo que la serie será descorrelacionada y el correlograma no mostrará ningún tipo de patrón.

Formalmente se define como $\{w_t : t = 1, 2, \dots, n\}$ será ruido blanco si las variables w_1, w_2, \dots, w_n son *independientes e idénticamente* distribuidas con media cero. Esto implica que las variables tienen la misma varianza σ^2 y $Cor(w_i, w_j) = 0$ para todo $i \neq j$. Adicionalmente si las variables siguen una distribución normal ($w_t \sim \mathcal{N}(\mu, \sigma^2)$) se conoce como *ruido blanco Gaussiano*.

Las propiedades de segundo orden de una serie de ruido blanco $\{w_t\}$ se define,

$$\begin{aligned} \mu_{wt} &= 0 \\ \gamma_k = Cov(w_t, w_{t+k}) &= \begin{cases} \sigma^2 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases} \end{aligned}$$

La función de autocorrelación se define como,

$$\rho_k = \begin{cases} 1 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$$

Un serie de ruido blanco, es construida como una serie de residuos después de ser ajustada correctamente a un modelo de series de tiempo. Si la serie tiene una longitud apropiada el correlograma da la suficiente información, para soportar la conjetura de que los residuos están bien aproximados con el ruido blanco. El único parámetro de esta serie es la varianza σ^2 , que es estimada por la varianza de los residuos.

2.3.2.1.3 Caminata Aleatoria

Una caminata aleatoria es un buen modelo para los datos estocásticos que tienen alguna tendencia. Sea $\{x_t\}$ una serie de tiempo con una marcada tendencia (creciente o decreciente).

Luego $\{x_t\}$ es una caminata aleatoria si,

$$x_t = x_{t-1} + w_t$$

donde $\{w_t\}$ es una serie de ruido blanco. En donde tenemos que $x_{t-1} = x_{t-2} + w_{t-1}$ y así sucesivamente lo que resulta en,

$$x_t = w_t + w_{t-1} + w_{t-2} + \dots$$

Para esta operación utilizamos el operador de retroceso \mathbf{B} , que se define como,

$$\mathbf{B}x_t = x_{t-1}$$

Este operador también se conoce como el operador de desfase. Si lo aplicamos n veces, da como resultado,

$$\mathbf{B}^n x_t = x_{t-n}$$

Las propiedades de segundo orden son,

$$\begin{aligned} \mu_x &= 0 \\ \gamma_k &= Cov(x_t, x_{t+k}) = t\sigma^2 \end{aligned}$$

Como la covarianza es función del tiempo, es un proceso no-estacionario. Para este caso, la varianza se incrementa al tiempo que t crece. De acuerdo con [48] es un modelo que se ajusta bien solamente para predicciones en una ventana de tiempo pequeña.

Ahora se debe introducir el operador de diferencias, su utilidad radica en que pueden convertir una serie no estacionaria en una serie estacionaria. Por ejemplo, la serie $\{x_t\}$ es una caminata aleatoria, esta será no estacionaria. Sin embargo si se aplica la diferencia de primer orden a $\{x_t\}$ produce una serie de ruido blanco estacionaria $\{w_t\}$, definida como $x_t - x_{t-1} = w_t$.

Este procedimiento puede ser visto como un filtro para estudiar las series de tiempo que no son estacionarias. El operador de diferencias ∇ se define como,

$$\nabla x_t = x_t - x_{t-1}$$

Se debe resaltar que $\nabla x_t = (1 - \mathbf{B})x_t$, por lo que ∇ puede ser expresado en termino del operador de retroceso \mathbf{B} . En general, para ordenes grandes se puede expresar como,

$$\nabla^n = (1 - \mathbf{B})^n$$

2.3.2.1.4 Modelos Autoregresivos

La serie $\{x_t\}$ es autoregresiva de orden p , abreviada como $AR(p)$, si

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + w_t$$

donde w_t es ruido blanco y α_i son los parámetros del modelo con $\alpha_p \neq 0$ para un proceso de orden p . Este modelo también se puede escrito como,

$$\theta_p(\mathbf{B})x_t = (1 - \alpha_1 \mathbf{B} - \alpha_2 \mathbf{B}^2 - \cdots - \alpha_p \mathbf{B}^p)x_t = w_t$$

Se debe resaltar los siguientes puntos,

1. La caminata aleatoria es un caso especial de grado $AR(1)$, con $\alpha = 1$.
2. El modelo es una regresión de x_t con la misma serie, de ahí su nombre autoregresivo.
3. La predicción al tiempo t esta dado por,

$$\hat{x}_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p}$$

4. Los parámetros del modelo se pueden obtener minimizando la suma de los errores al cuadrado.

Estos modelos tienen una ecuación característica, que se define como $\theta_p(\mathbf{B}) = 0$ y sus raíces deben mayores al valor absoluto de unidad para que el proceso sea *estacionario*. Por ejemplo, la caminata aleatoria tiene el polinomio característico $\theta = 1 - \mathbf{B}$ con raíz $\mathbf{B} = 1$ por lo tanto es *no estacionario*.

Las propiedades de segundo orden para un proceso $AR(1)$ se define,

$$\begin{aligned}\mu_x &= 0 \\ \gamma_k &= \frac{\alpha^k \sigma^2}{(1 - \alpha^2)}\end{aligned}$$

Autocorrelación Parcial

Para un proceso $AR(1)$ ⁶ la función de autocorrelación puede ser escrita como,

$$\rho_k = \alpha^k \quad (k \geq 0)$$

donde $|\alpha| < 1$. Por tanto, la autocorrelaciones son diferentes a cero para todos los desfases k , inclusive si el modelo x_t solamente depende del valor inmediatamente anterior x_{t-1} . Por este motivo, se genera la *autocorrelación parcial* en el desfase k que se puede ver como la autocorrelación que se genera después de remover los efectos de la correlación de los desfases k mas cercanos. Por ejemplo, para un proceso $AR(1)$, la autocorrelación será cero para todos los desfases mayores a 1. Generalizando, la autocorrelación parcial en el desfase k , es el $k - \text{esimo}$ coeficiente del modelo $AR(k)$, si este proceso es $AR(p)$ entonces los coeficientes α_k serán cero para $k > p$. Por tanto, un proceso $AR(p)$ tendrá un correlograma de autocorrelaciones parciales igual a cero para los desfases después del desfase p .

El anexo A presenta la solución del modelo autorregresivo AR de primer orden de manera exacta $AR(1)$.

2.3.2.1.5 Series estrictamente estacionarias

Un modelo de series de tiempo $\{x_t\}$ es *estrictamente estacionario* si la distribución conjunta de x_{t_1}, \dots, x_{t_n} es la misma que la distribución conjunta para $x_{t_1+m}, \dots, x_{t_n+m}$ para todo t_1, \dots, t_n y m , quiere decir que la distribución no cambia a pesar de cualquier adelanto o atraso. Esto implica que la media y la varianza sean constantes en el tiempo y que la autocovarianza $Cov(x_t, x_s)$ dependa solamente del desfase $k = |t - s|$ que puede ser escrita como $\gamma(k)$.

Si una serie no es estrictamente estacionaria, pero la media y la varianza son constantes en el tiempo y la autocovarianza solo depende del desfase, entonces se conoce estacionario de *segundo orden* también se conoce como *débilmente estacionaria*.

⁶ $x_t = \alpha x_{t-1} + w_t$

Estacionariedad es una idealización que es parte de las propiedades de los modelos. Si se logra ajustar un modelo estacionario a un conjunto de datos, se asume que esos datos son una realización de un proceso estacionario. El primer paso en el análisis debe ser identificar cualquier tendencia o efecto periódico y removerlo. Por ello es razonable tratar las series de tiempo de los residuos como una realización de una serie de errores estacionarios.

2.3.2.1.6 Modelos de media móvil

Un proceso de media móvil (*MA*) es de orden q si puede ser expresado como una combinación lineal del ruido blanco en el tiempo t y el mas reciente q^{avo} ruido blanco, se define como,

$$x_t = w_t + \beta_1 w_{t-1} + \cdots + \beta_q w_{t-q}$$

donde $\{w_t\}$ es ruido blanco con media cero y varianza σ_w^2 . Si utilizamos el operador de retroceso,

$$x_t = (1 + \beta_1 \mathbf{B} + \beta_2 \mathbf{B}^2 + \cdots + \beta_q \mathbf{B}^q) w_t = \phi_q(\mathbf{B}) w_t$$

donde ϕ_q es un polinomio de orden q . Ya que un proceso *MA* es una suma finita de términos de ruido blanco, es estacionario y además tiene una media y autocovarianza invariante en el tiempo.

La función de autocorrelación para $T \geq 0$, se define como,

$$\rho_k = \begin{cases} 1 & k = 0 \\ \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^q \beta_i^2} & k = 1, \dots, q \\ 0 & si \quad k > q \end{cases}$$

El resultado es producto de que la media es cero porque consiste en la suma de términos que tienen media cero. La varianza es $\sigma_w^2(1 + \beta_1^2 + \cdots + \beta_q^2)$ porque cada término es ruido blanco independiente con la misma varianza.

En general un proceso *MA*(q) es invertible cuando las raíces del polinomio característico $\phi_q(B)$ exceden la unidad en valor absoluto.

2.3.2.1.7 Proceso ARMA

Estos modelos son una combinación de los modelos *AR* y *MA* cuando sus términos son añadidos en una sola expresion. Se dice que una series $\{x_t\}$ sigue un modelo

autoregresivo con media móvil de orden (p, q) , denotado $ARMA(p, q)$, si

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + w_t + \beta_1 w_{t-1} + \beta_2 w_{t-2} + \cdots + \beta_q w_{t-q}$$

donde $\{w_t\}$ es ruido blanco. Si se representa con el operador de retroceso

$$\theta_p(\mathbf{B})x_t = \phi_q(\mathbf{B})w_t$$

. Puntos a tener en cuenta en un proceso $ARMA(p, q)$

1. El proceso es estacionario cuando las raíces de θ son estrictamente mayores en valor absoluto.
2. El proceso es invertible cuando las raíces de ϕ son estrictamente mayores en valor absoluto.
3. El modelo $AR(p)$ es un caso especial de $ARMA(p, 0)$.
4. El modelo $MA(q)$ es un caso especial de $ARMA(0, q)$.
5. De acuerdo al *principio de parsimonia* un modelo $ARMA$, suele requerir menos parámetros que un modelo AR ó MA .
6. Cuando θ y ϕ tienen un factor común, *redundancia de parámetros*, el modelo estacionario puede ser simplificado.

2.3.2.1.8 Proceso ARIMA

Los modelos que se describen ahora, son usados para las series de tiempo que *no son estacionarias* ya sea porque tienen tendencia o porque presentan un comportamiento periódico. Como se explicó anteriormente las caminatas aleatorias ⁷ puede ser transformado a una serie de tiempo estacionaria aplicando el operador de diferencias ∇ . Para generalizar este procedimiento a los términos autoregresivos y de media móvil es necesario agregar (*integrar*) un termino para volver a la serie original. Resumiendo, el proceso ahora se llamara, Autoregresivo *integrado* de media móvil o abreviado *ARIMA*

Una serie $\{x_t\}$ es *integrada* de orden d , $I(d)$, si la diferencia d^{ava} de $\{x_t\}$ es ruido blanco $\{w_t\}$, por ejemplo $\nabla^d x_t = w_t$. Dado que $\nabla^d \equiv (1 - \mathbf{B})^d$ la serie $\{x_t\}$ es integrada de orden d si

$$(1 - \mathbf{B})^d x_t = w_t$$

⁷No estacionario

Se concluye que la caminata aleatoria es un caso especial $I(1)$.

Una serie $\{x_t\}$ sigue un proceso $ARIMA(p, d, q)$ si la diferencia d^{ava} de $\{x_t\}$ es un proceso $ARMA(p, q)$. Si definimos $y_t = (1 - \mathbf{B})^d x_t$, entonces $\theta_p(\mathbf{B})y_t = \phi_q(\mathbf{B})w_t$. Ahora si se sustituye por y_t por un proceso $ARIMA(p, d, q)$

$$\theta_p(\mathbf{B})(1 - \mathbf{B})^d x_t = \phi_q(\mathbf{B})w_t$$

donde θ_p y ϕ_q son polinomios de orden p y q .

2.3.2.1.9 Proceso SARIMA

La estacionalidad en una serie de tiempo es un patrón regular que se repite durante S periodos de tiempo, donde S define el numero de periodos de tiempo hasta que patrón se repite nuevamente.

De acuerdo con [46], la diferenciación estacional se puede definir como una diferencia entre un valor y un valor de rezago que es múltiplo de S .

Este proceso usa el operador de diferenciación ∇ en el desfase igual al número de periodos (s) para remover los efectos de esta periodicidad. Este proceso introduce un término de media móvil. Este modelo incluye un termino autoregresivo y de media móvil en el desfase s . Este modelo $SARIMA(p, d, q)(P, D, Q)_s$ se puede reescribir usando el operador de retroceso,

$$\Theta_p(\mathbf{B}^s)\theta_p(\mathbf{B})(1 - \mathbf{B}^s)^D(1 - \mathbf{B})^d x_t = \Phi_Q(\mathbf{B}^s)\phi_q(\mathbf{B})w_t$$

donde $\Theta_P, \theta_p, \Phi_Q$ y ϕ_q son polinomios de orden P, p, Q y q respectivamente.

El componente periódico de un modelo $AR(p)$ ó $MA(q)$, se vera reflejado en los retrocesos de las funciones ACF y PACF. Por ejemplo, el modelo $(0, 0, 0)(0, 0, 1)_{12}$ tendrá.

1. Solamente un pico significativo en la unidad 12 de la ACF.
2. Un decaimiento exponencial en los retrocesos periódicos de la PACF $(12, 24, 36, \dots)$.

Ahora si el modelo es $(0, 0, 0)(1, 0, 0)_{12}$ tendrá,

1. Un decaimiento exponencial en los retrocesos periódicos de la PACF $(12, 24, 36, \dots)$ de la ACF.
2. Solamente un pico significativo en la unidad 12 de la PACF

En resumen cada termino, autoregresivo AR ó de media móvil MA , se abreviaba, como se presenta en las ecuaciones 2.1 y 2.2 para la parte no estacional y para la parte estacional las ecuaciones 2.3 y 2.4 respectivamente.

$$AR : \quad \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (2.1)$$

$$MA : \quad \theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (2.2)$$

$$AR : \quad \Phi(B^S) = 1 - \phi_1 B^S - \dots - \phi_P B^{PS} \quad (2.3)$$

$$MA : \quad \Theta(B^S) = 1 + \theta_1 B^S + \dots + \theta_Q B^{QS} \quad (2.4)$$

2.3.2.2 Coeficiente de Hurst

El estudio de las series de tiempo con un rango largo de dependencia, las cuales pueden exhibir patrones cíclicos que pueden ser una indicación de la existencia de una memoria de largo plazo; esto es medir la dependencia a largo plazo de una serie de tiempo. Vamos a empezar explicando el método R/S, el análisis de fluctuación sin tendencia (DFA) por sus siglas en inglés y el periodograma de regresión.

2.3.2.2.1 Análisis R/S

De acuerdo con [66] el análisis se realiza de la siguiente forma; dividimos la serie de tiempo de longitud L en d subseries de longitud n . Luego para cada una de las subseries $m = 1, \dots, d$: (1) encontrar la media (E_m) y la desviación estándar (S_m); (2) normalizar la data ($Z_{i,m}$) restando la media muestral $X_{i,m} = Z_{i,m} - E_m$ para $i = 1, \dots, n$ (3) crear ahora una serie de tiempo acumulativa $Y_{i,m} = \sum_{j=1}^i X_{j,m}$ para $i = 1, \dots, n$ (4) encuentre el rango $R_M = \max\{Y_{1,m}, \dots, Y_{n,m}\} - \min\{Y_{1,m}, \dots, Y_{n,m}\}$ (5) reescalar el rango R_m/S_m para finalmente calcular la media del rango reescalado para todas las subseries de longitud n

$$(R/S)_n = \frac{1}{d} \sum_{m=1}^d R_m/S_m$$

Se puede demostrar que el estadístico (R/S) , asintóticamente converge a la relación:

$$(R/S)_n \sim cn^H$$

De donde el valor de H se puede obtener, haciendo una regresión lineal simple sobre un horizonte de tiempo creciente.

$$(R/S)_n = \log c + H \log n$$

Al graficar la ecuación tenemos tres resultados importantes sobre el valor de H , según sea el caso:

Ruido blanco: tendríamos una línea recta con una pendiente $H = 0.5$.

Proceso persistente: La pendiente será $H > 0.5$.

Proceso anti-persistente: La pendiente será $H < 0.5$.

Igualmente el nivel de significancia se suele escoger como uno entre la raíz cuadrada de la longitud de la muestra. Ejemplo, la desviación estándar del ruido blanco Gaussiano. Para valores pequeños de n , $n \leq 340$ ver [66].

2.3.2.2.2 Análisis de fluctuación sin tendencia (DFA)

Este método propone dividir la serie de longitud L en d subseries de longitud n . Para cada una de las subseries $m = 1, \dots, d$: (1) crear una serie de tiempo acumulativa $Y_{i,m} = \sum_{j=1}^i X_{j,m}$ para $i = 1, \dots, n$; (2) ajustar la línea de mínimos cuadrados $\tilde{Y}_m(x) = a_m x + b_m$ para $\{Y_{1,m}, \dots, Y_{n,m}\}$; (3) calcular la raíz media cuadrada de la fluctuación (ejemplo: la desviación estándar) de la integración y de la serie de tiempo sin tendencia.

$$F(m) = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{i,m} - a_m i - b_m)^2}$$

Finalmente, calcular el valor medio de la raíz media cuadrada de la fluctuación para todas las subseries de longitud n .

$$\bar{F}(n) = \frac{1}{d} \sum_{m=1}^d F(m)$$

Igual que en el análisis R/S, existe una relación lineal de $\bar{F}(n)$ contra el tamaño del intervalo n indicando una relación de potencias de la forma cn^H , con los mismos resultados ya mencionados.

2.3.2.2.3 Periodograma

El periodograma es una herramienta utilizada en análisis de series temporales y procesamiento de señales para estimar la densidad espectral de potencia de una señal. En términos simples, muestra cómo la potencia de una señal está distribuida en diferentes frecuencias. El periodograma es una representación gráfica que puede proporcionar información sobre las componentes de frecuencia presentes en una señal.

La fórmula para calcular el periodograma de una serie temporal discreta $x(t)$ en el punto de tiempo t es:

$$P(\omega) = \frac{1}{N} \left| \sum_{t=0}^{N-1} x(t)e^{-i\omega t} \right|^2$$

Donde:

- $P(\omega)$ es el valor del periodograma en la frecuencia angular ω .
- N es la longitud de la serie temporal.
- $x(t)$ es el valor de la serie temporal en el punto de tiempo t .
- i es la unidad imaginaria ($i^2 = -1$).
- ω es la frecuencia angular.

El cálculo del periodograma implica tomar la transformada discreta de Fourier (DFT) de la serie temporal y luego calcular el módulo al cuadrado de esta transformada en diferentes frecuencias angulares ω .

En R, puedes usar la función `spec.pgram` del paquete `stats` para calcular y visualizar el periodograma de una serie temporal. Aquí hay un ejemplo básico:

```
# Crear una serie temporal sintética
set.seed(123)
n <- 100
time <- seq(0, 10, length.out = n)
signal <- sin(2 * pi * 1 * time) + 0.5 * sin(2 * pi * 3 * time)

# Calcular el periodograma
```

```
periodogram <- spec.pgram(signal, plot = FALSE)
```

```
# Graficar el periodograma
```

```
plot(periodogram, main = "Periodograma de la Serie Temporal")
```

Este código generará un gráfico del periodograma de una serie temporal sintética que contiene dos componentes de frecuencia.

Según Arthur Schuster, un periodograma es un estimado de la densidad espectral de una señal. Es una herramienta para examinar la amplitud vs la frecuencia.

Dados los datos x_1, \dots, x_n , el periodograma se define como:

$$I(\omega_j) = |d(\omega_j)|^2 \quad \text{para } j = 0, 1, 2, \dots, n-1 \quad (2.5)$$

$$d(\omega_j) = n^{(-1/2)} \sum_{t=1}^n (x_t - \bar{x}) e^{-2\pi i \omega_j t}, \quad j \neq 0 \quad (2.6)$$

$$I(\omega_j) = |d(\omega_j)|^2 = n^{-1} \sum_{t=1}^n \sum_{s=1}^n (x_t - \bar{x})(x_s - \bar{x}) e^{-2\pi i \omega_j (t-s)} \quad (2.7)$$

$$= n^{-1} \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-\bar{h}} (x_{t+h} - \bar{x})(x_t - \bar{x}) e^{-2\pi i \omega_j h} \quad (2.8)$$

$$= \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) e^{-2\pi i \omega_j h} \quad (2.9)$$

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t e^{-2\pi i \omega_j t} \quad (2.10)$$

$$x_t = n^{-1/2} \sum_{j=0}^{n-1} d(\omega_j) e^{2\pi i \omega_j t} \quad (2.11)$$

$$(x_t - \bar{x}) = \frac{2}{\sqrt{n}} \sum_{j=1}^m [d_c(\omega_j) \cos(2\pi \omega_j t) + d_s(\omega_j) \sin 2\pi \omega_j t] \quad (2.12)$$

$$\sum_{t=1}^n (x_t - \bar{x})^2 = 2 \sum_{j=1}^m [d_c^2(\omega_j) + d_s^2(\omega_j)] \quad (2.13)$$

$$= 2 \sum_{j=1}^m I(\omega_j) \quad (2.14)$$

$$\hat{a}_0 = \bar{x} \quad (2.15)$$

$$\hat{d}_j = \frac{\sum_{t=1}^n (x_t - \bar{x}) \cos\left(\frac{2\pi t_j}{n}\right)}{\sum_{t=1}^n \cos\left(\frac{2\pi t_j}{n}\right)^2} \quad (2.16)$$

$$\begin{aligned} &= \frac{\sum_{t=1}^n (x_t - \bar{x}) \cos\left(\frac{2\pi t_j}{n}\right)}{\frac{n}{2}} \\ &= \frac{2}{n} \sum_{t=1}^n (x_t - \bar{x}) \cos\left(\frac{2\pi t_j}{n}\right) \\ &= \frac{2}{n} \sum_{t=1}^n x_t \cos\left(\frac{2\pi t_j}{n}\right) \end{aligned} \quad (2.17)$$

$$= \frac{2}{n} \sum_{t=1}^n x_t \sin\left(\frac{2\pi t_j}{n}\right) \quad (2.18)$$

Finalmente, en el **periodograma** la frecuencia es medida en ciclos por unidad de tiempo, donde la unidad de tiempo es definida como la distancia entre dos puntos. Una frecuencia de 0 corresponde a un ciclo infinito, mientras que una frecuencia de 0.5 corresponde a un ciclo de dos datos vecinos (frecuencia de Nyquist). Las series que están equidistantes tendrán frecuencias entre 0 y 0.5. Es recomendable remover la tendencia de las series antes de graficar el espectro de frecuencias. [61]

Periodograma de regresión

Es un método semi-paramétrico para obtener un estimado de la diferencia fraccional del parámetro d . Se basa en la observación de la pendiente de la función de densidad espectral de una serie integrada fraccionalmente alrededor de la frecuencia angular $w = 0$. Dado que la función de densidad espectral de un modelo general fraccional integrado (ejemplo: FARIMA) con parámetro de diferencia d es idéntico al del fraccional del ruido blanco Gaussiano con exponente de Hurst $H = d + 0.5$.

El procedimiento para estimar H , inicia calculando el periodograma, el cual es análogo a la densidad espectral. Para un vector de observaciones $\{x_1, \dots, x_L\}$, el periodograma se define como

$$I_L(w_k) = \frac{1}{L} \left| \sum_{t=1}^L x_t e^{-2\pi i(t-1)w_k} \right|^2,$$

De donde, $w_k = k/L$, $k = 1, \dots, [L/2]$ y $[x]$ denota el entero mas grande $\leq x$. Resaltar que I_L es el valor absoluto cuadrado de la transformada de Fourier, y si el vector

de observación es de la longitud apropiada (par o una potencia de 2), podemos utilizar algoritmos rápidos para calcular la transformada de Fourier.

El siguiente y último paso es realizar una regresión lineal simple,

$$\log\{I_L(w_k)\} = a - \hat{d} \log\{4 \sin^2(w_k/2)\} + \varepsilon_k,$$

Para bajas frecuencias de Fourier w_k , $k = 1, \dots, K \leq [L/2]$. La estimación de la media por mínimos cuadrados se desempeña como el parámetro diferenciador d en la relación $d = \hat{d}$, de donde $H = \hat{d} + 0.5$. La complicación de este método es escoger el valor de K . De acuerdo con [21] recomiendan escoger un valor de K tal que $K = [L^{0.5}]$, sin embargo otros valores también han sido sugeridos (ejemplo: $K = [L^{0.45}]$, $[L^{0.2}] \leq K \leq [L^{0.5}]$).

2.3.2.3 Transformada de Fourier

Esta sección explora teóricamente la transformada de Fourier que, en este caso, es empleada para identificar las frecuencias más importantes de la serie de tiempo en análisis.

Se dice que una función $f(x)$ tiene un periodo P o es periódica con un periodo P para todo x , $f(x+P)=f(x)$ donde P es una constante positiva.

Sea $f(x)$ definida en un intervalo $(-L, L)$ y determinada fuera de este intervalo por $f(x+2L)=f(x)$.

Por ejemplo: para $f(x)$ con un periodo $2L$, la serie de Fourier se representa mediante la ecuación 2.19:

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right) \right) \quad (2.19)$$

Para determinar los coeficientes a_n y b_n , se tienen las ecuaciones 2.20 y 2.21:

$$\frac{1}{L} \int_C^{C+2L} f(x) \cos\left(\frac{n\pi x}{L}\right) dx = a_n = \frac{1}{L} \int_{-L}^L f(x) \cos\left(\frac{n\pi x}{L}\right) dx \quad (2.20)$$

$$\frac{1}{L} \int_C^{C+2L} f(x) \sin\left(\frac{n\pi x}{L}\right) dx = b_n = \frac{1}{L} \int_{-L}^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx \quad (2.21)$$

Para $n = 0, 1, 2, \dots$

Ahora bien,

- $f(x)$ es **impar** si $f(-x) = -f(x)$
- $f(x)$ es **par** si $f(-x) = f(x)$

Teniendo que:

$$\int_a^b \left(\sum_{n=1}^{\infty} u_n(x) \right) dx = \sum_{n=1}^{\infty} \int_a^b u_n(x) dx \quad (2.22)$$

$$\frac{d}{dx} \sum_{n=1}^{\infty} u_n(x) = \sum_{n=1}^{\infty} \frac{d}{dx} u_n(x) \quad (2.23)$$

Con:

$$e^{i\theta} = \cos \theta + i \sin \theta \quad y \quad e^{-i\theta} = \cos \theta - i \sin \theta \quad (2.24)$$

Entonces,

$$f(x) = \sum_{-\infty}^{\infty} C_n e^{\frac{in\pi x}{L}} \quad (2.25)$$

$$C_n = \frac{1}{2L} \int_{-L}^L f(x) e^{-\frac{in\pi x}{L}} dx$$

Ejemplo

$$f(x) = \begin{cases} 0, & -5 < x < 0 \\ 3, & 0 < x < 5 \end{cases} \quad \text{Periodo} = 10$$

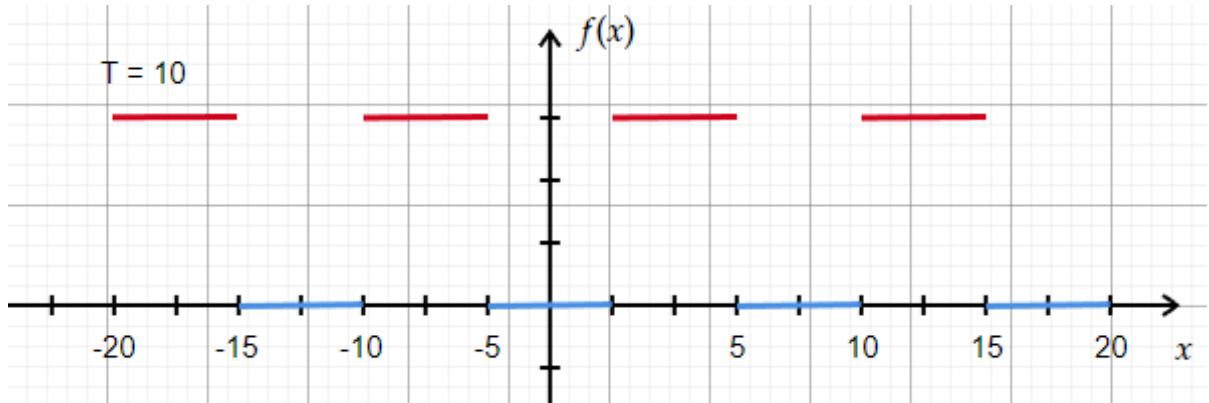


Figura 2.13: Función definida por partes con periodo 10.

$$\begin{aligned}
 a_n &= \frac{1}{L} \int_C^{C+2L} f(x) \cos\left(\frac{n\pi x}{L}\right) dx = \frac{1}{5} \int_{-5}^5 f(x) \cos\left(\frac{n\pi x}{5}\right) dx \\
 &= \frac{1}{5} \left\{ \int_{-5}^0 (0) \cos\left(\frac{n\pi x}{5}\right) dx + \int_0^5 (3) \cos\left(\frac{n\pi x}{5}\right) dx \right\} \\
 &= \frac{3}{5} \int_0^5 \cos\left(\frac{n\pi x}{5}\right) dx \\
 &= \frac{3}{5} \left(\frac{5}{n\pi} \sin\left(\frac{n\pi x}{5}\right) \right) \Big|_0^5 = 0 \quad \text{si } n \neq 0 \\
 a_n = a_0 &= \frac{3}{5} \int_0^5 (3) \cos\left(\frac{0\pi x}{5}\right) dx = \frac{3}{5} \int_0^5 dx = 3 \quad \text{si } n = 0
 \end{aligned}$$

$$\begin{aligned}
 b_n &= \frac{1}{L} \int_C^{C+2L} f(x) \sin\left(\frac{n\pi x}{L}\right) dx = \frac{1}{5} \int_{-5}^5 f(x) \sin\left(\frac{n\pi x}{5}\right) dx \\
 &= \frac{1}{5} \left\{ \int_{-5}^0 (0) \sin\left(\frac{n\pi x}{5}\right) dx + \int_0^5 (3) \sin\left(\frac{n\pi x}{5}\right) dx \right\} = \frac{0}{5} \int_0^5 \sin\left(\frac{n\pi x}{5}\right) dx \\
 &= \frac{3}{5} \left(\frac{-5}{n\pi} \cos\left(\frac{n\pi x}{5}\right) \right) \Big|_0^5 = \frac{3(1 - \cos n\pi)}{n\pi}
 \end{aligned}$$

$$\begin{aligned}
 \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right) \right) &= \frac{3}{2} + \sum_{n=1}^{\infty} \frac{3(1 - \cos n\pi)}{n\pi} \sin\left(\frac{n\pi x}{5}\right) \\
 &= \frac{3}{2} + \frac{6}{\pi} \left(\sin\left(\frac{\pi x}{5}\right) + \frac{1}{3} \sin\left(\frac{3\pi x}{5}\right) + \frac{1}{5} \sin\left(\frac{5\pi x}{5}\right) + \dots \right)
 \end{aligned}$$

2.3.2.3.1 Análisis Espectral

La descomposición de una serie de tiempo en senos y cosenos de diferentes frecuencias permite determinar aquellas frecuencias que son particularmente fuertes o importantes.

$$\begin{aligned}
 P(j/n) &= \left(\frac{2}{n} \sum_{t=1}^n x_t \cos \left(\frac{2\pi t j}{n} \right) \right)^2 + \left(\frac{2}{n} \sum_{t=1}^n x_t \sin \left(\frac{2\pi t j}{n} \right) \right)^2 \\
 \omega_j &= \frac{j}{n} \quad j \text{ ciclos en "n" puntos de tiempo} \\
 P(j/n) &= P(1 - \frac{j}{n}), \quad j = 0, 1, \dots, n-1
 \end{aligned}
 \tag{2.26}$$

La frecuencia Nyquist es igual a 1/2 de la frecuencia de muestreo. Se recomienda revisar el anexo B

Con $n = \text{impar}$

$$x_t = a_0 + \sum_{j=1}^{\frac{(n-1)}{2}} \left[a_j \cos \left(\frac{2\pi j t}{n} \right) + b_j \sin \left(\frac{2\pi j t}{n} \right) \right]
 \tag{2.27}$$

Con $n = \text{par}$

$$x_t = a_0 + \sum_{j=1}^{\frac{n}{2}-1} \left[a_j \cos \left(\frac{2\pi j t}{n} \right) + b_j \sin \left(\frac{2\pi j t}{n} \right) \right] + a_{n/2} \cos \left(\frac{2\pi t}{2} \right) = a_{n/2} (-1)^t
 \tag{2.28}$$

$$\begin{aligned}
 P(j/n) &= a_j^2 + b_j^2 \\
 a_j &= \frac{2}{n} \sum_{t=1}^n x_t \cos(2\pi\omega_j t) = \frac{2}{\sqrt{n}} d_c(\omega_j) \\
 b_j &= \frac{2}{n} \sum_{t=1}^n x_t \sin(2\pi\omega_j t) = \frac{2}{\sqrt{n}} d_s(\omega_j)
 \end{aligned}
 \tag{2.29}$$

Donde d_c y d_s , se definen con las ecuaciones 2.30 y 2.31 respectivamente:

$$d_c(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t \cos 2\pi\omega_j t \quad (2.30)$$

$$d_s(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t \sin 2\pi\omega_j t \quad (2.31)$$

Con $\omega_j = j/n$ para $j = 0, 1, \dots, n - 1$

$$d(\omega_j) = d_c(\omega_j) - id_s(\omega_j)$$

$$I(\omega_j) = d_c^2(\omega_j) + d_s^2(\omega_j)$$

CAPÍTULO 3

METODOLOGÍA

El presente capítulo explora y describe detalladamente la metodología seguida para el desarrollo del modelo y su respectiva predicción.

Inicialmente, es importante destacar los pasos básicos tanto para resolver problemas físicos mediante formulación matemática como aquellos mínimos requeridos para desarrollar predicciones.

En este orden de ideas, para abordar y darle solución a cualquier problema físico mediante formulación matemática, se deben seguir, de forma general, los siguientes pasos:

- Proponer un modelo matemático - Formulación matemática per-se
- Solucionar el modelo - Solución matemática
- Interpretar los resultados matemáticos aplicados al fenómeno físico en estudio

Mientras que, por su lado, [55] propone los pasos básicos para la realización de predicciones, así:

1. **Definición del problema:** Este paso se basa en generar un entendimiento profundo del uso futuro del pronóstico por parte de los potenciales usuarios.
2. **Recolectar información:** En esta fase se requieren tanto datos estadísticos como el conocimiento y experiencia de quienes recogen los datos y usan la información. La data debe ser cuidadosamente seleccionada para no usar información desactualizada que no agregue valor al pronóstico o para que, por el contrario, se deseche información relevante.
3. **Análisis exploratorio:** En este paso, se recomienda graficar los datos disponibles para identificar patrones, tendencias, estacionalidades, relación entre variables, etc.

4. **Escoger y ajustar los modelos:** El modelo a utilizar depende de los datos disponibles, la relación entre las variables analizadas, y el uso futuro del pronóstico en cuestión. Es posible necesitar probar varios modelos antes de determinar cuál es el que mejor se ajusta al set de datos en estudio.
5. **Usar y evaluar el modelo de pronóstico:** Una vez se ha seleccionado el modelo y se ha determinado un pronóstico inicial, es necesario evaluar la precisión del mismo. Existen técnicas específicas que pueden utilizarse con este fin y que deben ser incluidas en cualquier proceso de pronóstico.

La misma fuente [55], presenta una descripción detallada del flujo de trabajo que debe seguirse para producir una predicción mediante series de tiempo.

1. **Preparación de los datos:** Uno de los pasos más importantes que hay que llevar a cabo previamente tiene que ver con la preparación de los datos de forma que tengan el formato apropiado. Este proceso puede incluir la identificación de datos faltantes, la aplicación de cierto tipo de filtros, entre otros. Esta preparación varía dependiendo del modelo a aplicar y es de vital importancia para el exitoso desarrollo de la predicción en estudio.
2. **Graficar los datos:** Este paso permite visualizar y con esto entender los datos disponibles. Ver la información en una gráfica, permite identificar patrones comunes y con esto, seleccionar apropiadamente el modelo a utilizar.
3. **Definir el modelo:** En este paso se debe escoger y especificar el modelo a utilizar con base en el set de datos disponibles. De la correcta selección del modelo, dependerá el desempeño y exactitud de la predicción. El presente trabajo de investigación explora diferentes modelos y, de acuerdo con la revisión bibliográfica, estima que el modelo ARIMA es apropiado para la predicción de los precios horarios de la electricidad.
4. **Entrenar el modelo:** En esta etapa, se utiliza parte de los datos disponibles para entrenar el modelo. Con esto, ya es posible ajustar el modelo y obtener unas predicciones preliminares.
5. **Evaluar el desempeño del modelo:** Una vez el modelo ha sido ajustado, es fundamental aplicar alguna de las técnicas disponibles para la evaluación del desempeño del modelo.
6. **Generar la predicción:** Una vez completados los pasos anteriores, es posible correr el modelo para generar una predicción. En este punto, es necesario especificar el número de observaciones futuras a predecir.

3.1 TIPO DE ESTUDIO

El presente estudio es una investigación de tipo cuantitativo basados en el planteamiento de [58], que la define como el “conjunto de métodos de recopilación e interpretación de datos que se utiliza para estudiar fenómenos... Se basa principalmente en la estadística y las matemáticas, utilizando estas herramientas y otras similares para crear hipótesis y modelos teóricos sobre aquello que se está investigando.” y que tiene como principal objetivo “analizar grandes cantidades de datos para poder construir modelos y teorías a partir de ellos”.

Igualmente se puede considerar como una investigación aplicada, de acuerdo con [58] que la define como aquella cuyo “énfasis del estudio está en la resolución práctica de problemas. Se centra específicamente en cómo se pueden llevar a la práctica las teorías generales. Su motivación va hacia la resolución de los problemas que se plantean en un momento dado. El objetivo de la investigación aplicada es predecir un comportamiento específico en una situación definida.”

3.2 MUESTREO

La técnica de muestreo utilizado corresponde a una de carácter no probabilístico entendido como “aquel donde no todos los sujetos de la población estadística tienen la misma probabilidad de ser elegidos para formar parte del estudio que se está desarrollando” [67]

Las 8 regiones en las cuales se divide el Sistema Interconectado Nacional (SIN), figuras 3.1, 3.2, 3.3, 3.4, se sometieron a un análisis de correlación y se identificó que, para el periodo de cuatro años analizado (1ro de enero de 2018 a 29 de marzo de 2022), con una frecuencia horaria, cinco de ellas presentaban correlaciones 1:1 tal y como se evidencia en la matriz de correlación de la figura 3.21a y en el gráfico de correlación con histogramas, funciones de densidad, líneas de regresión y coeficientes de correlación con nivel de significancia del 1 % de la figura 3.22.

3.3 ESTRUCTURA DE LOS DATOS

Los datos que se analizan de este trabajo presentan la estructura que se resume en la tabla 3.1.

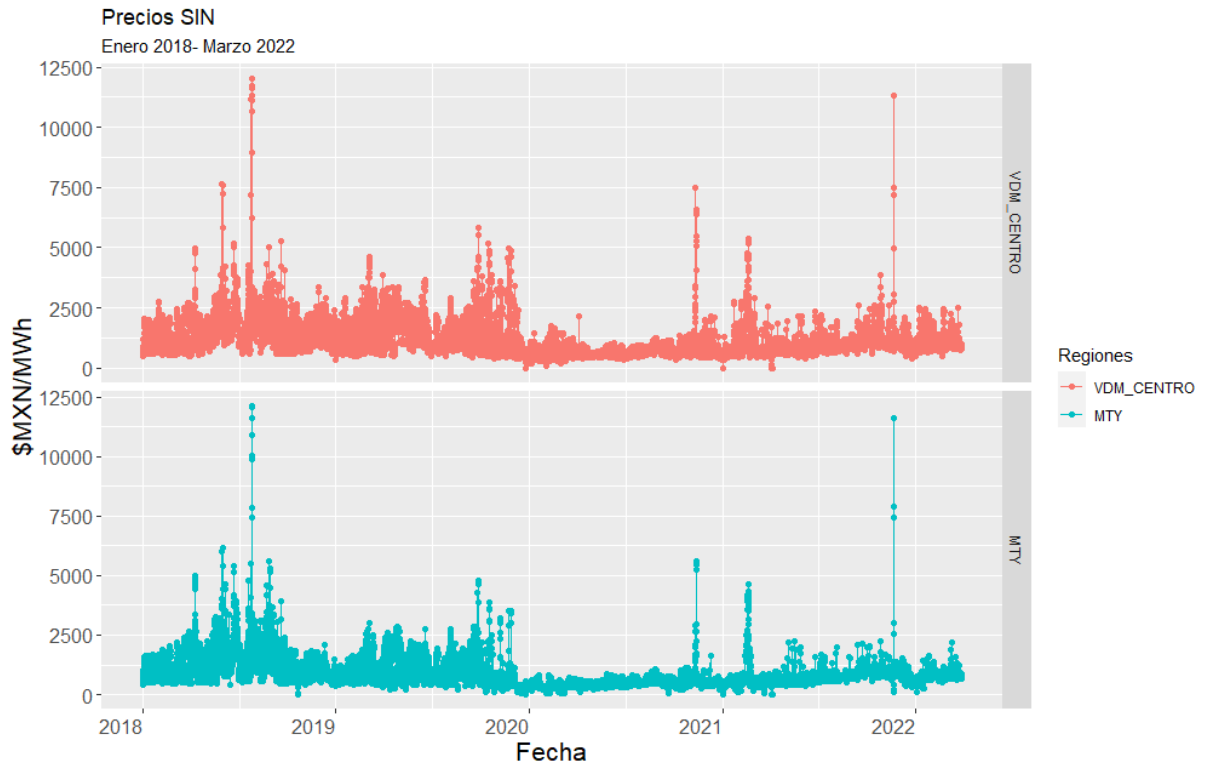


Figura 3.1: Región Valle de México Centro y Monterrey como una serie de tiempo

3.3.1 DATOS ATÍPICOS

Una forma de identificar los datos atípicos a partir del análisis de datos que se ha hecho es la descomposición de los datos.

$$x_t = M_t + S_t + R_t$$

Como se explico anteriormente, se debe remover el componente periódico y la tendencia de la serie, para así encontrar los datos atípicos de la serie de residuos R_t [27].

Fuerza de la periodicidad

La fuerza de la periodicidad en una serie temporal se refiere a cuán prominente o evidente es la repetición de patrones a lo largo de ciertos intervalos de tiempo, es decir, cuán claramente se observa la presencia de una periodicidad en los datos. En el análisis de series temporales, la fuerza de la periodicidad se evalúa a menudo mediante el uso de métodos como el periodograma, la autocorrelación y modelos específicos como el modelo aditivo de descomposición estacional (ADAMS).

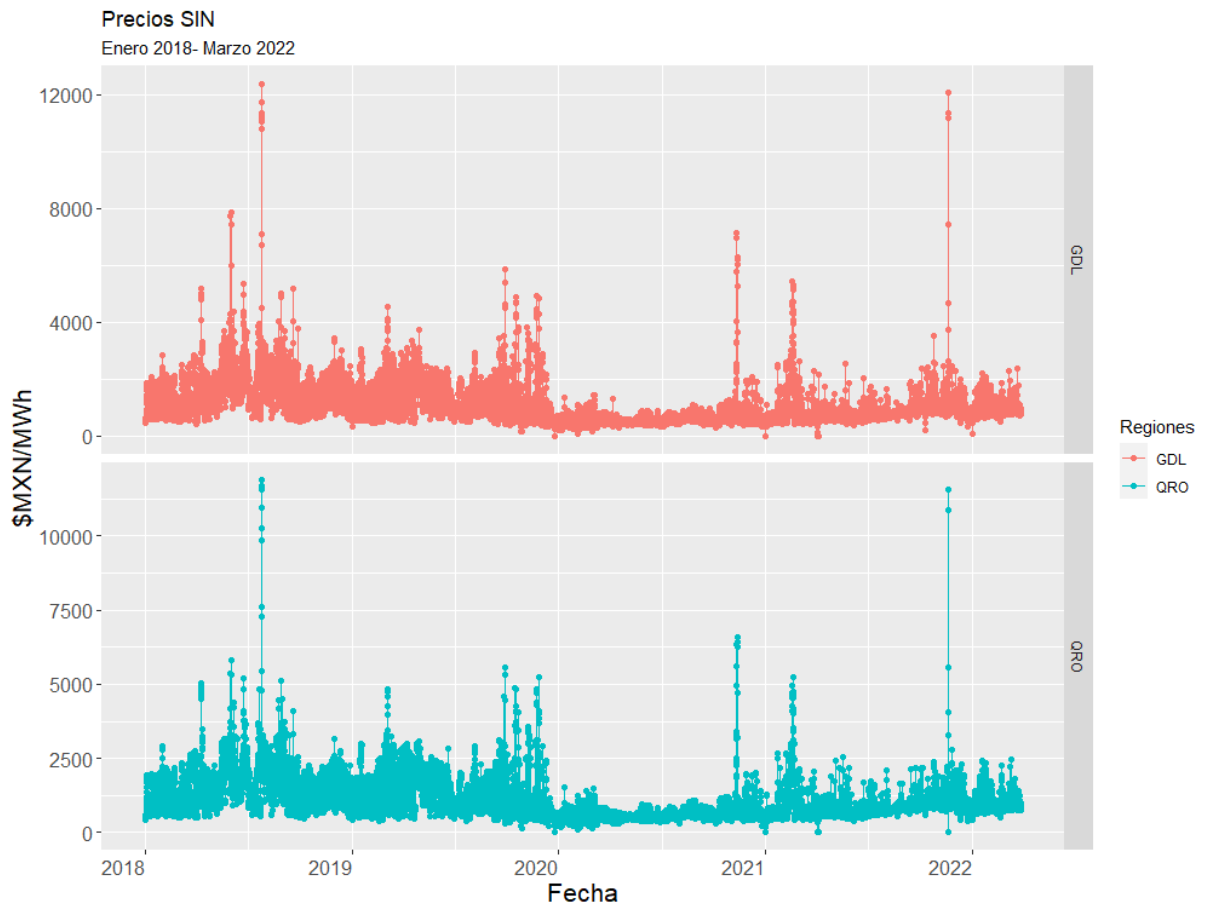


Figura 3.2: Región Guadalajara - Querétaro como una serie de tiempo

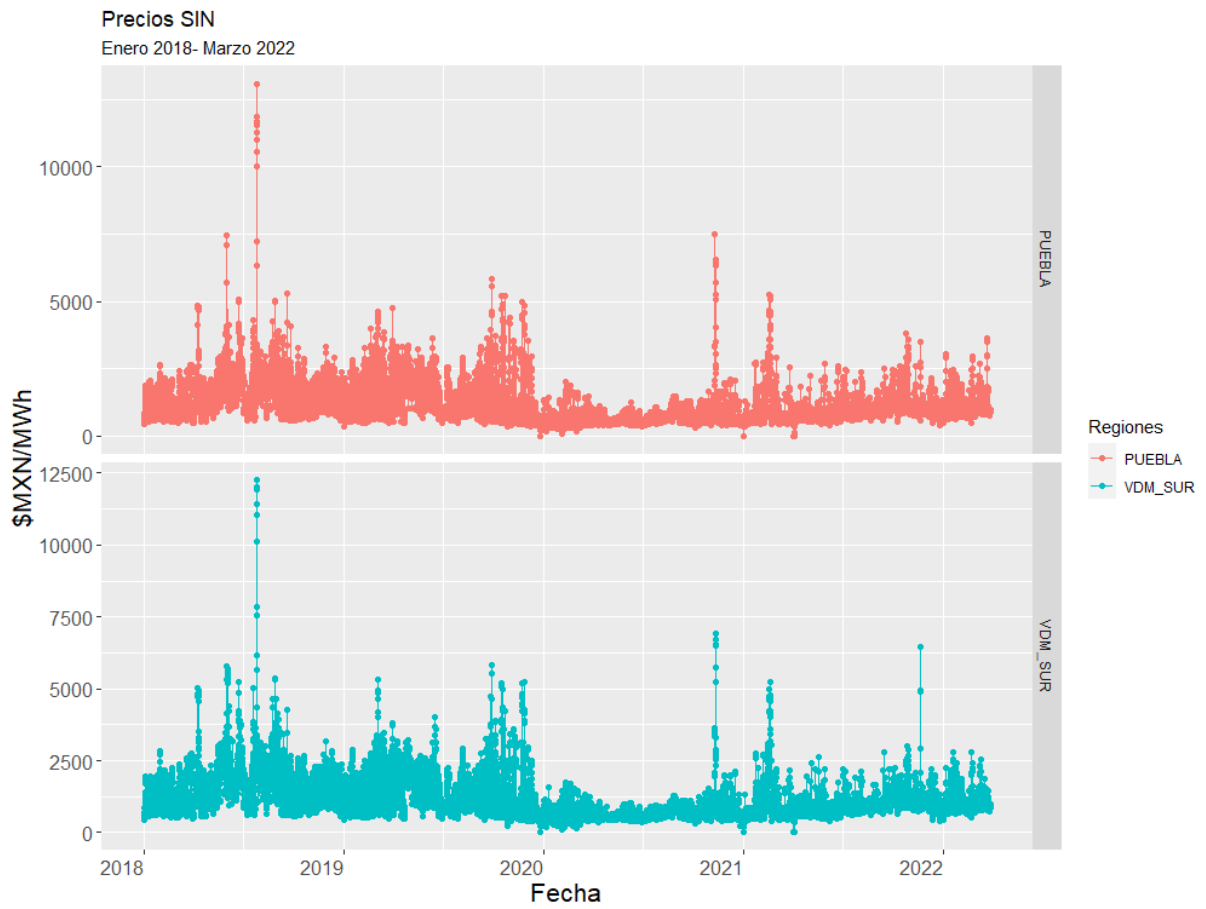


Figura 3.3: Región Valle de México Sur y Puebla como una serie de tiempo

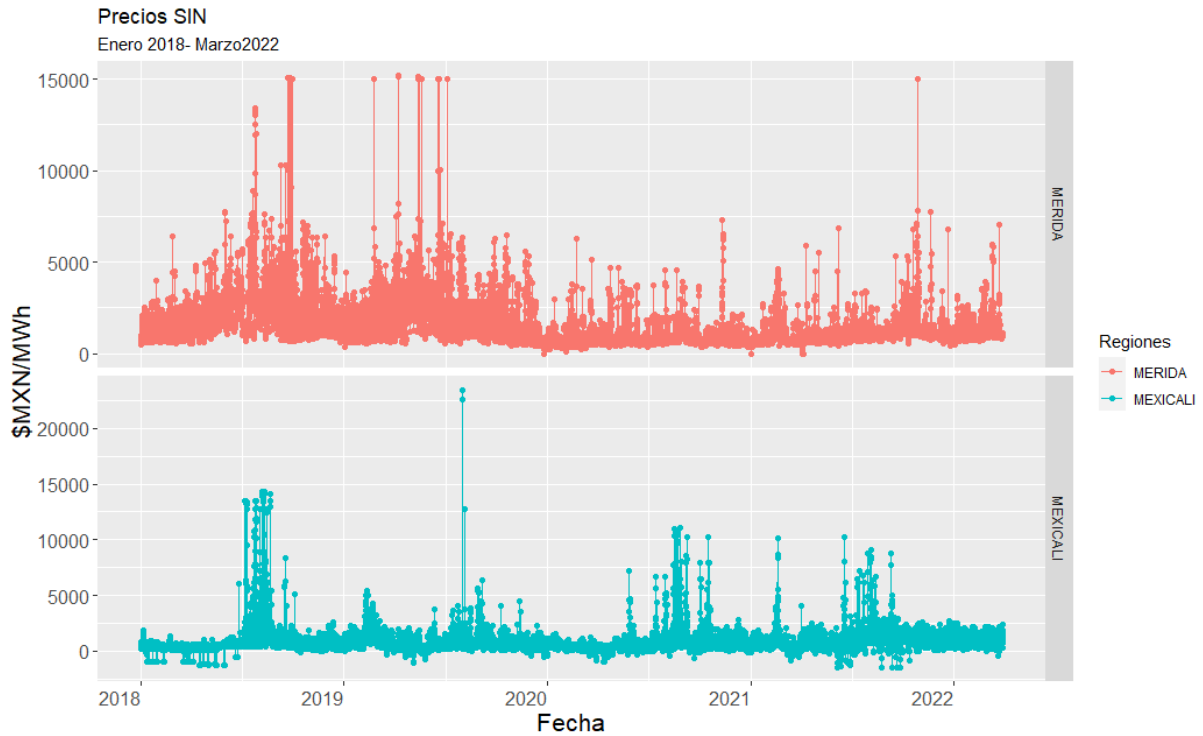


Figura 3.4: Región Mérida y Mexicali como una serie de tiempo

Un alto valor en el periodograma en una frecuencia específica indica una fuerte periodicidad en esa frecuencia en los datos. De manera similar, en un gráfico de autocorrelación, un patrón repetitivo de valores altos en ciertos rezagos indica una fuerte periodicidad. La fuerza de la periodicidad puede variar en diferentes series temporales y puede influir en la elección de métodos de modelado apropiados.

En resumen, la fuerza de la periodicidad en una serie temporal se refiere a cuán evidente y pronunciada es la repetición de patrones a intervalos regulares, y se evalúa mediante análisis visuales y técnicas cuantitativas.

Para ello se calcula la fuerza de la periodicidad usando,

$$F_s = 1 - \frac{Var(x_t - \hat{M}_t - \hat{S}_t)}{Var(x_t - \hat{M}_t)}$$

La ecuación se utiliza para decidir si se debe realizar un ajuste a la serie temporal. Si $F_s > 0.6$, se realiza un ajuste a la serie aplicando la fórmula $R_t = x_t^* - \hat{M}_t$, donde x_t^* es el valor ajustado de la serie.

Donde:

Tabla 3.1: Resumen de la base de datos utilizada

Características				
Regiones	NA	[%] NA	Atípico	[%] Atípico
VDM Centro	28	0.075	547	1.472
Monterrey	28	0.075	1880	5.060
Guadalajara	28	0.075	750	2.018
Queretaro	28	0.075	730	1.965
Puebla	28	0.075	557	1.499
VDM Sur	28	0.075	564	1.518
Mérida	28	0.075	1305	3.512
Méxicali	75	0.202	1490	4.010

- F_s es la fuerza de la periodicidad.
- $\text{Var}(x_t - \hat{M}_t - \hat{S}_t)$ es la varianza del residuo después de ajustar la serie.
- $\text{Var}(x_t - \hat{M}_t)$ es la varianza de la serie original después de ajustar la tendencia y la estacionalidad.
- x_t es el valor de la serie en el tiempo t.
- \hat{M}_t es una estimación de la tendencia en el tiempo t.
- \hat{S}_t es una estimación de la estacionalidad en el tiempo t.

El valor de 0.6 utilizado en la fórmula tiene una amplia aceptación en el análisis de series temporales como un criterio para determinar si la fuerza de la estacionalidad es lo bastante marcada como para justificar la aplicación de un ajuste estacional.

En la tabla 3.2, se calculo para cada una de las regiones. De aquí se concluye que la estacionalidad juega un rol importante en el composición de la serie. La elección de 0.6 como umbral se basa en investigaciones empíricas [55] y se ha consolidado como una guía útil para evaluar si la estacionalidad tiene una influencia sustancial en la serie de datos. Si la relación entre la variabilidad de la serie original después de eliminar la componente estacional y la variabilidad después de eliminar únicamente la tendencia supera el valor de 0.6, se considera que la estacionalidad tiene un impacto relevante en la serie.

$Q1$ Denota el percentil 25 y $Q3$ el percentil 75 de los residuos, ver la figura 3.5 por lo tanto el rango intercuartílico se puede definir como $IQR = Q3 - Q1$. De acuerdo con [27]¹ una observación se clasifica como atípica si es menor que $Q1 - 3XIQR$ o mayor que

¹Referencia el libro: Exploratory Data Analysis, John W. Tukey, 1977, p44. Donde Tukey propone para para los diagramas de cajas los valores atípicos.

Región	Fuerza de la Periodicidad
VDM_CENTRO	0.677
MTY	0.599
GDL	0.674
QRO	0.670
PUEBLA	0.650
VDM_SUR	0.675
MERIDA	0.649
MEXICALI	0.672

Tabla 3.2: Fuerza de la periodicidad por región

$Q3 + 3XIQR$. Si se los residuos están normalmente distribuidos, la probabilidad de una observación ser clasificada como atípica es de 1 en 427000 de acuerdo con [27]. Finalmente, los datos atípicos identificados de esta forma son reemplazados por valores interpolados linealmente, usando los vecinos cercanos.

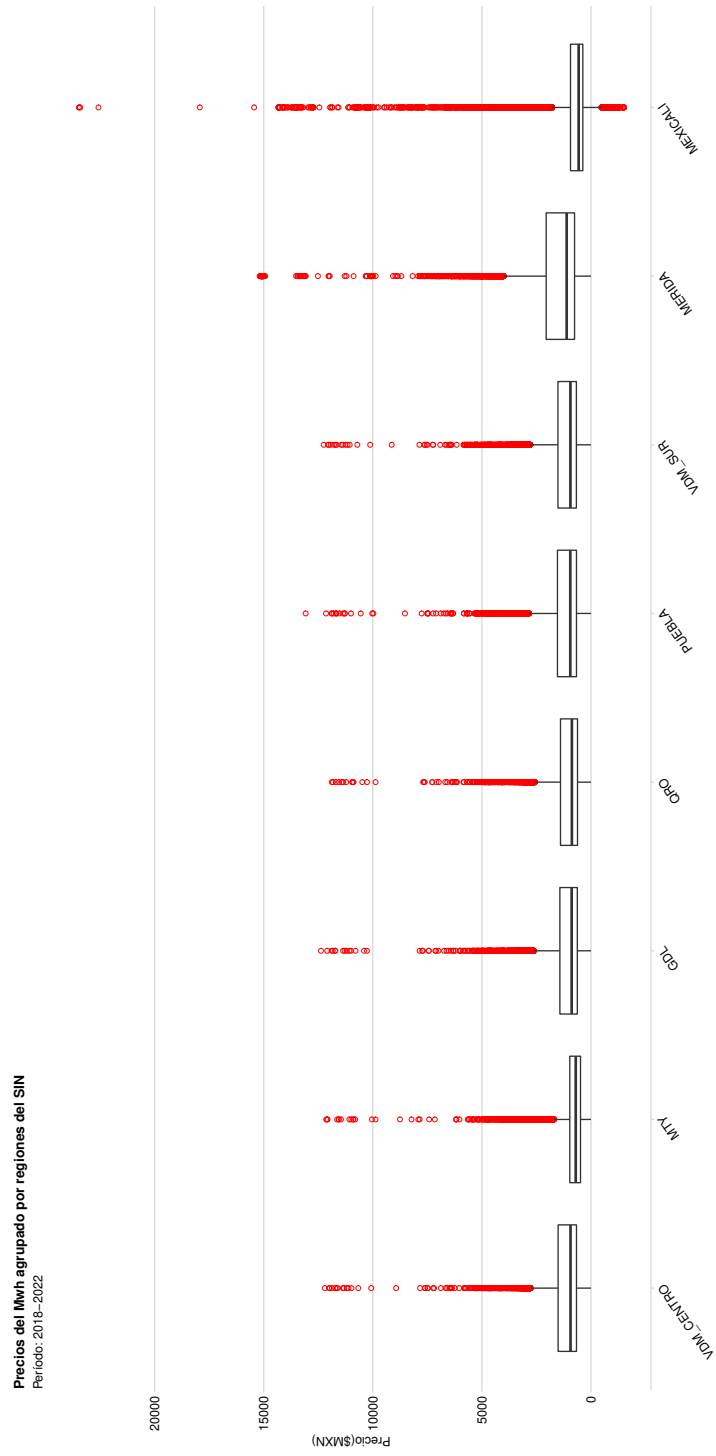


Figura 3.5: Diagrama de cajas para los datos analizados. Los puntos rojos indican los valores atípicos.

Se aplicó el procedimiento descrito anteriormente para eliminar los datos que se puedan considerar como atípicos, ver figuras 3.11, 3.12, 3.13, 3.14, 3.16, 3.15, 3.17 y 3.18. Los puntos azules, son aquellas horas en que se consideraron precios atípicos, y los puntos rojos (muy pocos) aquellos en donde no existían datos ver tabla 3.1.

Loess

La fórmula matemática que respalda el método de suavizado Loess (Local Polynomial Regression Fitting) se puede presentar en términos generales de la siguiente manera:

$$y_i = \sum_{j=1}^n w_{ij} \cdot y_j$$

Donde:

- y_i representa el valor observado en el punto i .
- y_j son los valores observados de los puntos vecinos en el conjunto de datos.
- w_{ij} son los pesos asignados a cada punto vecino j en función de su cercanía al punto i .
- n es el número total de puntos en la vecindad de i .

La clave en esta fórmula reside en cómo se determinan los pesos w_{ij} basados en la distancia entre los puntos i y j . Estos pesos se calculan de manera que los puntos cercanos a i tengan una contribución más significativa, mientras que los puntos más lejanos tengan una influencia reducida, lo que da como resultado un ajuste localizado.

En el contexto del suavizado Loess, se emplean comúnmente funciones kernel, como el kernel tricúbico o el kernel triangular, para calcular los pesos w_{ij} de tal forma que los puntos cercanos reciban un mayor peso. La elección de la función kernel influye en la distribución de los pesos en función de la distancia.

La fórmula proporcionada encapsula la esencia del método de suavizado Loess. Ver 3.6, 3.7, 3.8 y 3.9, en donde se emplea un kernel tricúbico.

El kernel tricúbico es una función matemática utilizada en estadísticas y métodos de suavizado no paramétrico, como el suavizado Loess. Su principal propósito es asignar pesos a los puntos de datos en función de su distancia a un punto central. Esto permite realizar un ajuste suavizado a los datos, enfatizando los valores cercanos y atenuando la

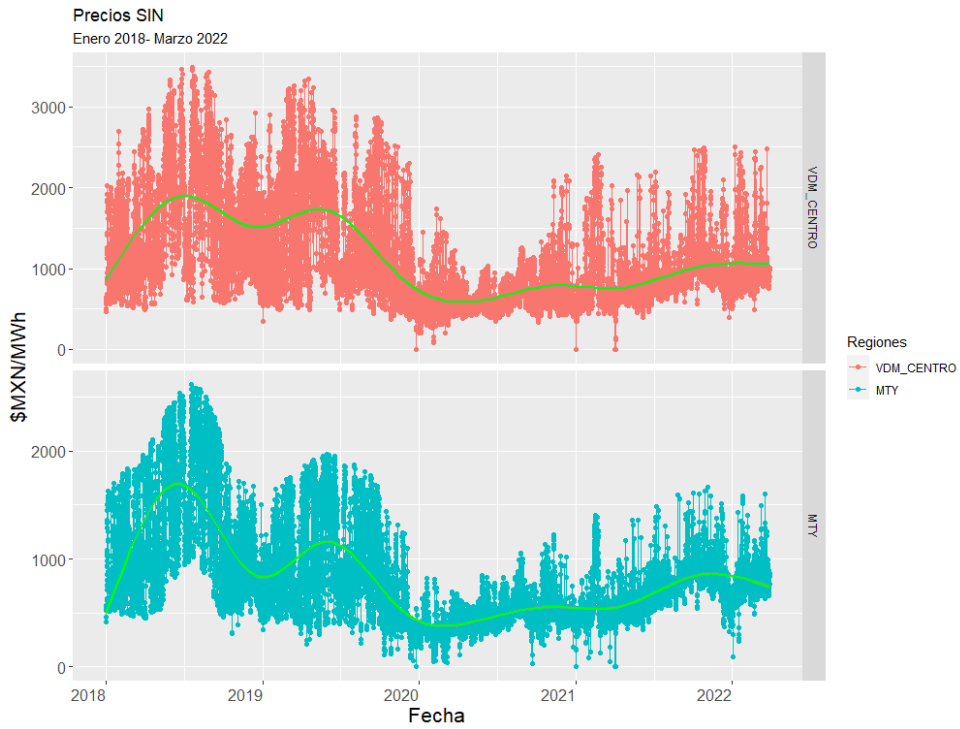


Figura 3.6: Región Valle de México Centro y Monterrey como serie de tiempo con tendencia y sin datos atípicos

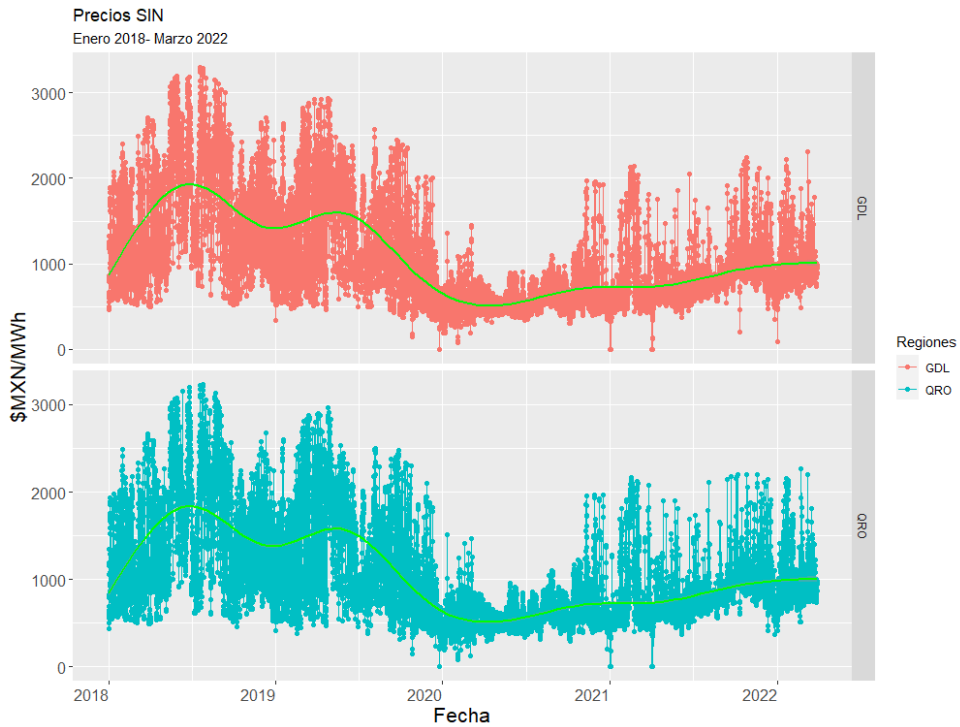


Figura 3.7: Región Guadalajara y Querétaro como serie de tiempo con tendencia y sin datos atípicos

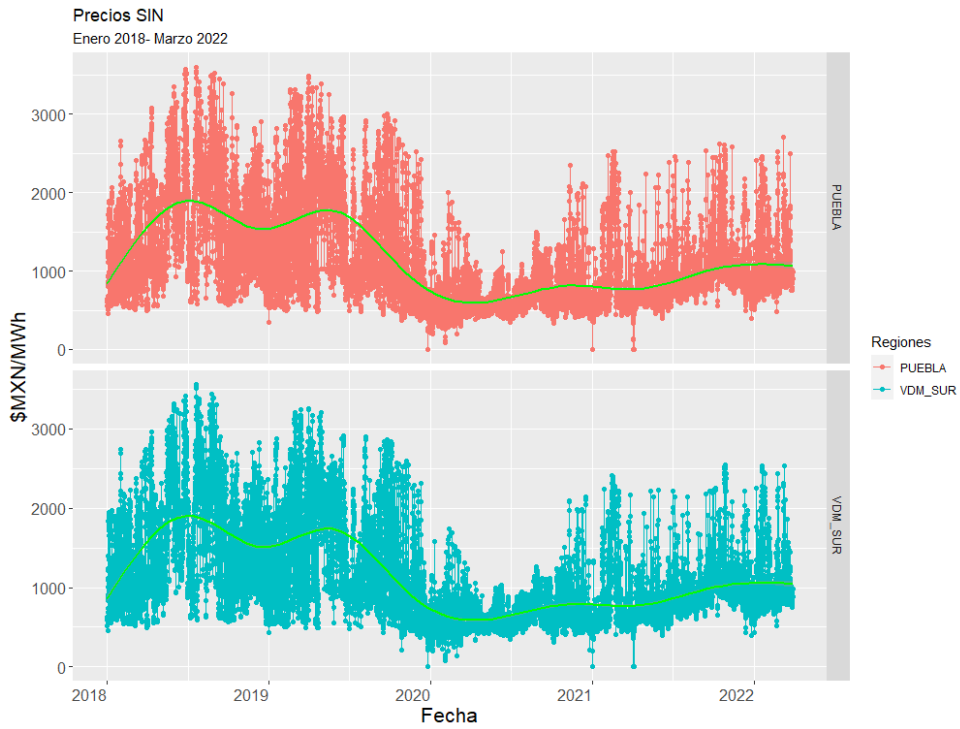


Figura 3.8: Región Valle de México Sur y Puebla como serie de tiempo con tendencia y sin datos atípicos

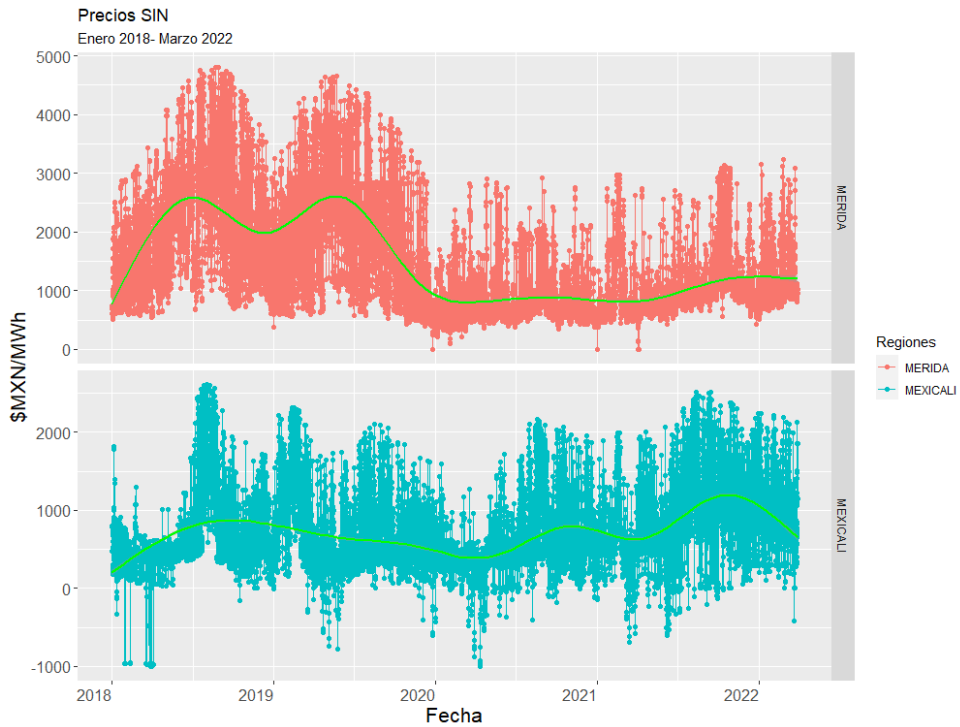


Figura 3.9: Región Valle de Mérida y Mexicali como serie de tiempo con tendencia y sin datos atípicos

influencia de los valores más distantes. El término "tricúbico" hace referencia al hecho de que la función se basa en un polinomio de tercer grado.

La función kernel tricúbico tiene dos componentes clave:

1. **Cálculo de pesos:** Asigna pesos a los puntos vecinos en función de su proximidad al punto central. Los puntos más cercanos tienen un peso mayor, mientras que los puntos más lejanos tienen un peso menor. Esto asegura que los valores cercanos tengan una mayor influencia en el cálculo del valor suavizado.
2. **Decrecimiento rápido:** Los puntos a una distancia mayor que 1 del punto central tienen un peso de 0. Esto significa que los puntos muy lejanos no contribuyen al valor suavizado en ese punto. Este decrecimiento rápido es lo que permite que el suavizado sea localizado y se ajuste a patrones específicos en la vecindad de cada punto.

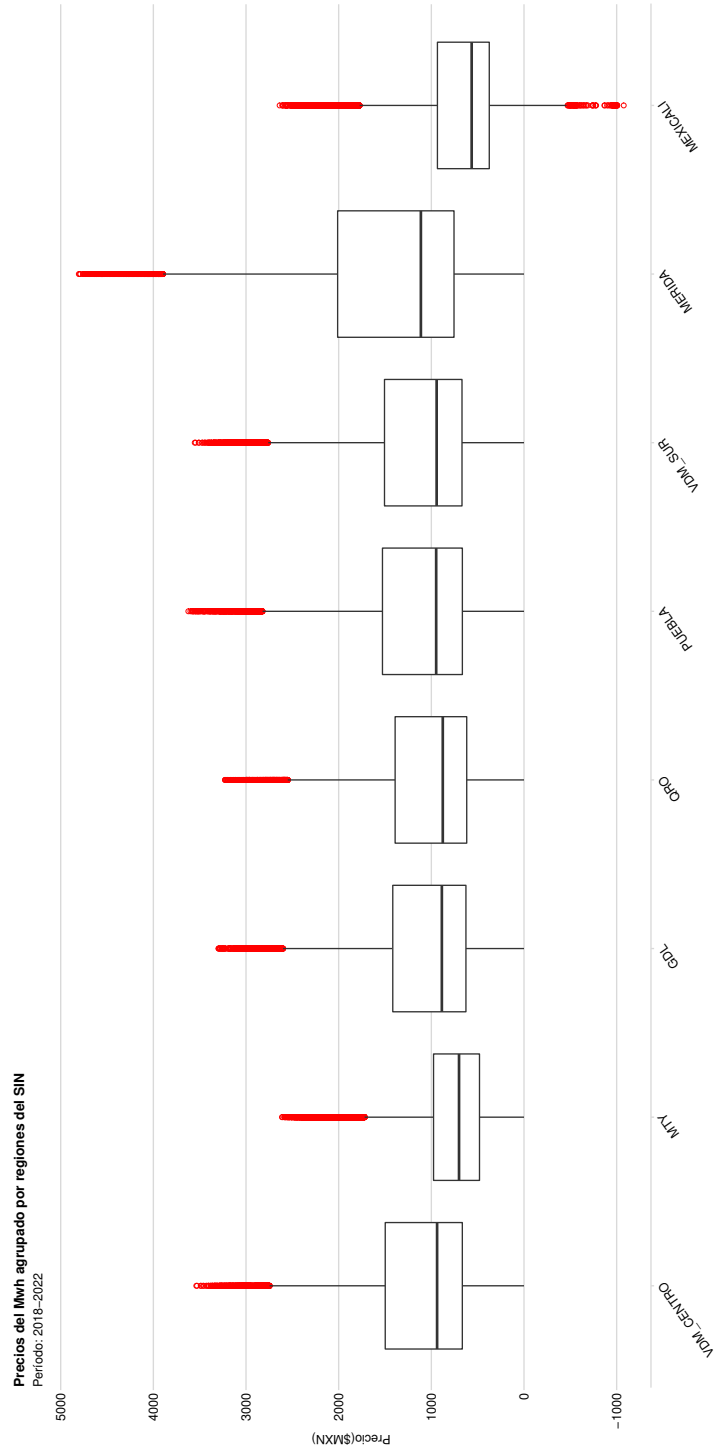


Figura 3.10: Diagrama de cajas para los datos analizados eliminando valores extremos. Los puntos rojos indican los valores atípicos.

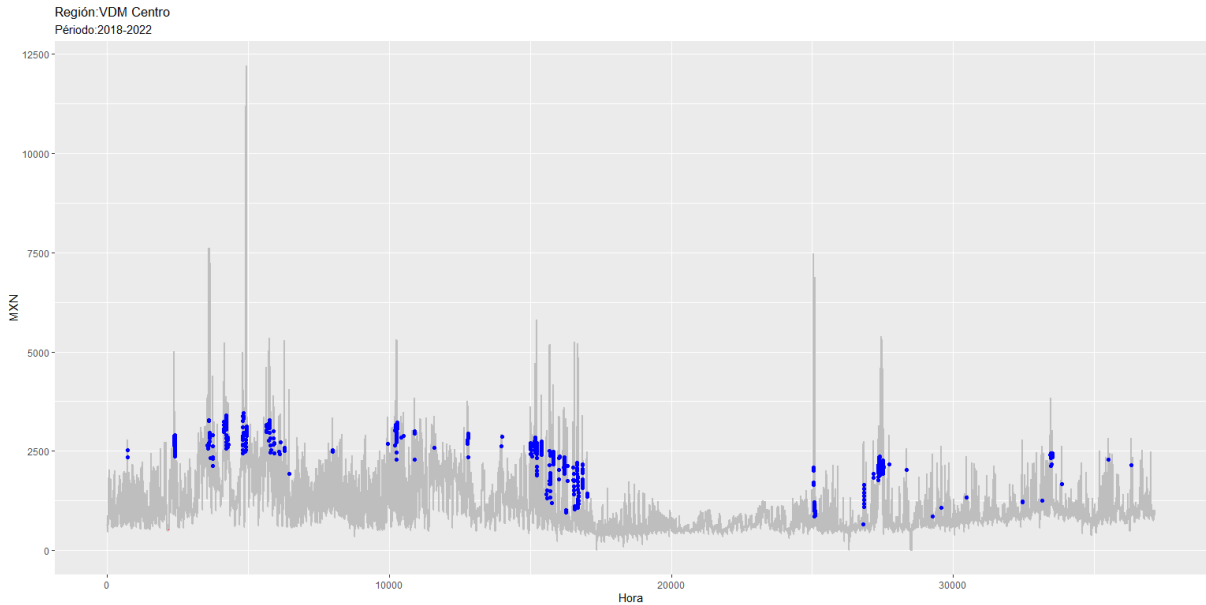


Figura 3.11: Valle de México Centro con valores atípicos y datos faltantes

La autocorrelación a largo plazo en una serie de tiempo se refiere a la relación entre los valores pasados y presentes de la serie con un retraso considerable, como días, semanas o incluso meses. Cuando existe autocorrelación a largo plazo, significa que los valores en un punto dado en el tiempo están correlacionados con los valores que ocurrieron mucho antes en el pasado. Esto puede indicar la presencia de patrones o tendencias que se repiten a lo largo del tiempo.

En el contexto de los precios de energía por región, existe autocorrelación a largo plazo, significa que los precios en un momento dado están influenciados por los precios que ocurrieron varias semanas o incluso meses atrás. Esta observación puede ser importante para tomar decisiones informadas, ya que te permite identificar patrones cíclicos o estacionales en los precios de energía.

En las figuras 3.19 y 3.20 donde se muestra la autocorrelación para todas las regiones, se observa picos significativos en los retrasos de aproximadamente 7 días (una semana), esto indica una autocorrelación semanal en los precios de energía. Esto podría sugerir que hay ciertos patrones recurrentes en los precios de energía que se repiten cada semana.

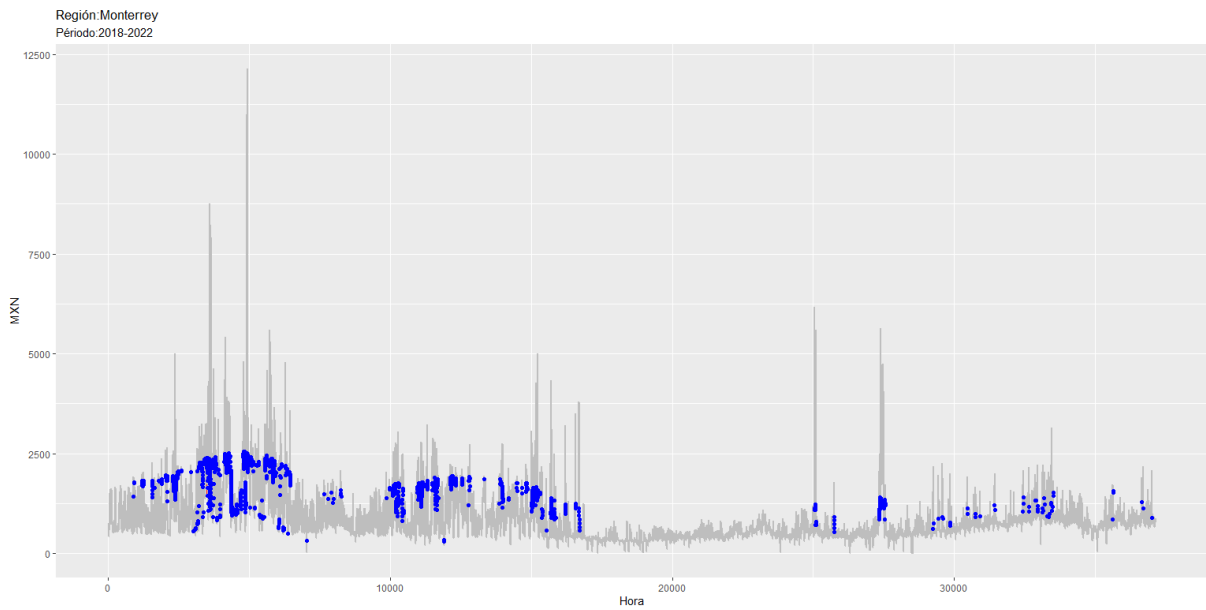


Figura 3.12: Monterrey con valores atípicos y datos faltantes

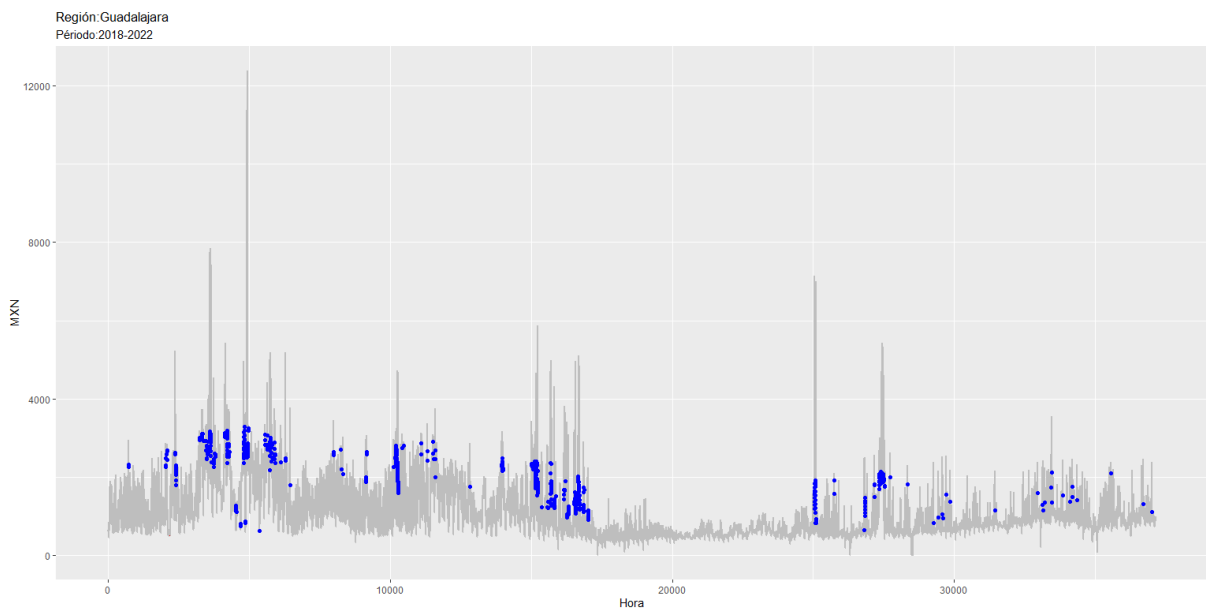


Figura 3.13: Guadalajara con valores atípicos y datos faltantes

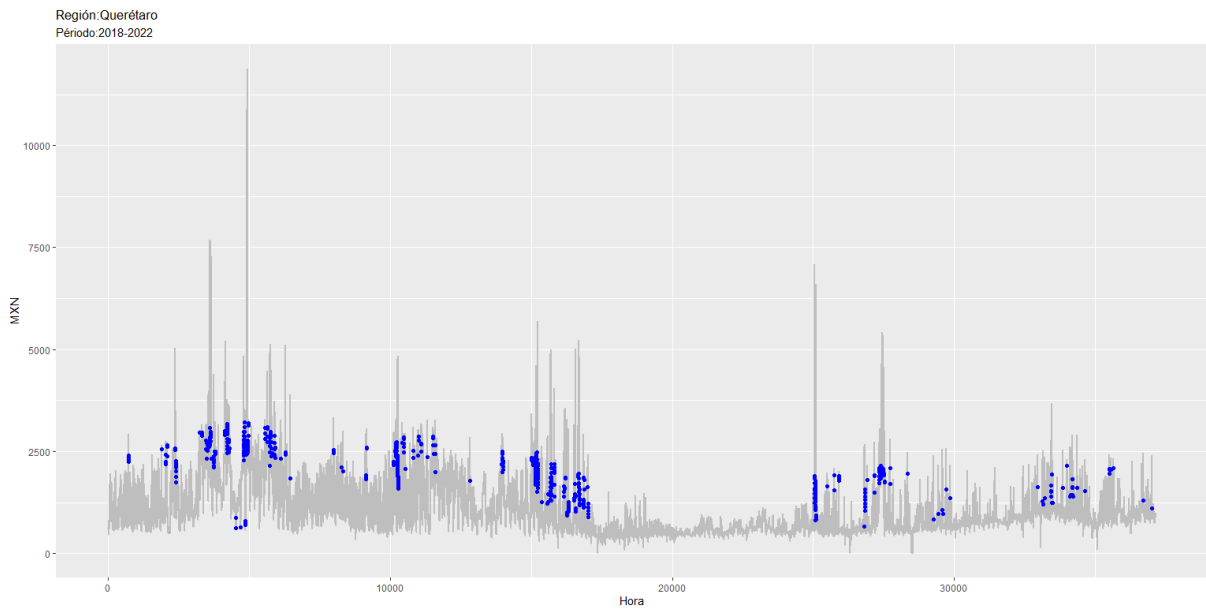


Figura 3.14: Queretaro con valores atípicos y datos faltantes

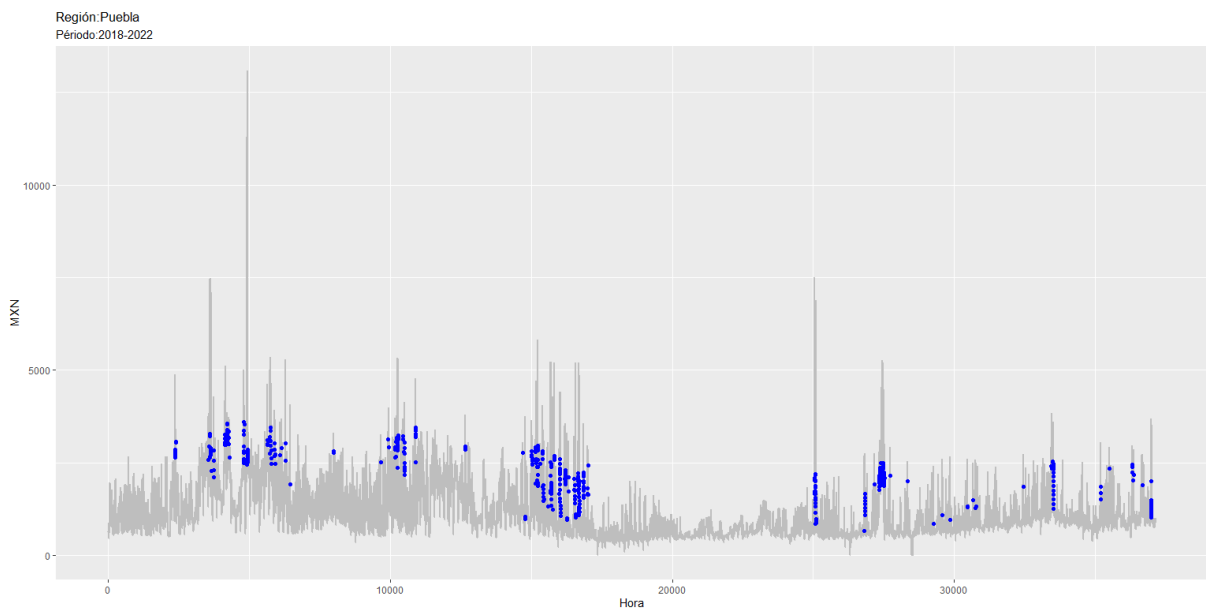


Figura 3.15: Puebla con valores atípicos y datos faltantes

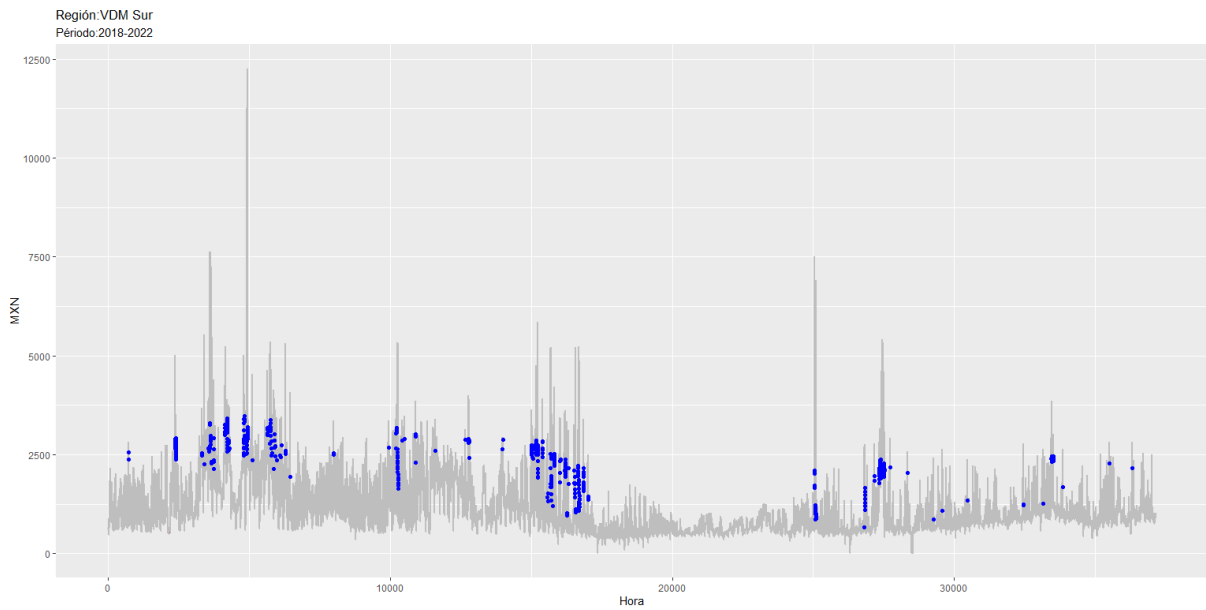


Figura 3.16: Valle de México Sur con valores atípicos y datos faltantes

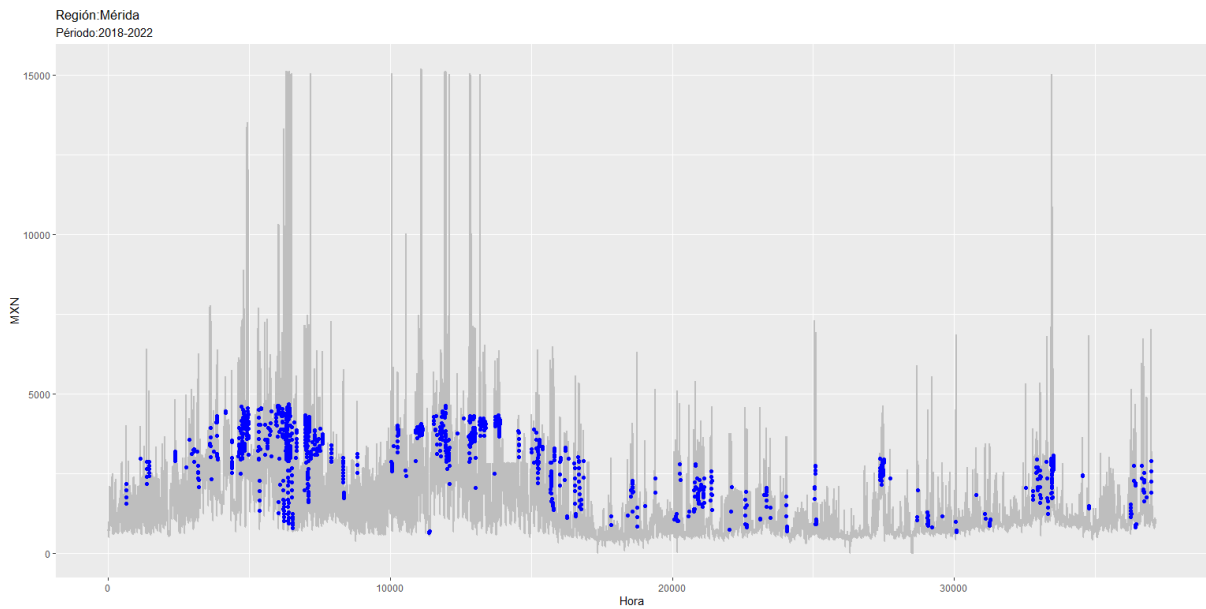


Figura 3.17: Mérida con valores atípicos y datos faltantes

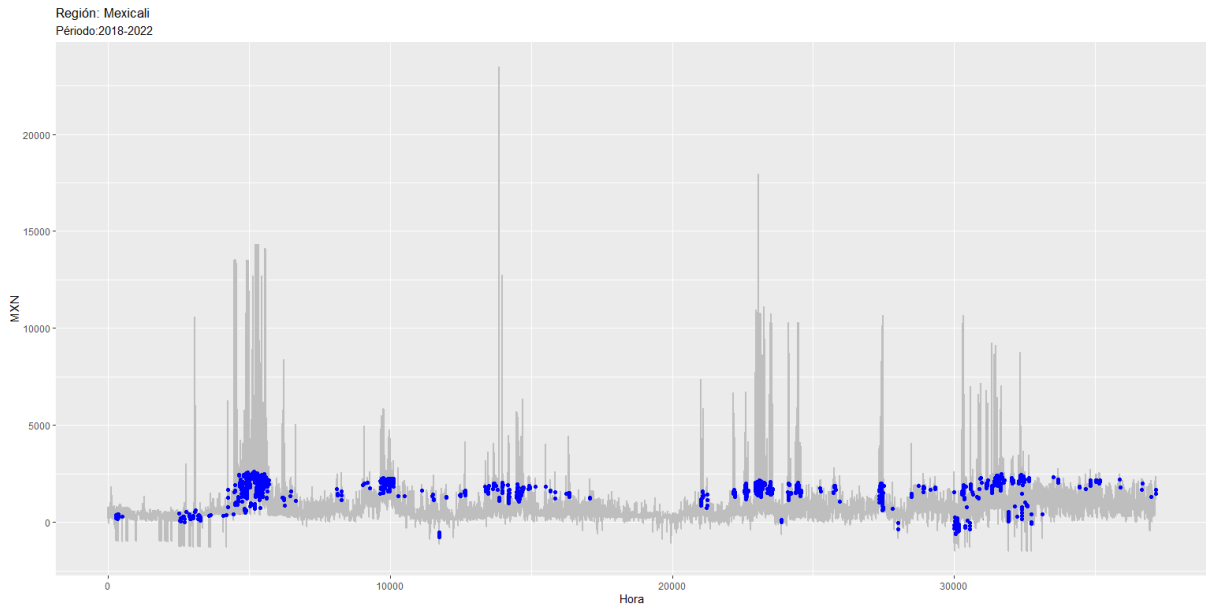


Figura 3.18: Mexicali con valores atípicos y datos faltantes

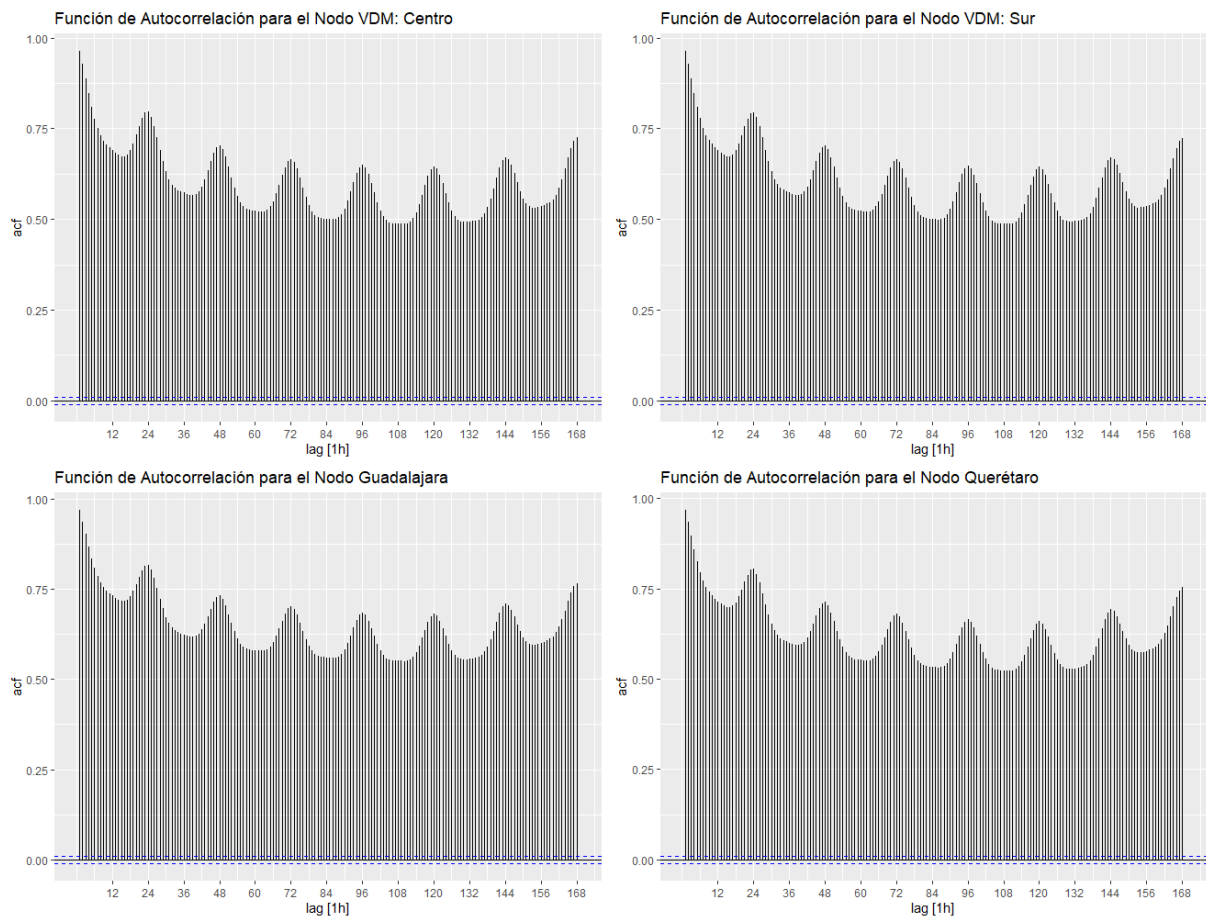


Figura 3.19: Función de Autocorrelación para las series de tiempo. Se observa autocorrelación significativa en los datos con un lag de 1 semana

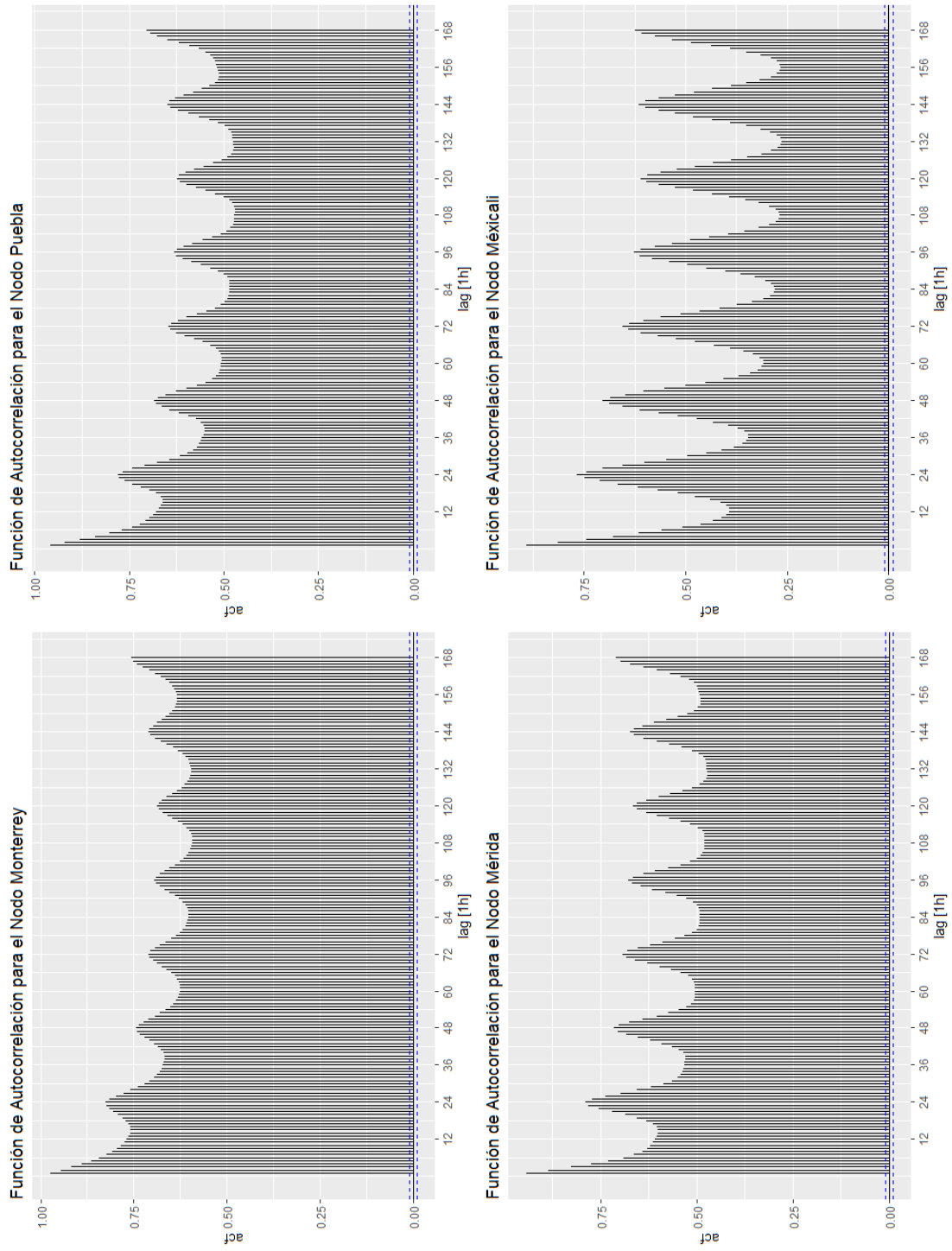


Figura 3.20: Función de Autocorrelación para las series de tiempo. Se observa autocorrelación significativa en los datos con un lag de 1 semana

3.3.2 CORRELACIÓN

La **correlación de Pearson** es una medida estadística que evalúa la relación lineal entre dos variables continuas. Es ampliamente utilizada para determinar el grado y la dirección de la asociación entre dos variables cuantitativas.

El coeficiente de correlación de Pearson, denotado como r , varía entre -1 y 1. Un valor de 1 indica una correlación positiva perfecta, lo que significa que a medida que una variable aumenta, la otra también lo hace en proporción constante. Un valor de -1 indica una correlación negativa perfecta, lo que significa que a medida que una variable aumenta, la otra disminuye en proporción constante. Un valor de 0 indica la ausencia de una relación lineal entre las variables.

El cálculo del coeficiente de correlación de Pearson se basa en las desviaciones de los valores individuales con respecto a sus respectivas medias. La fórmula general es:

$$r = \frac{\sum [(X_i - \bar{X})(Y_i - \bar{Y})]}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}}$$

Donde:

- r : Coeficiente de correlación de Pearson.
- X_i, Y_i : Valores individuales de las dos variables.
- \bar{X}, \bar{Y} : Medias de las dos variables.
- \sum : Sumatoria sobre todos los valores.

Es importante tener en cuenta que la correlación de Pearson solo mide la relación lineal entre las variables y no captura otras formas de relación no lineal. Además, la correlación de Pearson puede ser influenciada por valores atípicos o por la falta de una relación lineal verdadera entre las variables.

Por otro lado la **correlación de Spearman** es una medida estadística que evalúa la relación entre dos variables ordenadas o clasificatorias. La correlación de Spearman es una medida estadística que evalúa la relación entre dos variables ordenadas o clasificatorias. A diferencia de la correlación de Pearson, que evalúa la relación lineal entre variables continuas, la correlación de Spearman se utiliza para variables que pueden estar relacionadas de manera no lineal o cuando los datos se presentan en forma de rangos o clasificaciones.

La correlación de Spearman se basa en los rangos de los datos en lugar de en los valores reales. Primero, se asignan rangos a los datos de cada variable, donde el rango más bajo corresponde al valor más pequeño y el rango más alto al valor más grande. Luego, se calcula el coeficiente de correlación de Spearman utilizando la fórmula:

$$\rho = 1 - \frac{(6 * \sum d^2)}{n * (n^2 - 1)}$$

Donde:

- ρ : Coeficiente de correlación de Spearman.
- $\sum d^2$: Suma de los cuadrados de las diferencias entre los rangos de las dos variables.
- n : Tamaño de la muestra.

El coeficiente de correlación de Spearman puede tener valores entre -1 y 1. Un valor de 1 indica una correlación positiva perfecta, -1 indica una correlación negativa perfecta, y 0 indica la ausencia de correlación.

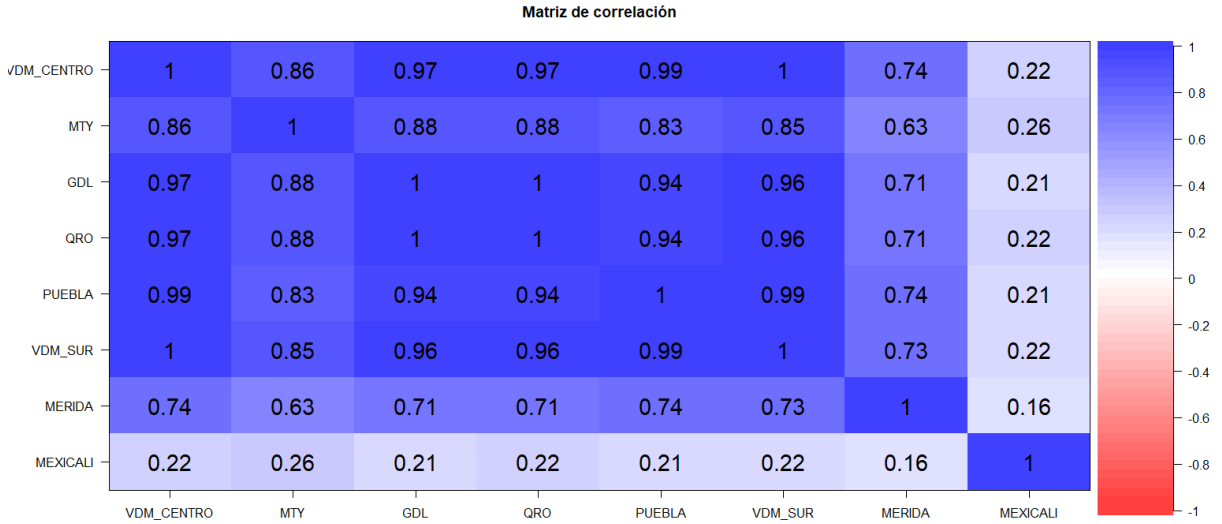
La correlación de Spearman es útil cuando los datos no siguen una distribución normal o cuando la relación entre las variables no es lineal.

La correlación de Spearman puede ser menos sensible a los datos atípicos en comparación con la correlación de Pearson, ya que se basa en los rangos de los datos en lugar de los valores reales. Sin embargo, es importante tener en cuenta que la presencia de datos atípicos puede afectar cualquier medida de correlación.

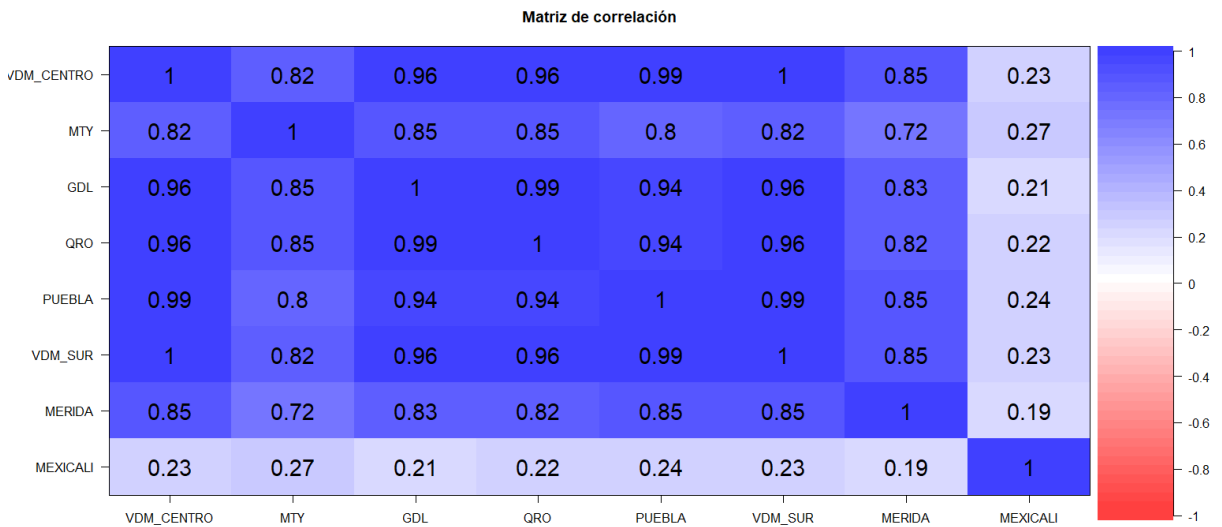
Si hay datos atípicos en la muestra, la correlación de Spearman puede verse afectada debido a los cambios en los rangos. Los datos atípicos extremos pueden tener un impacto considerable en los rangos y, por lo tanto, en el coeficiente de correlación de Spearman. En algunos casos, los datos atípicos pueden influir en la magnitud o la dirección de la correlación.

Cuando se analizan datos con presencia de valores atípicos, es recomendable tener precaución y considerar la interpretación de la correlación de Spearman en el contexto de los datos en general. Además, es importante examinar visualmente los datos y considerar otras medidas de asociación o realizar análisis más robustos si se sospecha que los datos atípicos están distorsionando los resultados.

En resumen, aunque la correlación de Spearman puede ser menos afectada por los datos atípicos en comparación con la correlación de Pearson, es importante tener en cuenta



(a) Matriz de correlación de los precios horarios de la energía eléctrica en las 8 regiones que conforman el SIN



(b) Matriz de correlación de los precios horarios de la energía, eliminando datos atípicos para las 8 regiones que conforman el SIN

que los valores atípicos pueden tener un impacto en cualquier medida de correlación y deben considerarse de manera apropiada en el análisis de los resultados.

Una correlación cercana a 1, como 0.94, 0.96 y 0.99, indica una relación muy fuerte y positiva entre las dos variables que se están comparando. En términos simples, significa que cuando una variable aumenta, la otra variable tiende a aumentar también, y cuando una variable disminuye, la otra variable tiende a disminuir en la misma dirección.

Más específicamente:

- Una correlación de 0.94 sugiere que hay una relación muy fuerte y positiva entre las dos variables. Los puntos en un gráfico de dispersión estarían muy cerca de una línea recta ascendente, lo que indica una relación lineal positiva bastante fuerte.
- Una correlación de 0.96 indica una relación aún más fuerte entre las variables. Los puntos en un gráfico de dispersión estarían aún más cerca de una línea recta ascendente.
- Una correlación de 0.99 señala una correlación extremadamente fuerte y prácticamente perfecta entre las variables. Los puntos en un gráfico de dispersión estarían prácticamente en una línea recta ascendente, lo que sugiere que las dos variables están muy relacionadas y tienden a moverse juntas en casi todas las observaciones.

En resumen, estas cifras de correlación cercanas a 1 indican una relación lineal positiva muy fuerte entre las variables. Sin embargo, es importante recordar que la correlación solo mide una relación lineal y no necesariamente implica causalidad. Otras variables o factores podrían estar influyendo en la relación observada.

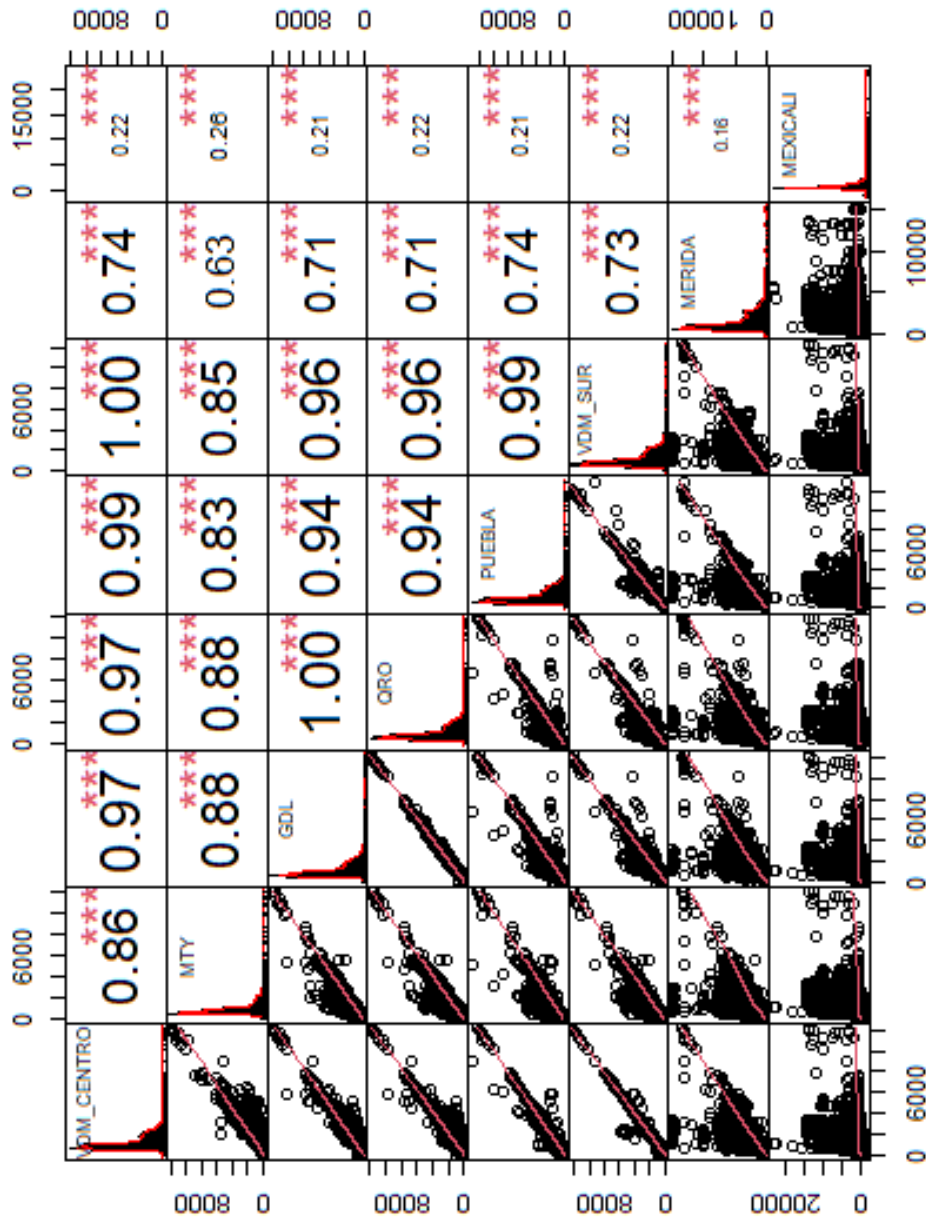


Figura 3.22: Gráfico de correlación con histogramas, funciones de densidad, líneas de regresión y coeficientes de correlación con nivel de significancia del 1 % para los precios horarios de la energía eléctrica en las 8 regiones del SIN

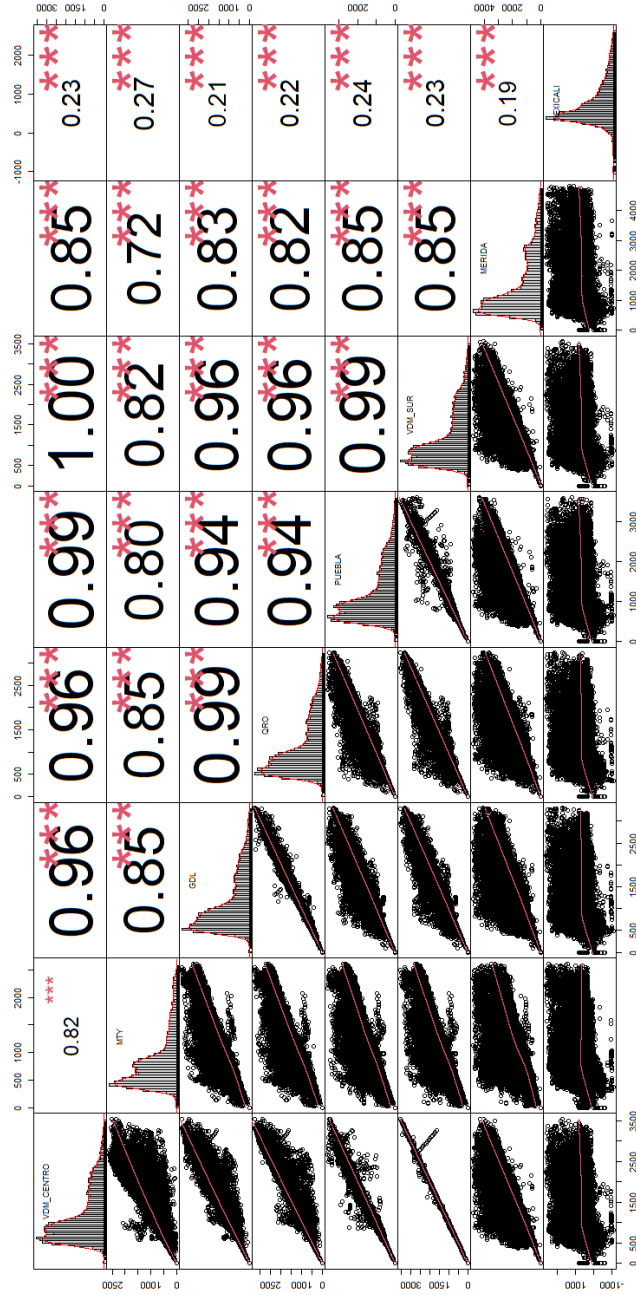


Figura 3.23: Gráfico de correlación con histogramas, funciones de densidad, líneas de regresión y coeficientes de correlación con nivel de significancia del 1 % para los precios horarios de la energía eléctrica en las 8 regiones del SIN, sin datos atípicos

Además, se identificó cualitativamente que las regiones que presentaban correlaciones, se encuentran geográficamente próximas y presentan cierta semejanza demográfica e industrial. Así pues, el presente estudio agrupará las regiones en los 5 grupos que se listan a continuación:

- Valle de México Centro, Sur y Puebla
- Guadalajara y Querétaro
- Monterrey
- Mérida
- Mexicali

Con base en lo anterior, se seleccionan 5 regiones representativas que son las que serán modeladas en el presente estudio. A estas regiones se les aplicó un análisis de dispersión x_t vs x_{t-k} , esto quiere decir que se toma la serie original y se gráfica contra la serie retrasada en k unidades de tiempo. Ver las figuras 3.24, 3.25, 3.26, 3.27 y 3.28 que corresponden a los gráficos de dispersión para las variables x_t vs x_{t-k} con el k^{th} retraso.

Los patrones observados en las gráficas de autocorrelación pueden estar relacionados con los comportamientos típicos de la serie de tiempo y la naturaleza de los datos. Las diferencias en la distribución de puntos en horas cercanas al mediodía y en la madrugada podrían explicarse por varios factores:

- **Ciclos Diarios:** Muchas series de tiempo, especialmente aquellas relacionadas con la actividad humana, como los precios de la energía, tienen patrones cíclicos diarios. Por ejemplo, es común que los precios de la energía sean más altos durante las horas pico de consumo durante el día y más bajos durante la noche. Esto puede crear patrones en la autocorrelación que reflejen estos ciclos diarios.
- **Comportamientos de Consumo:** Las horas cercanas al mediodía suelen ser momentos de mayor consumo de energía debido a la actividad industrial, comercial y doméstica. Esto podría influir en los patrones de autocorrelación, ya que los cambios en el consumo pueden influir en los precios.
- **Cambios en la Demanda:** Los patrones de consumo de energía pueden variar en diferentes momentos del día. Por ejemplo, la demanda de energía podría ser más baja durante la madrugada cuando la mayoría de las personas están durmiendo. Estos cambios en la demanda pueden llevar a patrones de autocorrelación específicos en ciertos momentos del día.

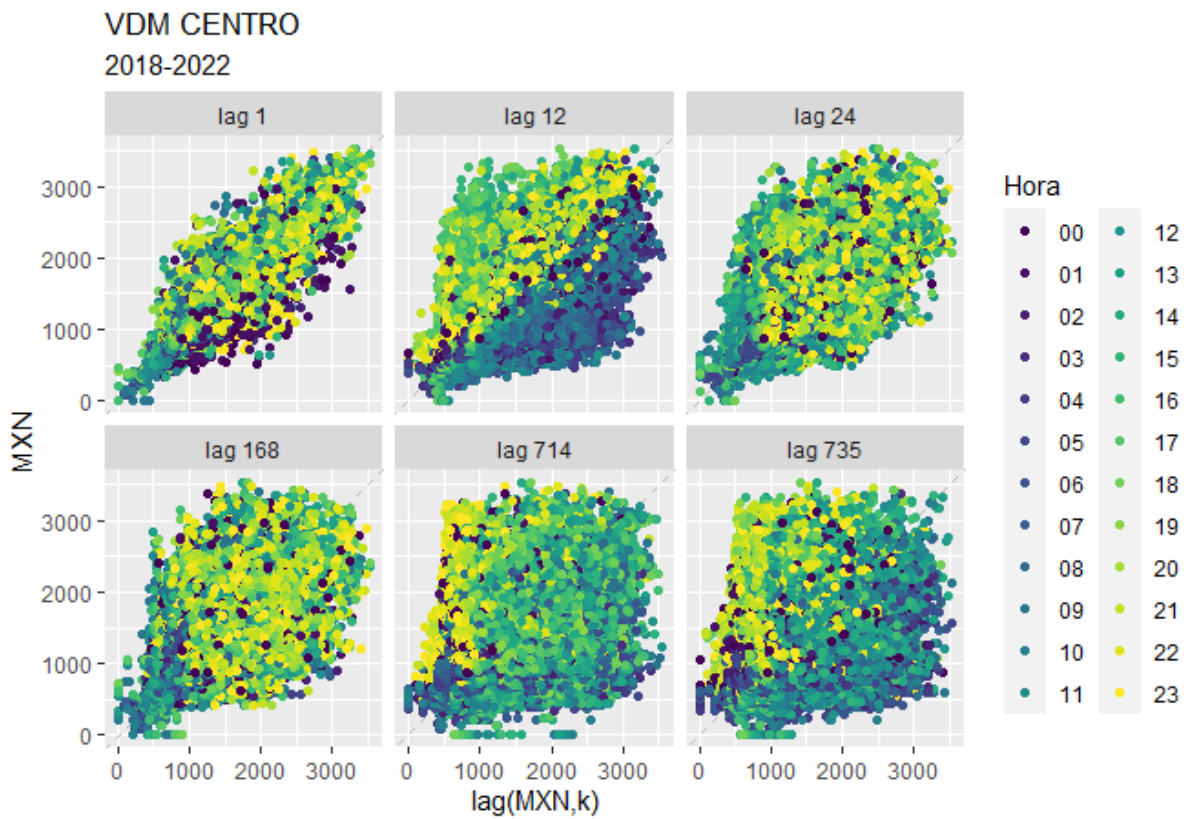


Figura 3.24: Gráfica de Autocorrelación por horas para el nodo VDM Centro y a diferentes retrasos

- Factores Externos:** Otros factores externos, como las condiciones climáticas, la disponibilidad de recursos energéticos y los eventos económicos, pueden influir en los precios de la energía. Estos factores podrían tener un impacto diferencial en diferentes momentos del día y podrían contribuir a los patrones observados en la gráfica.

En resumen, los patrones de autocorrelación en la serie de tiempo pueden reflejar los comportamientos cíclicos y las variaciones en la demanda que son típicos de los precios de la energía. Las diferencias en la distribución de puntos entre horas cercanas al mediodía y en la madrugada podrían estar relacionadas con cómo estos factores influyen en los precios en diferentes momentos del día.

Estas gráficas muestran la autocorrelación de la serie de tiempo de precios de energía para cada una de las cinco regiones seleccionadas. Se representan los diferentes retrasos en el eje x y los valores de autocorrelación en el eje y. Los puntos en la gráfica indican la relación entre el valor actual y los valores pasados de la serie de tiempo. Se observan

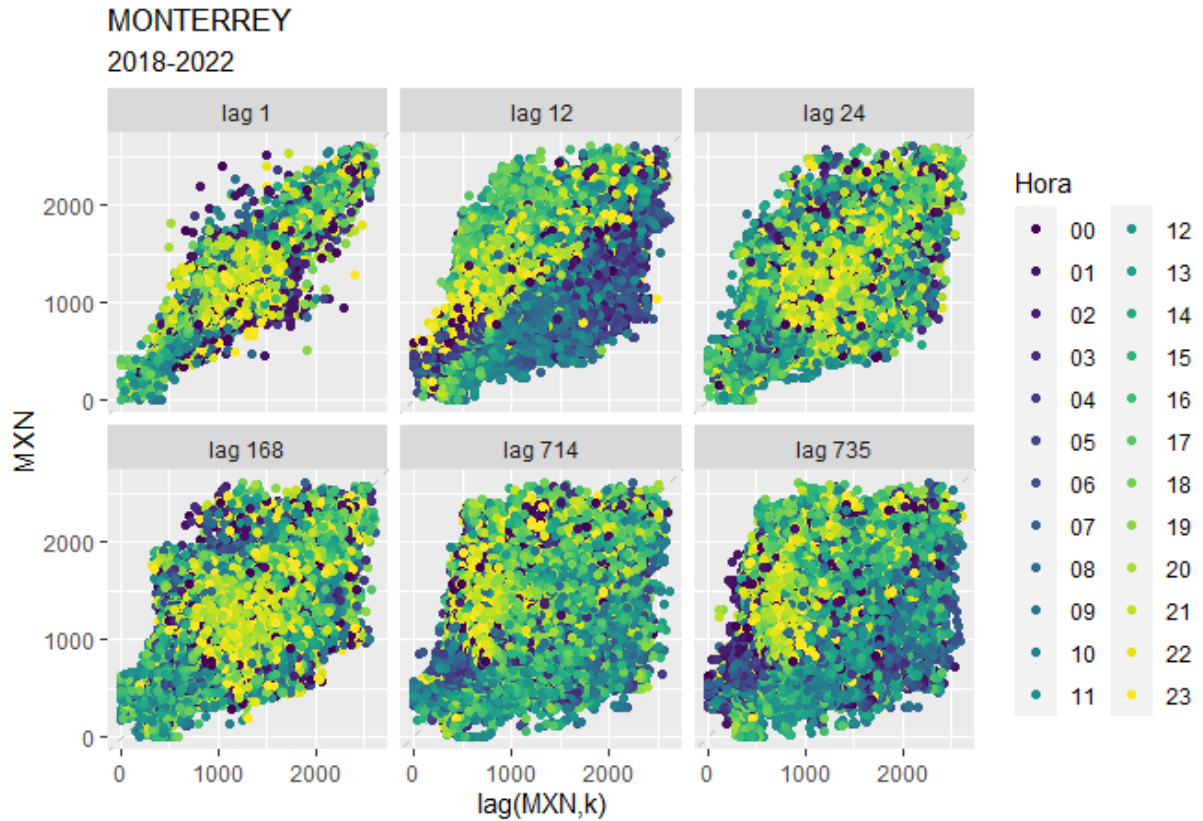


Figura 3.25: Gráfica de Autocorrelación por horas para el nodo Monterrey y a diferentes retrasos

patrones significativos en la autocorrelación a diferentes retrasos, lo que puede sugerir dependencias temporales en los precios de la energía en esta región. La variación en los patrones de autocorrelación en diferentes retrasos puede indicar la influencia de diferentes factores a lo largo del día y las estaciones del año en los precios de la energía durante el período de 2018 a 2022."

Adicionalmente, se descartan los datos del año 2020 debido al comportamiento atípico de los mismos dada la situación del COVID 19. Según [23] la pandemia tuvo impactos no lineales que afectaron tanto el ámbito económico como el energético en México. Durante este periodo, se evidenció una reducción considerable en la demanda y consumo de energía eléctrica con patrones de consumo diferentes en cantidad y fuentes.

Por otro lado, aunque el dataset disponible para precios horarios de energía en México comprende más de 37.000 datos recogidos durante un periodo mayor a 4 años, tal y como se especifica en la sección posterior, el presente estudio estima probar el modelo con un grupo de datos que comprenda al menos los de un año. Este periodo de tiempo y el conjunto de datos a considerar, se determinó con base en la recomendación de [33] en

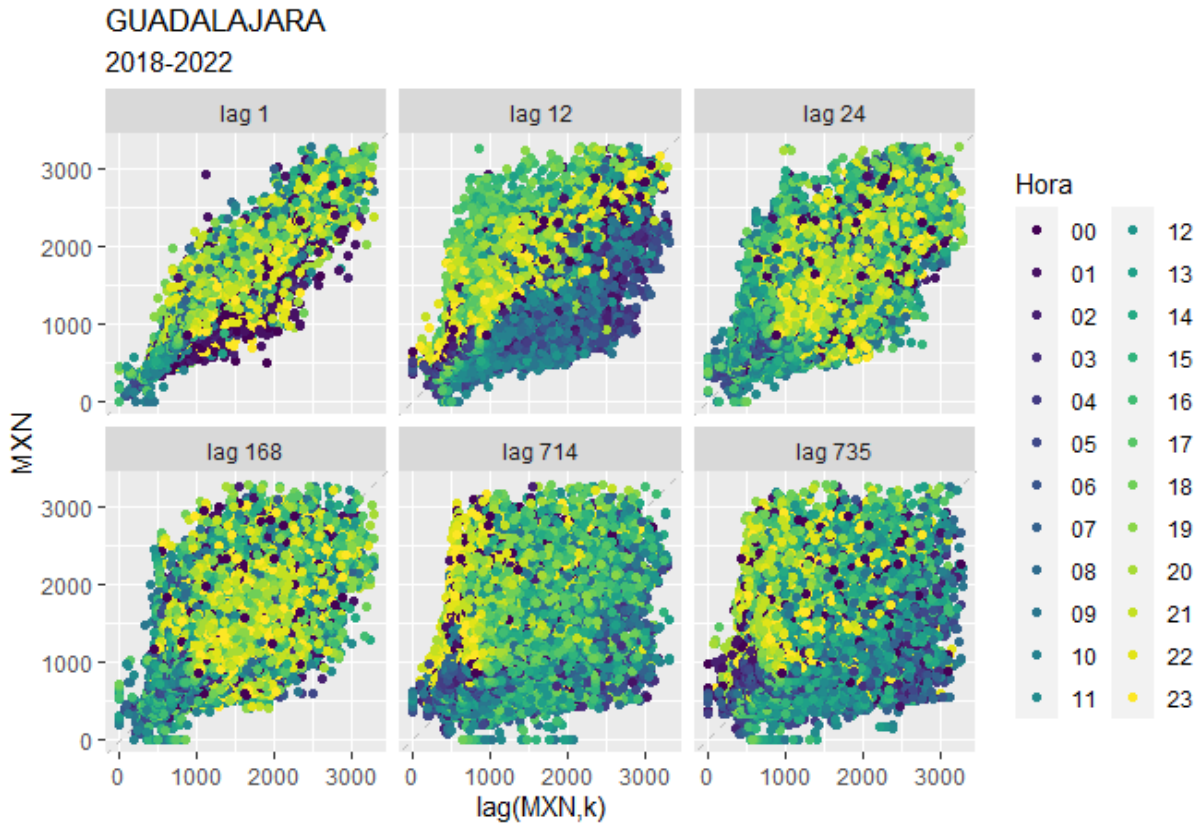


Figura 3.26: Gráfica de Autocorrelación por horas para el nodo Guadalajara y a diferentes retrasos

el cual se estima un *benchmark* y se dan una serie de recomendaciones o *check list* para la correcta evaluación de modelos de predicción de precios de energía eléctrica.

3.4 TÉCNICAS DE RECOLECCIÓN DE LA INFORMACIÓN

La presente investigación se fundamentará en la recopilación de información secundaria proveniente de bases de datos disponibles en Bloomberg [1], compuesta por un total de 37.152 datos de precios horarios de energía eléctrica por cada región que comprende el periodo desde el 1ro de enero de 2018 al 29 de marzo de 2022. Este conjunto de datos presenta los precios horarios de la energía en México para cada una de las regiones que componen el SEN y que fueron detalladas previamente.

¹Bloomberg es una herramienta que presenta datos en tiempo real así como registros históricos de diferentes mercados alrededor del mundo y que es de gran utilidad para los participantes del sector financiero; los datos que entrega se consideran altamente confiables al ser depurados, verificados y actualizados continuamente según evoluciona el mercado, los instrumentos y las entidades [17] [7]

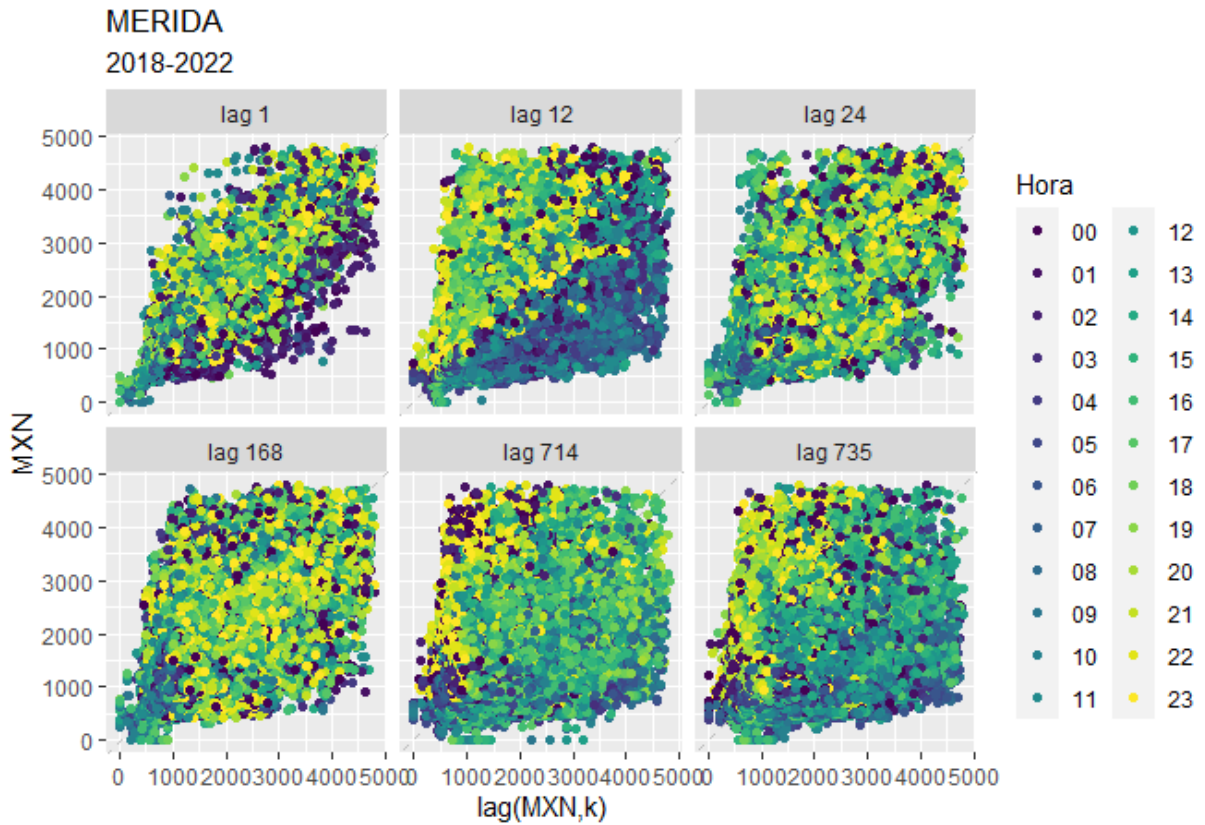


Figura 3.27: Gráfica de Autocorrelación por horas para el nodo Mérida y a diferentes retrasos

Para interpretar la gráfica de precios de energía por hora en diferentes escalas (hora, día, semana y mes), puedes observar los siguientes puntos:

- Interpretación por hora: En este nivel de detalle, se puede identificar las fluctuaciones de precios de energía a lo largo del día. Se puede observar si hay patrones recurrentes, como horas pico con precios más altos o horas de menor demanda con precios más bajos. También se puede identificar horas con precios estables o momentos de volatilidad.
- Interpretación por día: Al observar la gráfica a nivel diario, se puede identificar las tendencias generales de los precios de energía. Fijarse si hay variaciones significativas entre días de la semana, como precios más altos durante días laborables y precios más bajos durante los fines de semana. También se puede observar si hay patrones estacionales o cambios en los precios de energía a lo largo de diferentes estaciones del año.

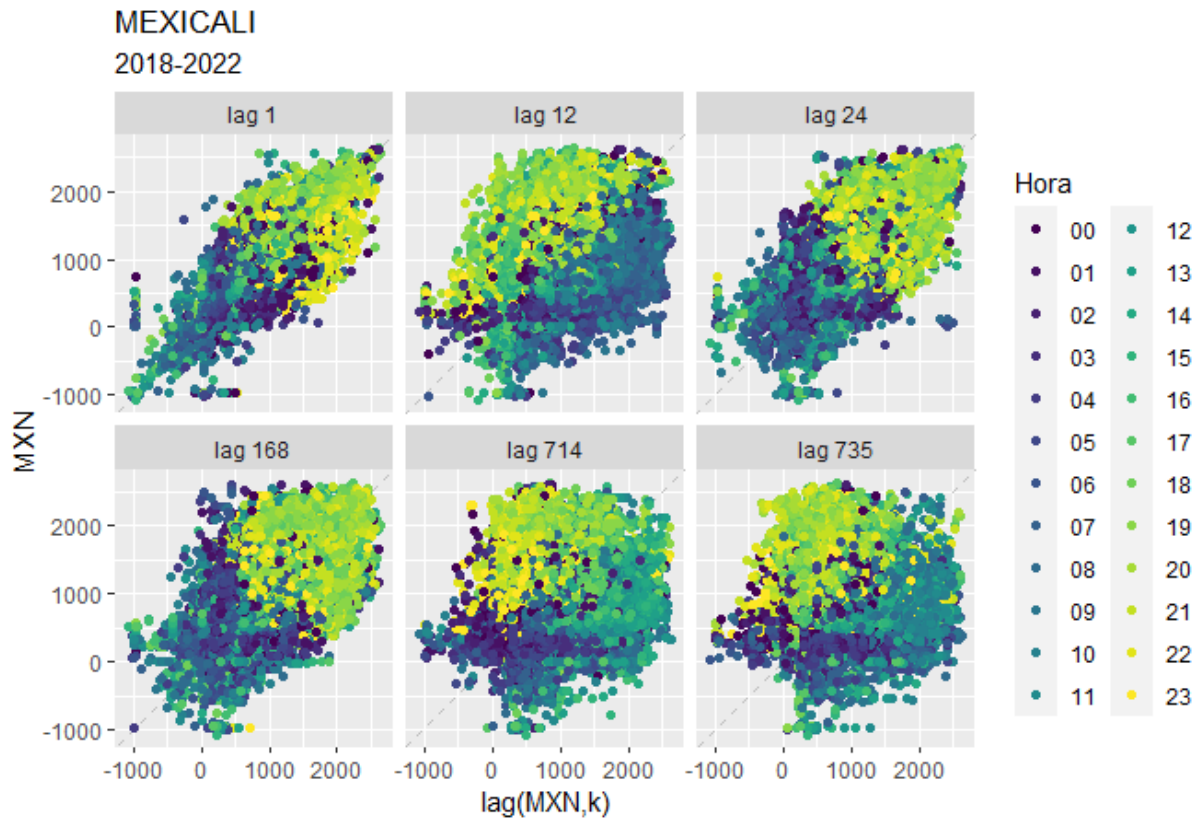


Figura 3.28: Gráfica de Autocorrelación por horas para el nodo Méxicali y a diferentes retrasos

- Interpretación por semana: Al ampliar la escala a nivel semanal, se puede identificar patrones de variación de precios a lo largo de una semana. Esto puede revelar fluctuaciones regulares, como precios más altos durante días laborables y precios más bajos durante los fines de semana. También se puede observar cambios en los precios de energía entre semanas y detectar posibles tendencias a largo plazo.
- Interpretación por mes: Al visualizar los precios de energía a nivel mensual, se puede identificar patrones estacionales y cambios a lo largo del año. Observar si hay meses con precios consistentemente más altos o más bajos en comparación con otros. También se puede detectar tendencias a largo plazo, como aumentos o disminuciones en los precios de energía a lo largo de los meses.

3.5 PRUEBA PILOTO

Se realiza prueba piloto usando datos de la región centro correspondiente a 3 meses, para el periodo comprendido entre 1 de febrero de 2021 al 30 de abril.

3.5.1 DESCOMPOSICIÓN

Para este caso se utilizó la técnica Error-Tendencia-Estacionalidad(ETS), que se puede resumir como:

- **Error (E):** El componente de error representa las fluctuaciones aleatorias y no sistemáticas en los datos de la serie de tiempo. Estos errores pueden ser debidos a factores impredecibles o ruidos en los datos. El componente de error también se conoce como el término residual", y se refiere a la diferencia entre el valor observado y el valor pronosticado.
- **Tendencia (T):** La tendencia se refiere a los cambios sistemáticos y a largo plazo en una serie de tiempo. Puede ser ascendente, descendente o estar ausente. Una tendencia positiva indica que la serie de tiempo está aumentando a lo largo del tiempo, mientras que una tendencia negativa indica una disminución. La tendencia puede ser lineal o no lineal.
- **Estacionalidad (S):** La estacionalidad se refiere a los patrones regulares y repetitivos que ocurren en intervalos fijos de tiempo, como estaciones, meses, días de la semana, entre otros. Por ejemplo, las ventas de abrigos pueden aumentar en invierno y disminuir en verano debido a la estacionalidad. La estacionalidad puede ser aditiva (las diferencias entre los valores observados y los valores estacionales son constantes) o multiplicativa (las diferencias varían proporcionalmente a los valores estacionales).

Cada componente (ver 3.29) se calcula de la siguiente manera:

1. **Nivel (Level):** El nivel en el período de tiempo t , denotado como L_t , se actualiza utilizando una combinación ponderada del valor observado Y_t en ese período y el nivel estimado previo L_{t-1} , según la fórmula:

$$L_t = \alpha \cdot Y_t + (1 - \alpha) \cdot L_{t-1}$$

El valor de α controla la ponderación relativa de los valores observados recientes y el nivel estimado previo en la actualización del nivel. Si α es cercano a 1, se da más importancia a Y_t , lo que hace que el modelo sea más reactivo a las fluctuaciones recientes en los datos. Si α es cercano a 0, el modelo se basa más en el nivel previo y es menos sensible a las fluctuaciones recientes.

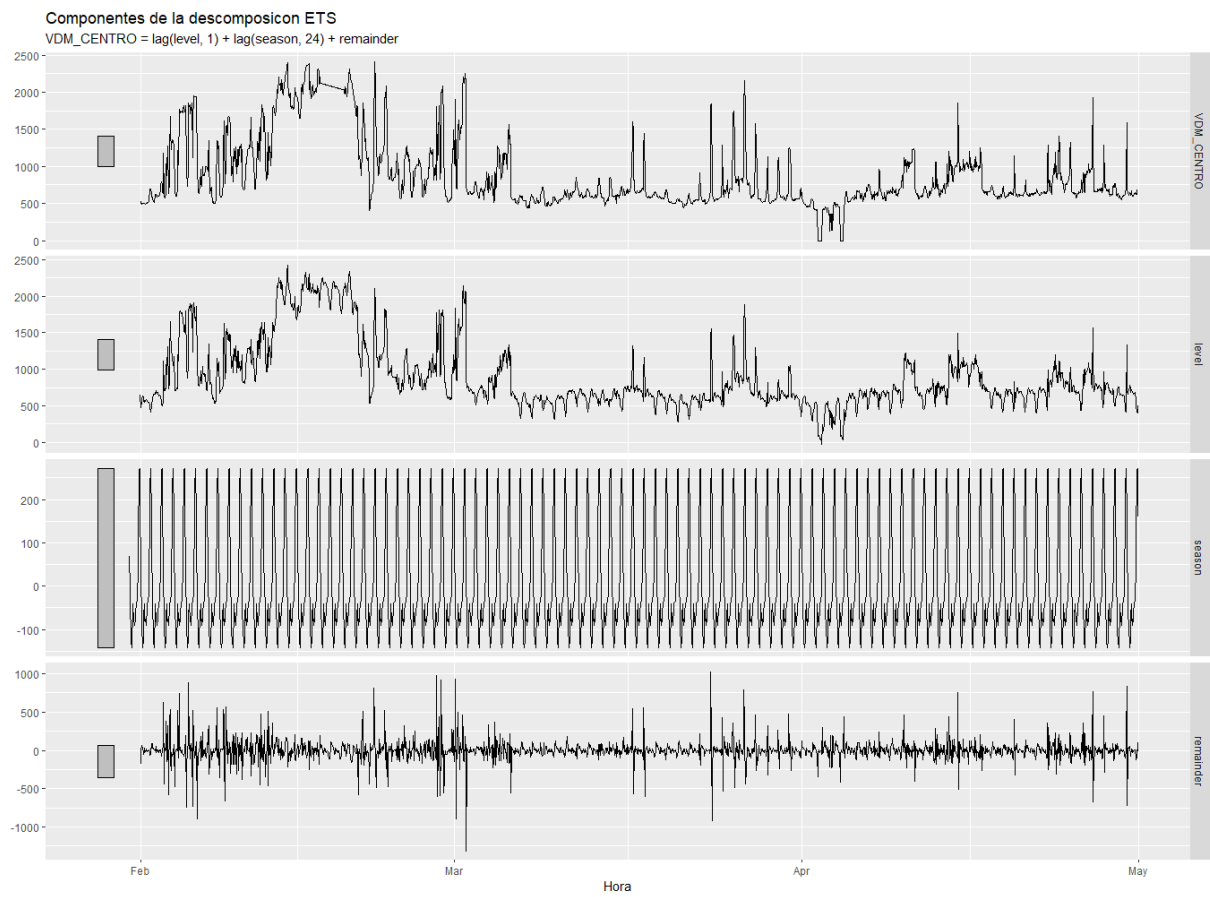


Figura 3.29: Descomposición por medio de ETS

2. **Tendencia (Trend):** En un modelo de suavización exponencial simple.

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

3. **Estacionalidad (Seasonality):** serie tiene estacionalidad, es necesario considerar modelos como Holt-Winters, que incorporan componentes de estacionalidad.

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-m}$$

4. **Error (Error):** La componente de error en el período t , denotada como e_t , se calcula como la diferencia entre el valor observado Y_t y la suma de las componentes de nivel L_t , tendencia T_t y estacionalidad S_t :

$$E_t = Y_t - L_t - T_t - S_t$$

Donde:

- α, β y γ son los parámetros de suavización correspondientes a nivel, tendencia y estacionalidad, respectivamente.
- Y_t es el valor observado en el período t .
- L_t es el nivel en el período t .
- L_{t-1} es el nivel estimado en el período anterior.
- S_t es la componente de estacionalidad en el período t (si aplica).
- m es el número de períodos en una estacionalidad. Por ejemplo, 24 para datos horarios con estacionalidad diaria).

En resumen, los componentes en un modelo de suavización exponencial simple se calculan principalmente a través de la actualización del nivel utilizando el valor observado y el nivel previo, controlada por el parámetro de suavización α .

Parámetros de Suavización para estos datos

- Parámetro de suavización alfa (α): 0.6882047
- Parámetro de suavización gamma (γ): 0.0001000661

- Valores Iniciales de los Estados: Los valores iniciales de los estados representan los valores iniciales estimados para las componentes de nivel y estacionalidad en diferentes retrasos ($-1, -2, \dots, -23$). Por ejemplo, el valor inicial para el nivel es 641.8445 y para la estacionalidad en el retraso -1 es 112.8504, y así sucesivamente.
- Varianza del Error (σ^2): La varianza del error (σ^2) es 22769.12. Representa la cantidad de variabilidad no explicada por las componentes de nivel, estacionalidad y error del modelo.
- Criterios de Información:
 1. AIC (Criterio de Información de Akaike): 194589.1
 2. AICc (Criterio de Información de Akaike corregido): 194589.2
 3. BIC (Criterio de Información Bayesiano): 194784.1

Estos criterios de información son útiles para evaluar la calidad del ajuste del modelo. Valores más bajos indican un mejor ajuste.

En resumen, el modelo ETS(A,N,A) aplicado a la serie "VDM_CENTRO" utiliza una combinación de suavización exponencial para modelar las componentes de nivel y estacionalidad, con una varianza del error de 22769.12. Los valores iniciales de los estados y los criterios de información proporcionan información sobre la calidad del ajuste del modelo a los datos observados.

3.5.2 PREDICCIÓN

El proceso de descomposición es especialmente útil para analizar datos de series temporales y explorar los cambios históricos a lo largo del tiempo, pero también puede emplearse con fines de pronóstico.

Suponiendo una descomposición aditiva, la serie temporal descompuesta se puede escribir como

$$y_t = \hat{S}_t + \hat{A}_t,$$

donde

$$\hat{A}_t = \hat{T}_t + \hat{E}_t$$

es la componente ajustada estacionalmente. O, si se ha utilizado una descomposición multiplicativa, podemos escribir

$$y_t = \hat{S}_t \cdot \hat{A}_t,$$

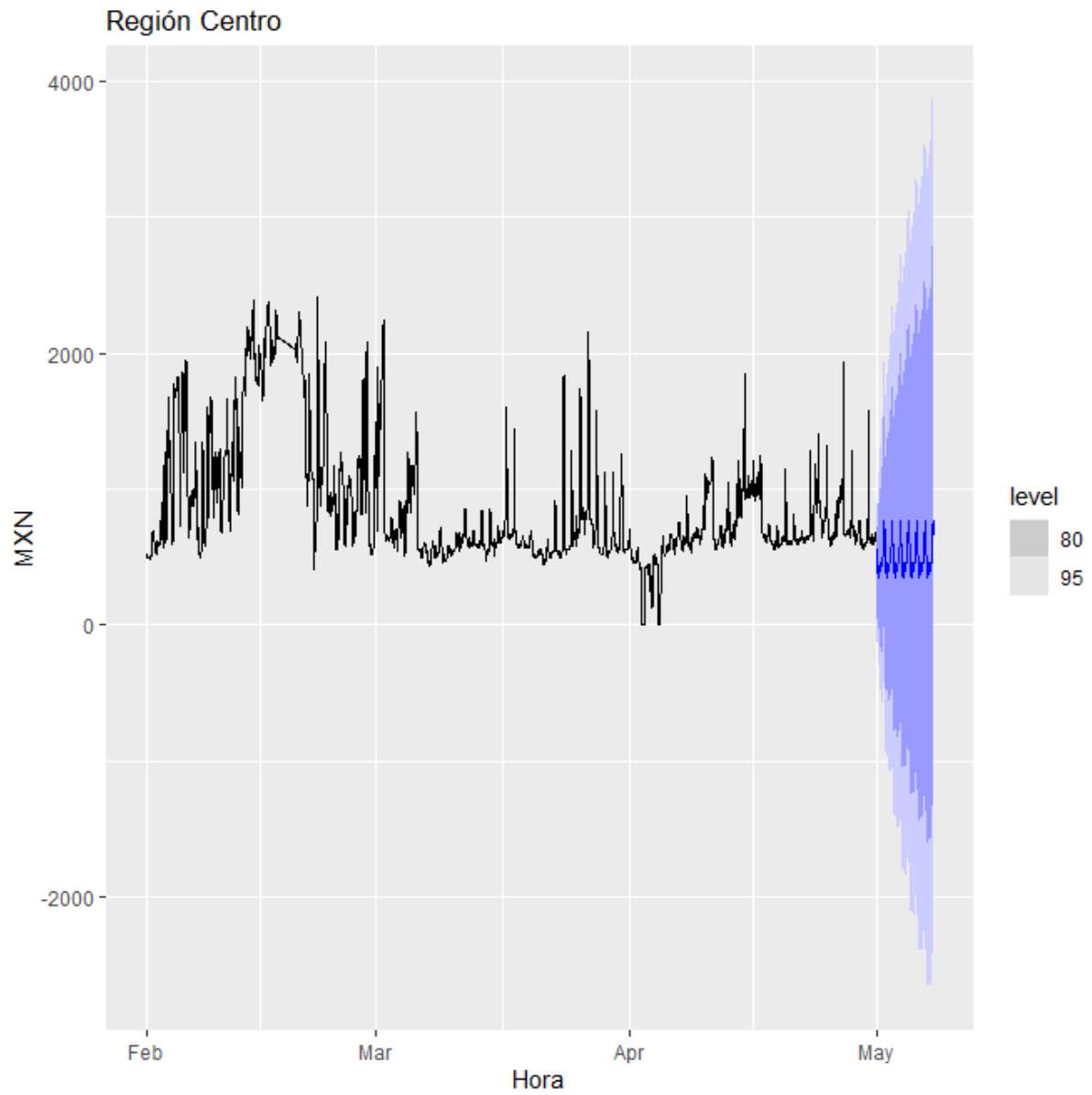


Figura 3.30: Predicción a una semana

donde

$$\hat{A}_t = \hat{T}_t \cdot \hat{E}_t.$$

Para pronosticar una serie temporal descompuesta, pronosticamos la componente estacional, \hat{S}_t , y la componente ajustada estacionalmente, \hat{A}_t , por separado. Por lo general, se asume que la componente estacional no cambia, o cambia extremadamente lentamente, por lo que se pronostica simplemente tomando el último año de la componente estimada.

La predicción generada por este modelo es efectiva para capturar y reflejar de manera precisa los patrones estacionales presentes en los datos. Sin embargo, es importante señalar que su capacidad para anticipar eventos de tendencia resulta menos satisfactoria. Esto puede atribuirse a la naturaleza intrínseca del modelo, que tiende a enfocarse en las variaciones estacionales en lugar de las tendencias a largo plazo.

3.6 TÉCNICAS DE ANÁLISIS DE LA INFORMACIÓN

3.6.1 ESPECTOGRAMA

El espectrograma obtenido mediante la Transformada Rápida de Fourier (FFT) es una herramienta valiosa en el análisis de precios horarios de energía. Esta representación gráfica descompone la serie temporal de precios en sus componentes de frecuencia, permitiendo identificar patrones y comportamientos cíclicos en la variabilidad de los precios a lo largo del tiempo. Mediante la conversión de los datos de precio desde el dominio del tiempo al dominio de la frecuencia, el espectrograma brinda una idea de las fluctuaciones y tendencias presentes en los precios horarios de energía. La visualización resultante facilita la detección de patrones de temporada, ciclos recurrentes y posibles relaciones entre eventos específicos y cambios en la estructura de precios. En definitiva, el uso del espectrograma basado en la FFT se presenta como una herramienta poderosa para desentrañar la complejidad subyacente en los datos de precios horarios de energía y proporcionar información clave para la toma de decisiones en el ámbito energético.

A continuación se representan los espectrogramas para cada una de las regiones seleccionadas. Ver Figuras 3.31, 3.32, 3.33, 3.34, 3.35

Dada la resolución de los datos por hora y la medición de la frecuencia en ciclos por segundo, la frecuencia más baja discernible en el espectrograma corresponde a un ciclo a lo largo de los cuatro años. Esto se traduce en el período más largo detectable en el espectrograma, que equivale a cuatro años.

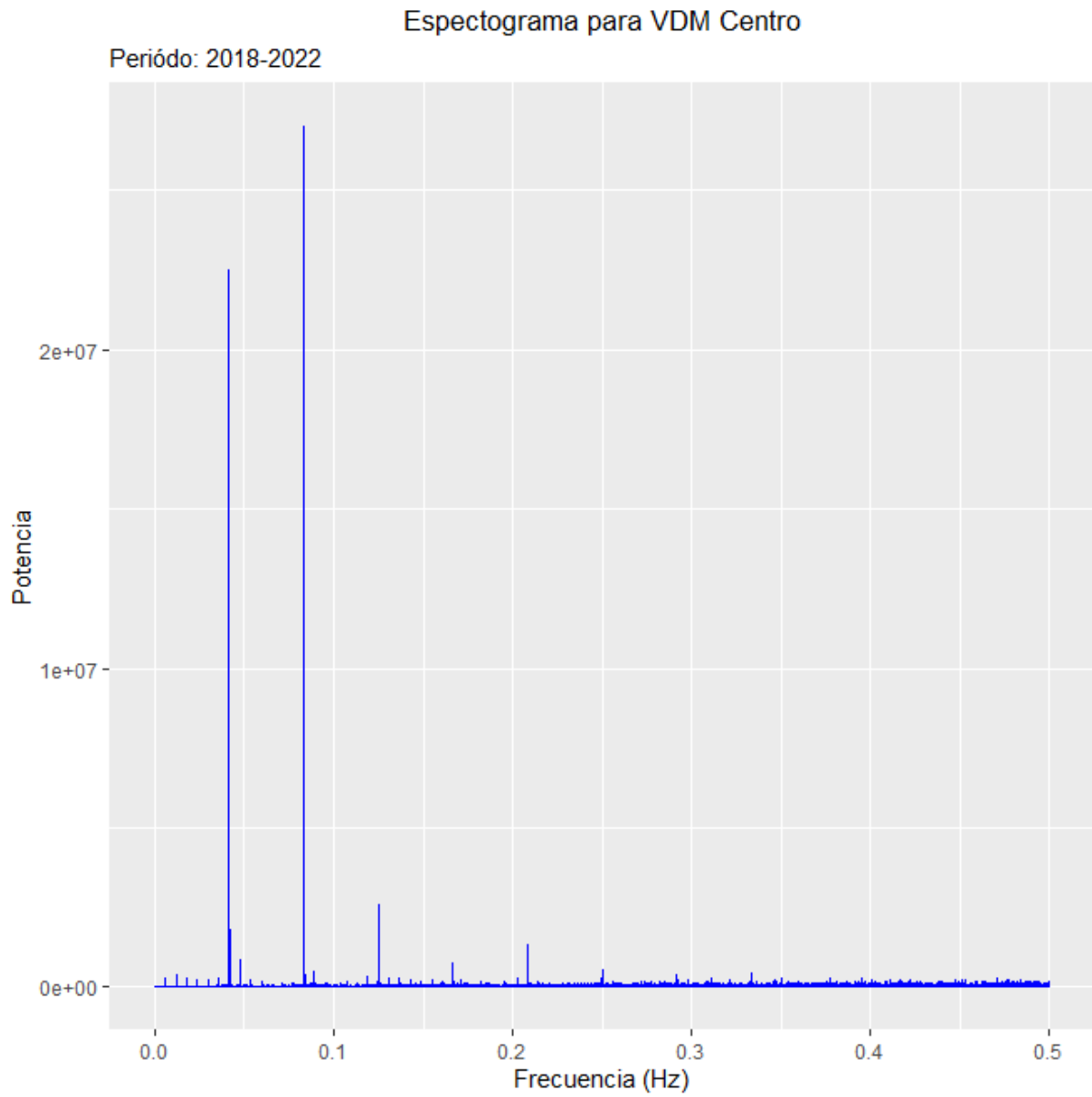


Figura 3.31: Espectograma de frecuencias para Valle de México Centro

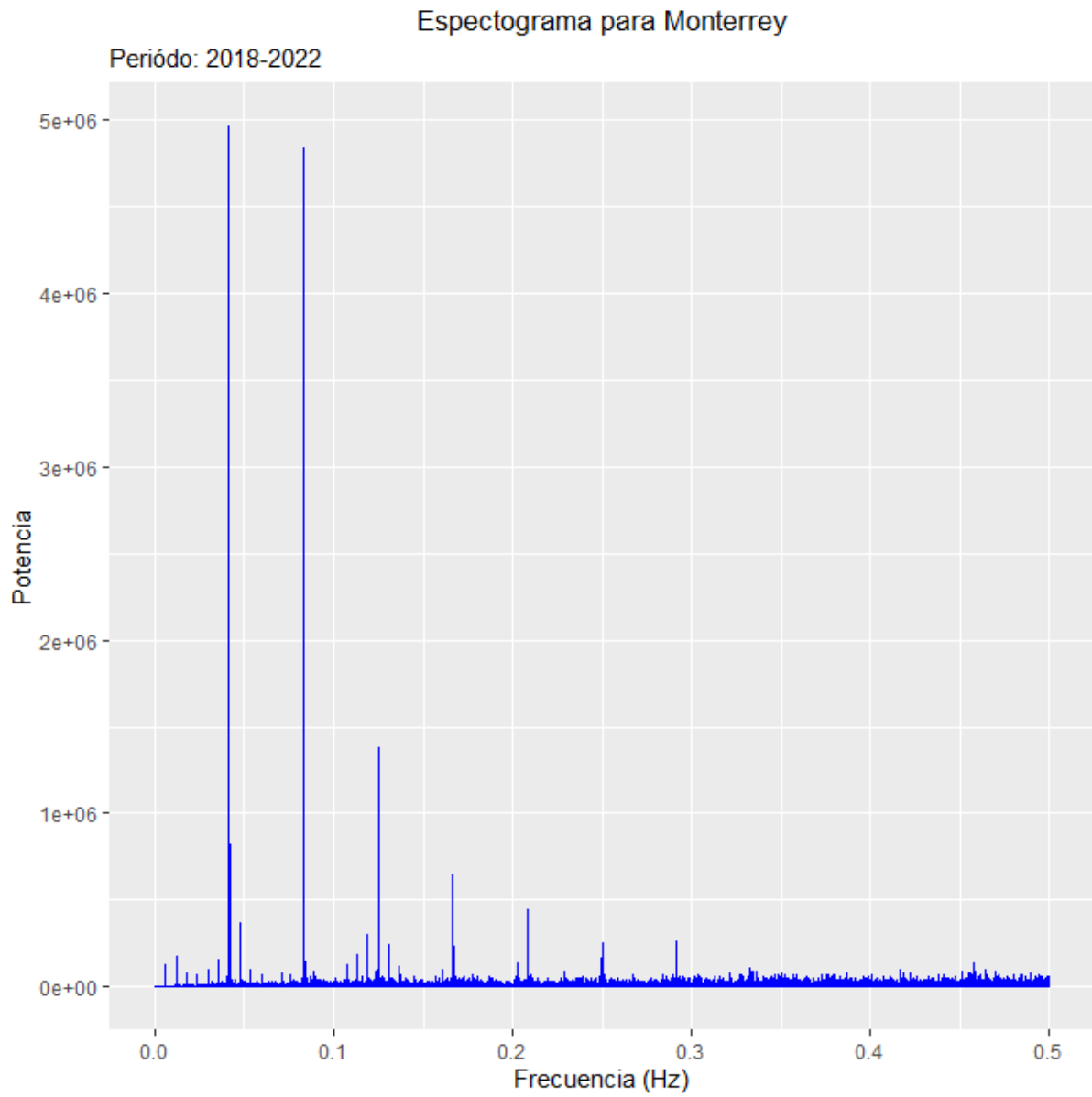


Figura 3.32: Espectograma de frecuencias para Monterrey

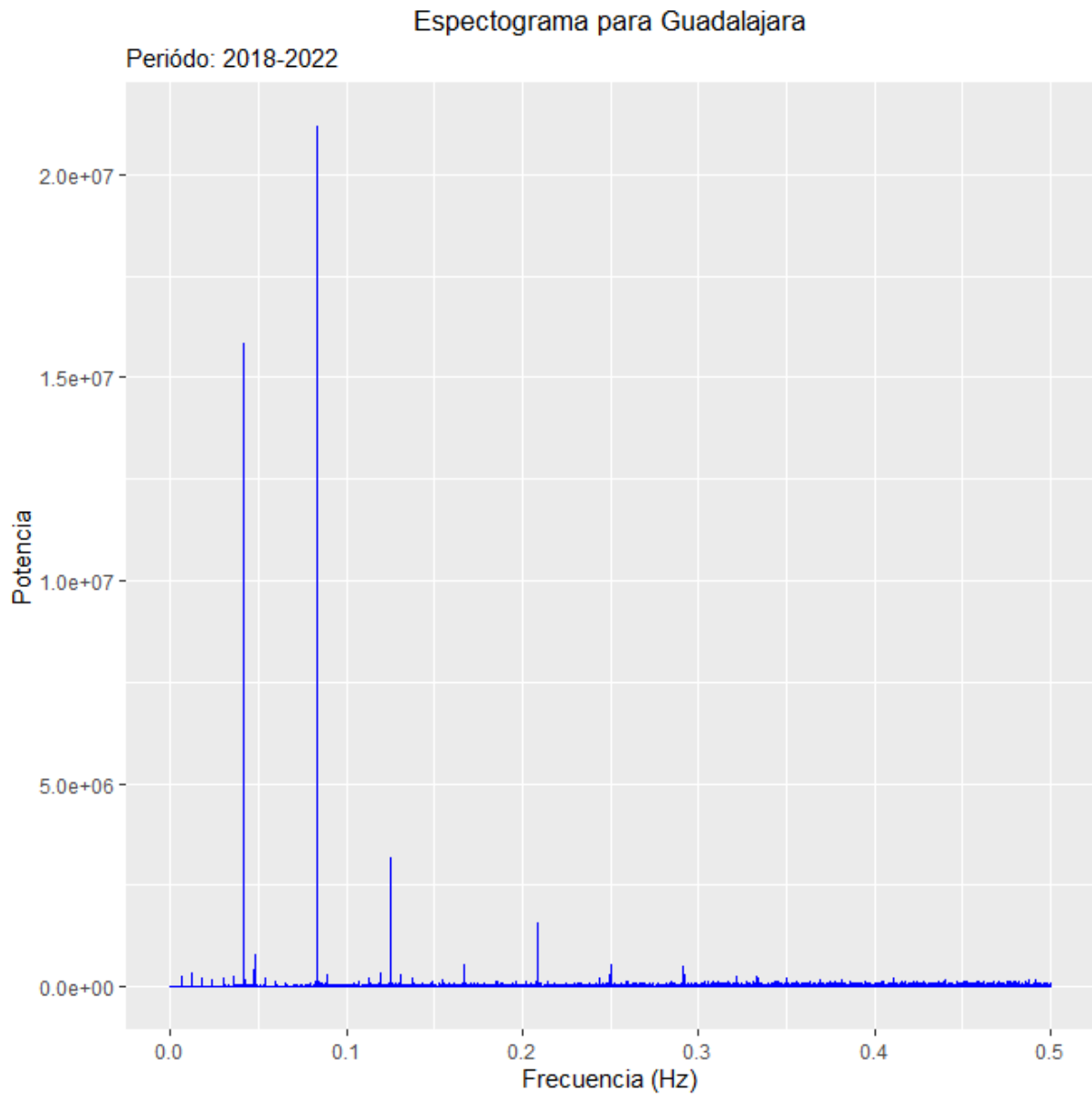


Figura 3.33: Espectograma de frecuencias para Guadalajara

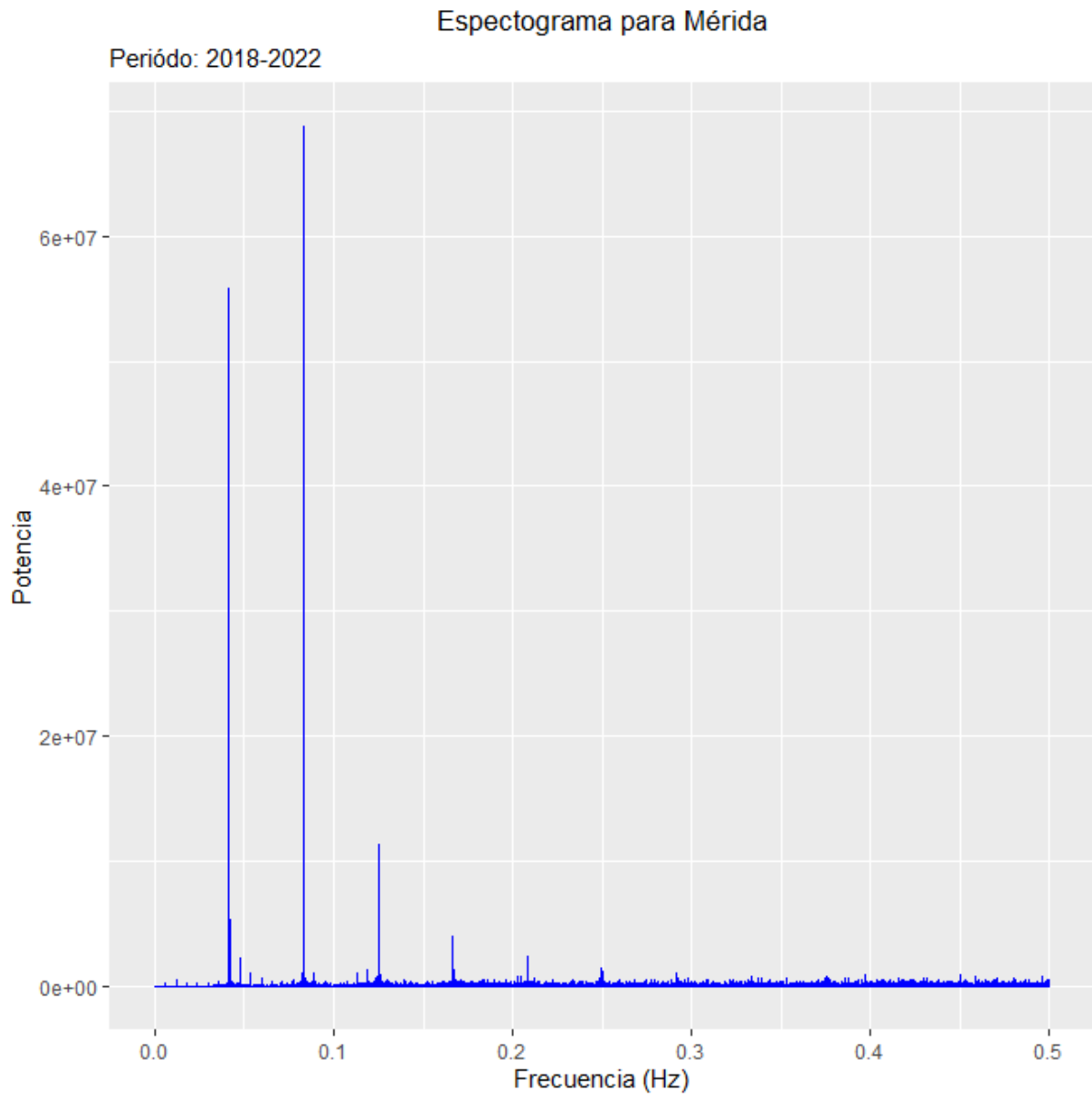


Figura 3.34: Espectograma de frecuencias para Mérida

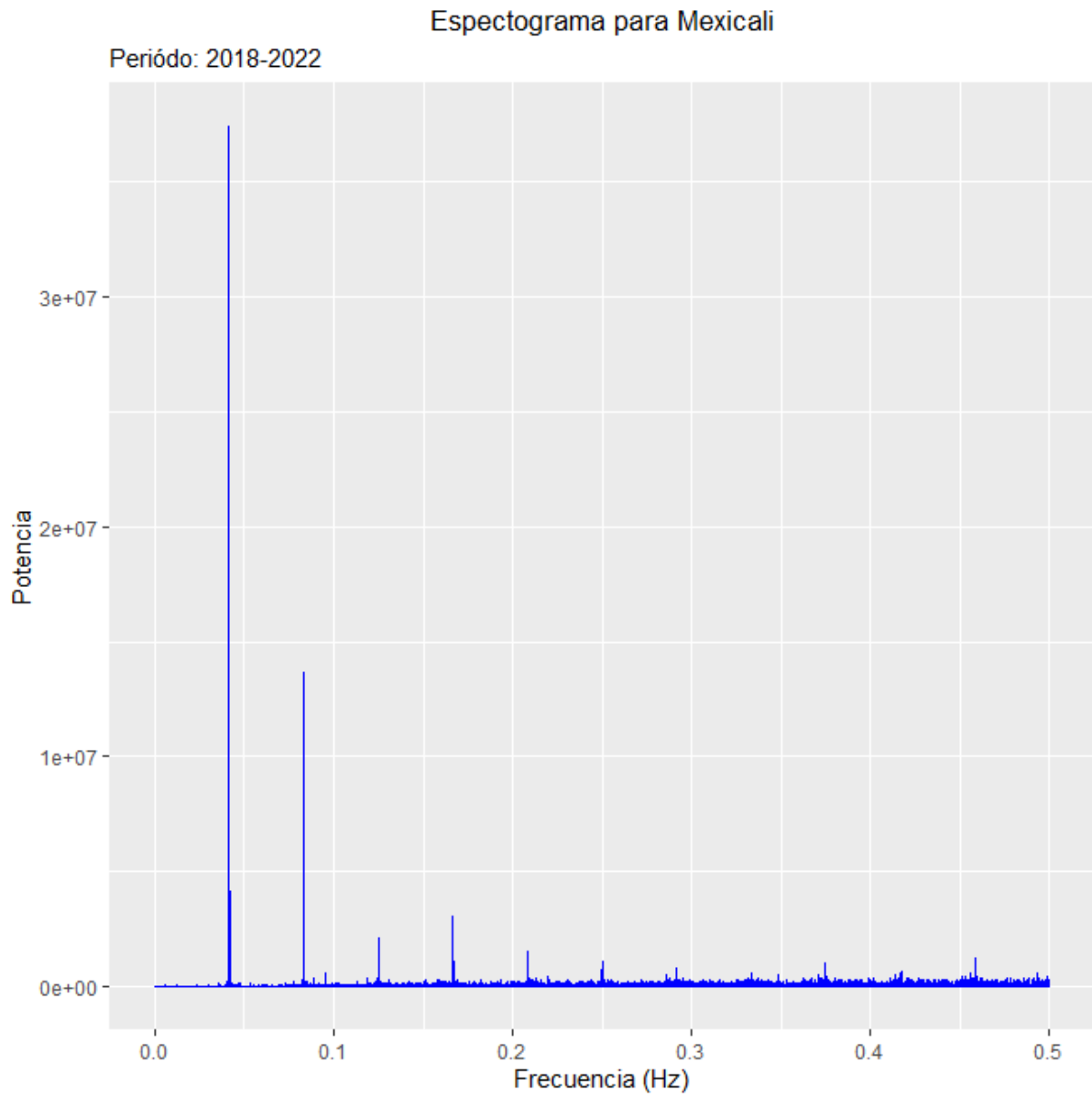


Figura 3.35: Espectograma de frecuencias para Mexicali

Como aproximación general para establecer un contexto temporal en horas basado en las frecuencias del espectrograma para el conjunto de datos de cuatro años:

- La frecuencia más baja (1 ciclo por 4 años) correspondería a un período de 4 años, lo que se traduce en $4 \text{ años} * 365 \text{ días/año} * 24 \text{ horas/día}$.
- La frecuencia más alta en el espectrograma (0.5 ciclos por hora) correspondería a un período de 2 horas.

Para las frecuencias entre estos dos extremos, se podría emplear un enfoque de interpolación lineal para estimar el período correspondiente en horas.

Es importante destacar que esta aproximación simplificada asume una relación lineal entre las frecuencias y las horas, lo que podría no capturar con precisión posibles patrones no lineales debido a las características intrincadas de tus datos y a los posibles comportamientos estacionales subyacentes. Si tienes información adicional sobre los patrones cíclicos dentro de tus datos, ajustar esta relación de manera más precisa podría lograr una alineación más exacta entre las frecuencias del espectrograma y sus equivalentes temporales. Este esfuerzo analítico ejemplifica la intrigante interacción entre el análisis de frecuencias y la interpretación temporal en el ámbito de la investigación de series de tiempo.

3.6.2 ANÁLISIS POLAR

Las gráficas polares se revelan como una herramienta visual esencial en el análisis de los precios de la energía, ya que ofrecen una perspectiva única al presentar la distribución de precios en función de múltiples escalas temporales. Al agrupar los precios por hora, día, semana y mes, estas gráficas trazan patrones circulares que capturan las variaciones en el comportamiento de los precios a lo largo del tiempo. Cada anillo concéntrico representa una de estas escalas, y los puntos en la circunferencia del anillo indican los precios en un período específico. Este enfoque permite identificar tendencias recurrentes, fluctuaciones estacionales y posibles correlaciones temporales con mayor claridad. Las gráficas polares, al combinar la información de múltiples intervalos temporales en una visualización cohesiva, ofrecen una visión holística y rica de los patrones de precios de la energía, brindando una perspectiva invaluable para la toma de decisiones en el sector energético.

- Las gráficas agrupadas por horas (D) son las figuras: 3.36, 3.40, 3.44, 3.48, 3.52.
- Las gráficas agrupadas por días (W) son las figuras: 3.37, 3.41, 3.45, 3.49, 3.53.

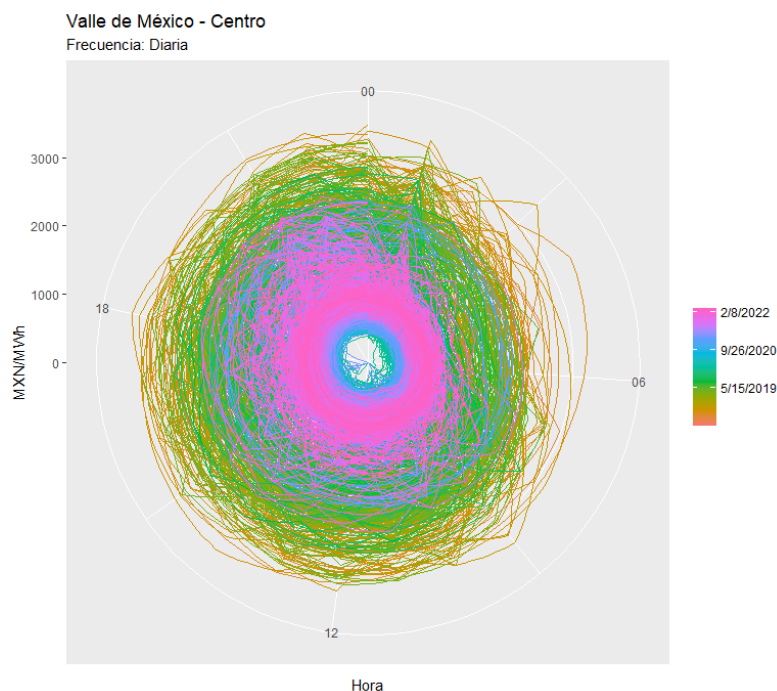


Figura 3.36: Distribución de precio por hora del día región Valle de México Centro

- Las gráficas agrupadas por semanas (M) son las figuras: 3.38, 3.42, 3.46, 3.50, 3.54.
- Las gráficas agrupadas por meses (Y) son las figuras: 3.39, 3.43, 3.47, 3.51, 3.55.

Los gráficos estacionales son una forma de visualización que se utiliza para identificar patrones repetitivos y estacionales en los datos a lo largo del tiempo. No hay una fórmula matemática única para representar gráficos estacionales, ya que estos gráficos se basan en datos observados en diferentes momentos del tiempo.

La idea principal detrás de un gráfico estacional es resaltar patrones recurrentes en los datos a lo largo de intervalos regulares de tiempo, como días, semanas, meses o años. Por lo tanto, en lugar de una fórmula matemática específica, se utiliza una representación gráfica que permite visualizar estos patrones.

3.7 PREDICCIÓN

Falta ahora calcular de los escenarios futuros, proceso conocido como predicción puntual. Consta de 3 pasos.

- La ecuación ARIMA debe tener a x_t como término izquierdo, y el resto de términos en el lado derecho.

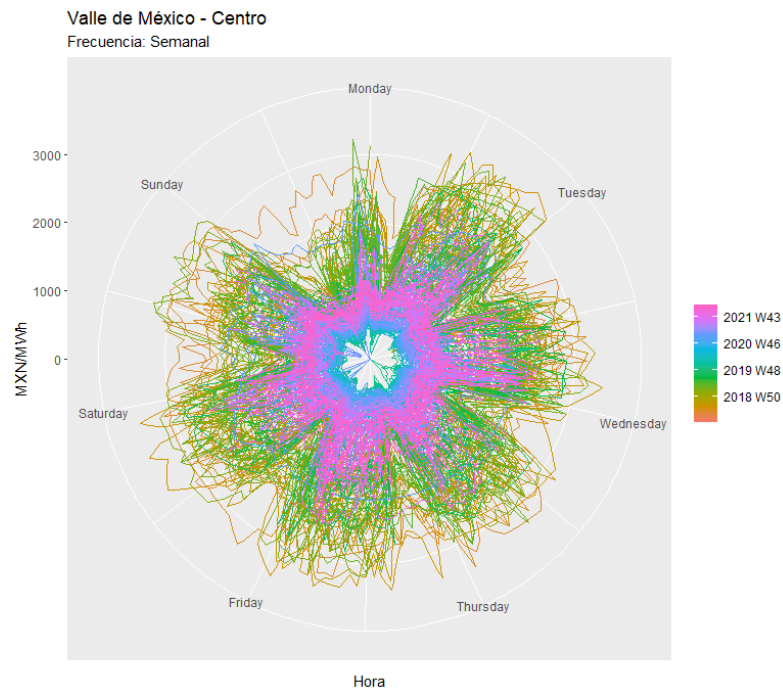


Figura 3.37: Distribución de precio por día de la semana región Valle de México Centro

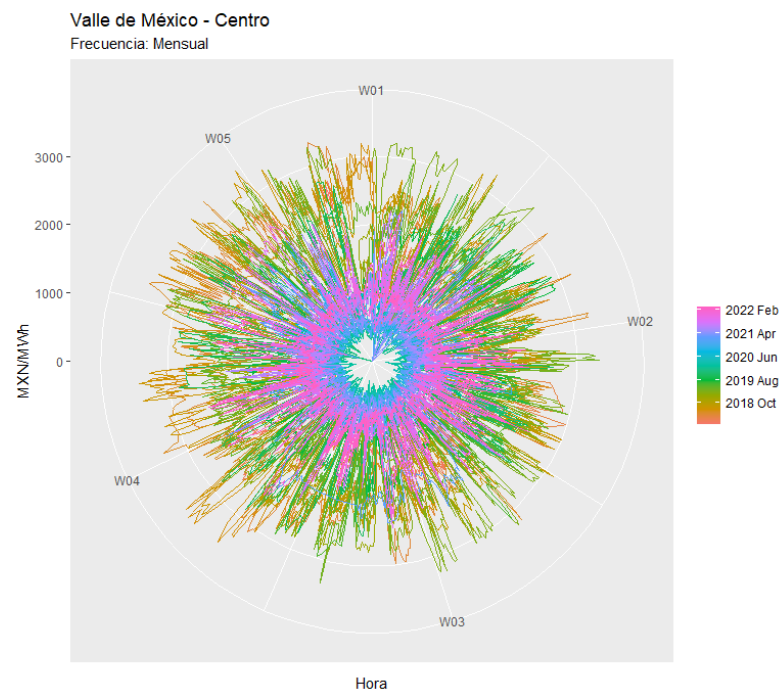


Figura 3.38: Distribución de precio por semana del mes región Valle de México Centro

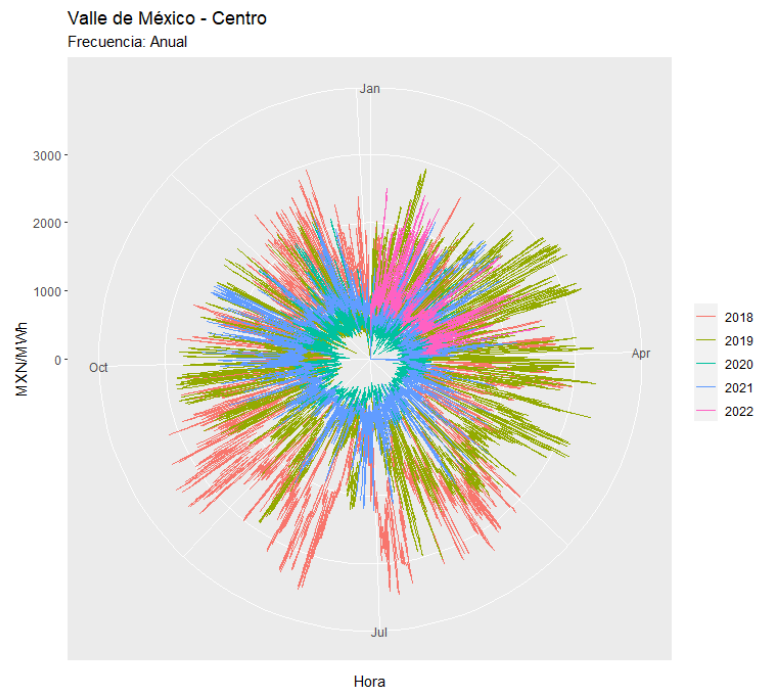


Figura 3.39: Distribución de precio por mes del año región Valle de México Centro

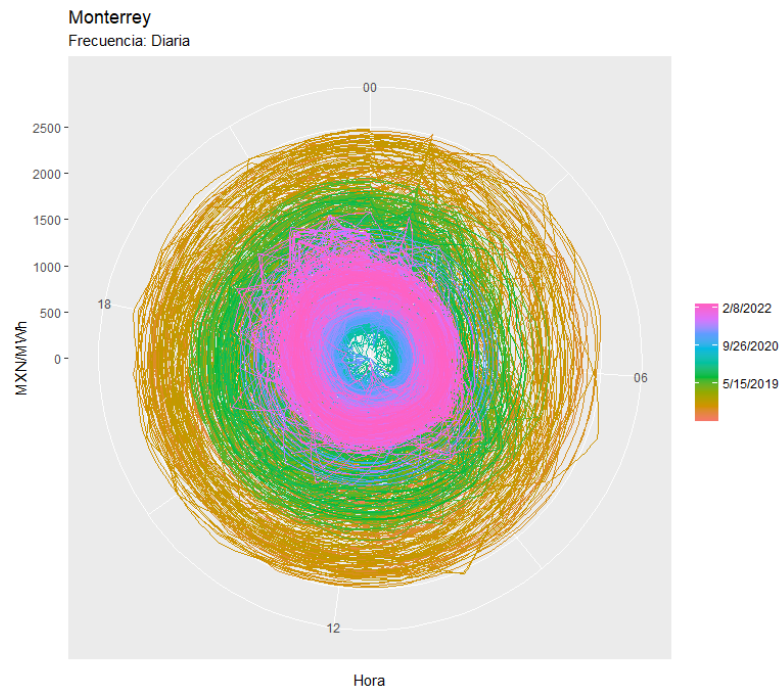


Figura 3.40: Distribución de precio por hora del día región Monterrey

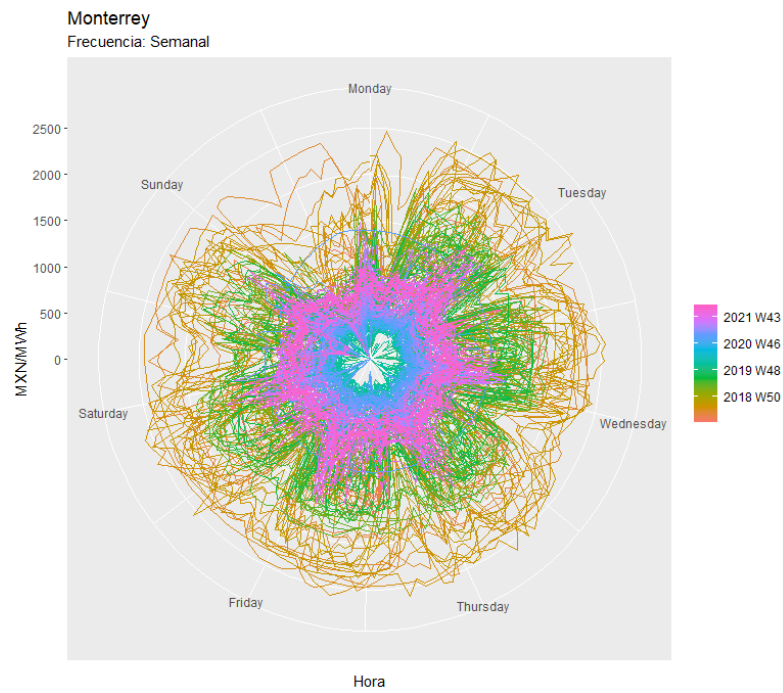


Figura 3.41: Distribución de precio por día de la semana región Monterrey

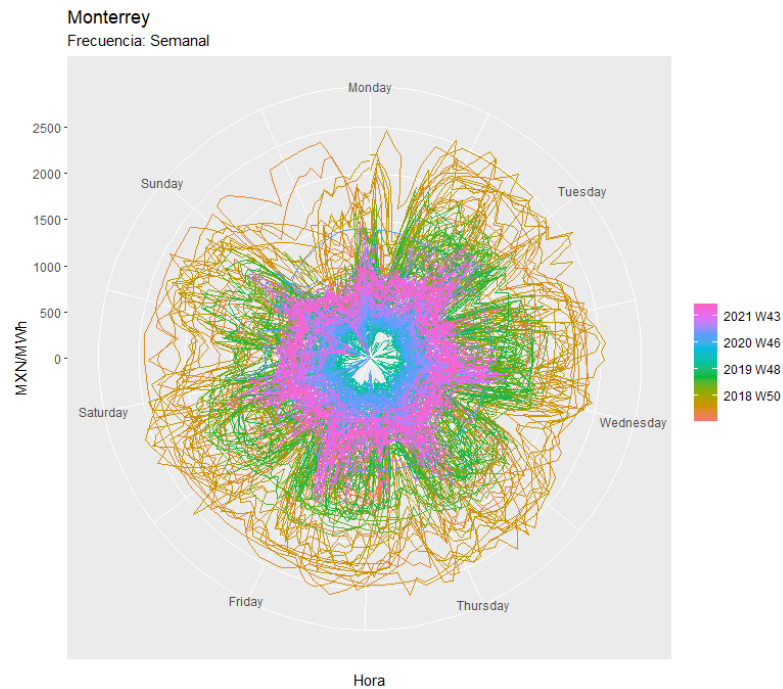


Figura 3.42: Distribución de precio por semana del mes región Monterrey

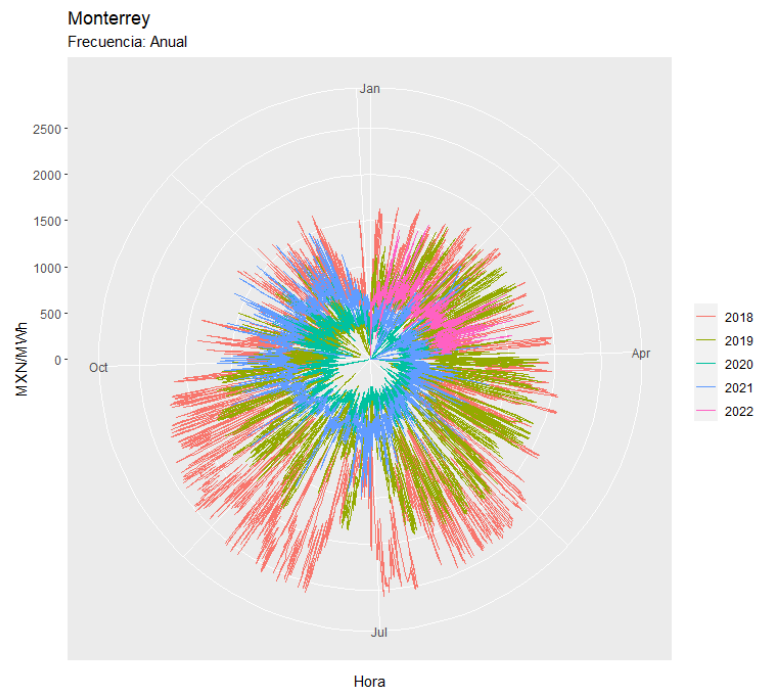


Figura 3.43: Distribución de precio por mes del año región Monterrey

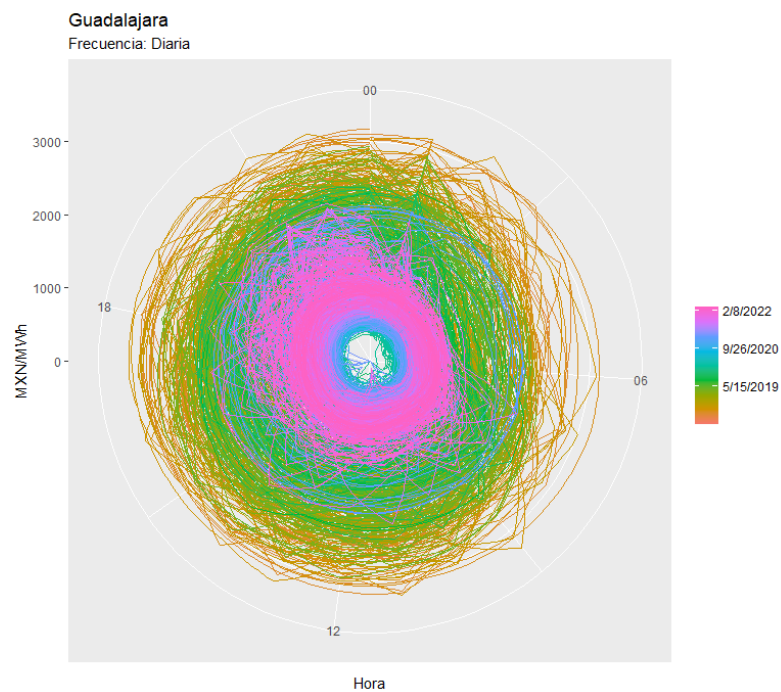


Figura 3.44: Distribución de precio por hora del día región Guadalajara

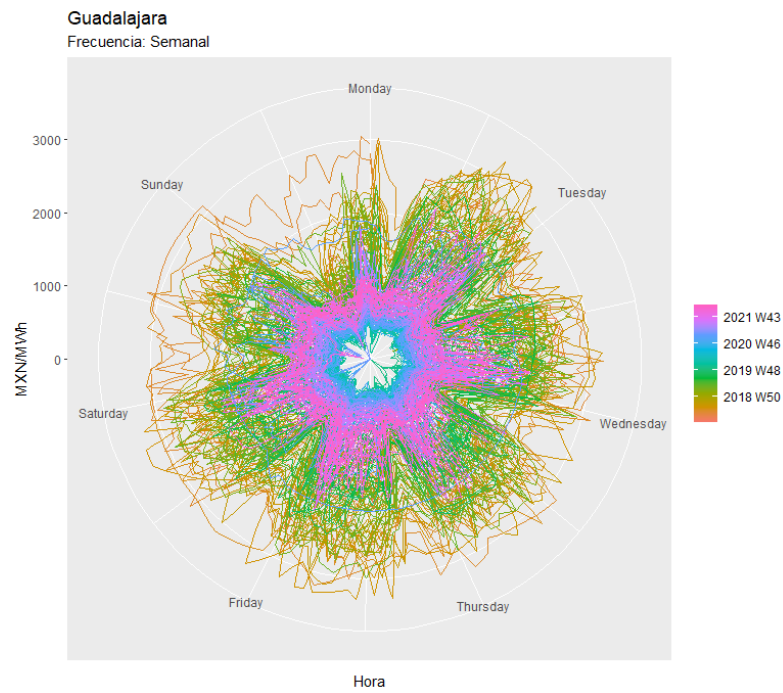


Figura 3.45: Distribución de precio por día de la semana región Guadalajara

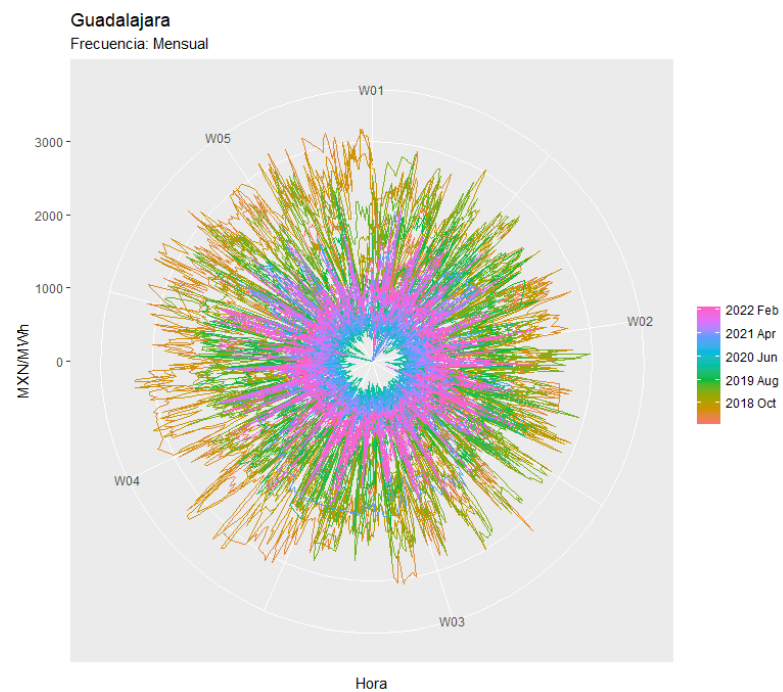


Figura 3.46: Distribución de precio por semana del mes región Guadalajara

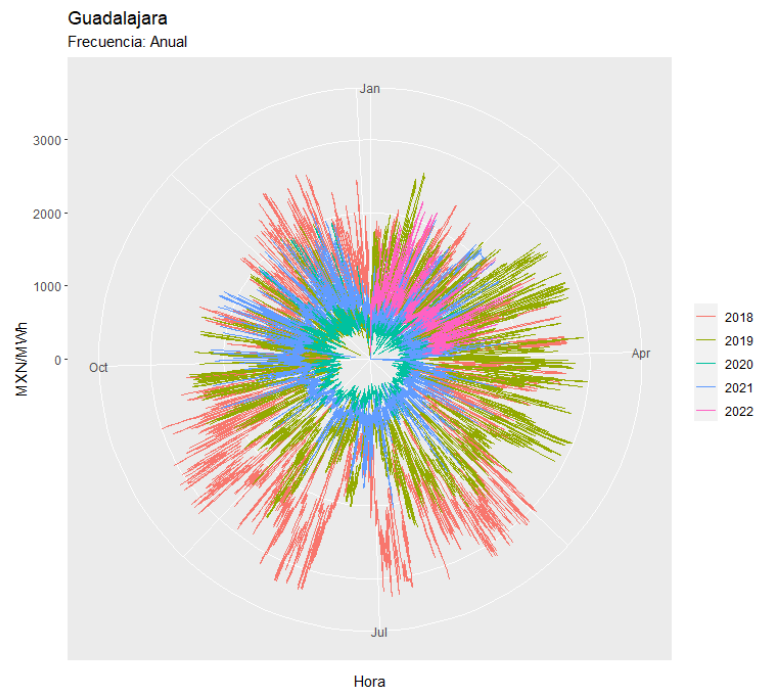


Figura 3.47: Distribución de precio por mes del año región Guadalajara

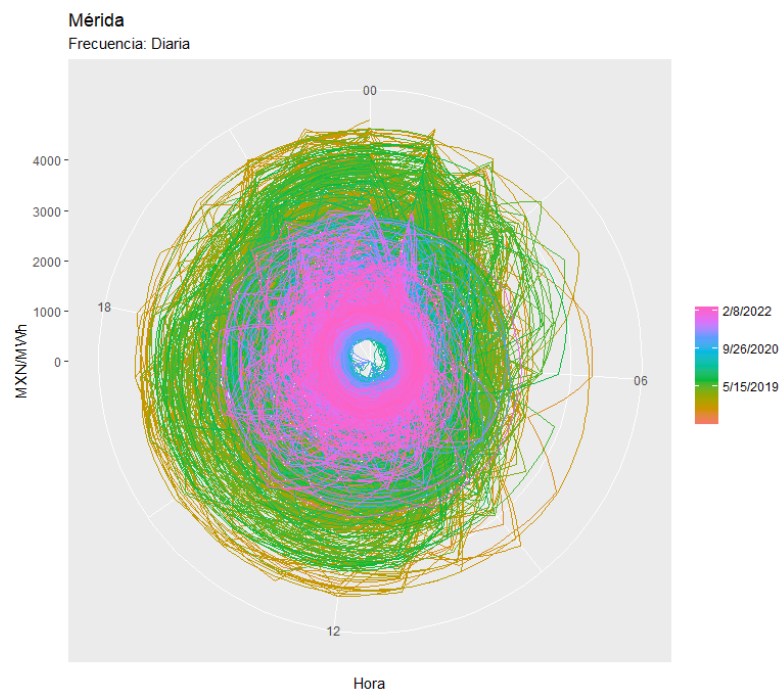


Figura 3.48: Distribución de precio por hora del día región Mérida

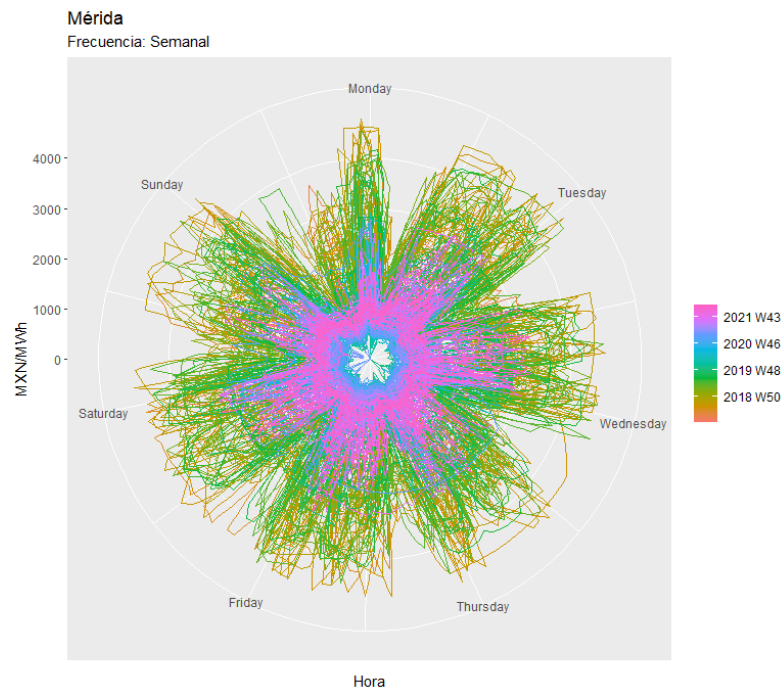


Figura 3.49: Distribución de precio por día de la semana región Mérida

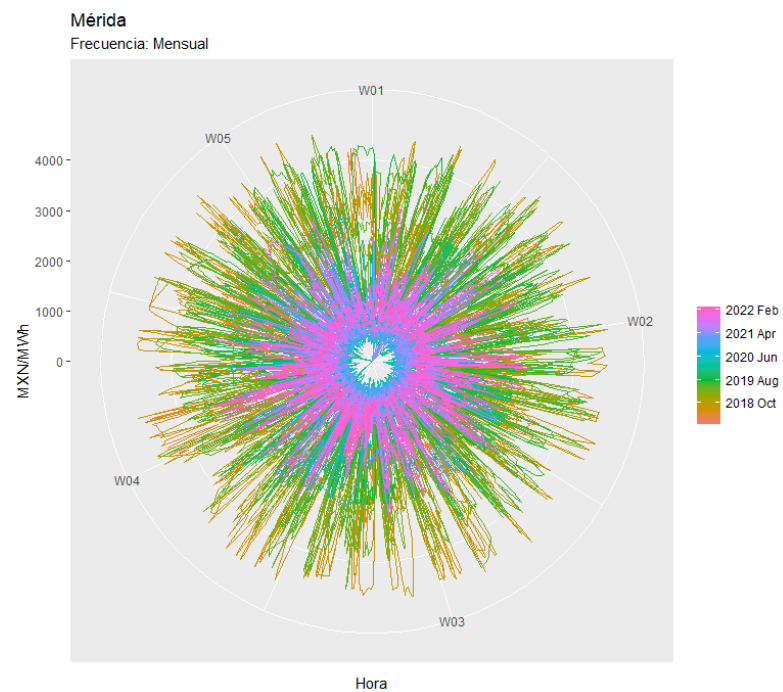


Figura 3.50: Distribución de precio por semana del mes región Mérida

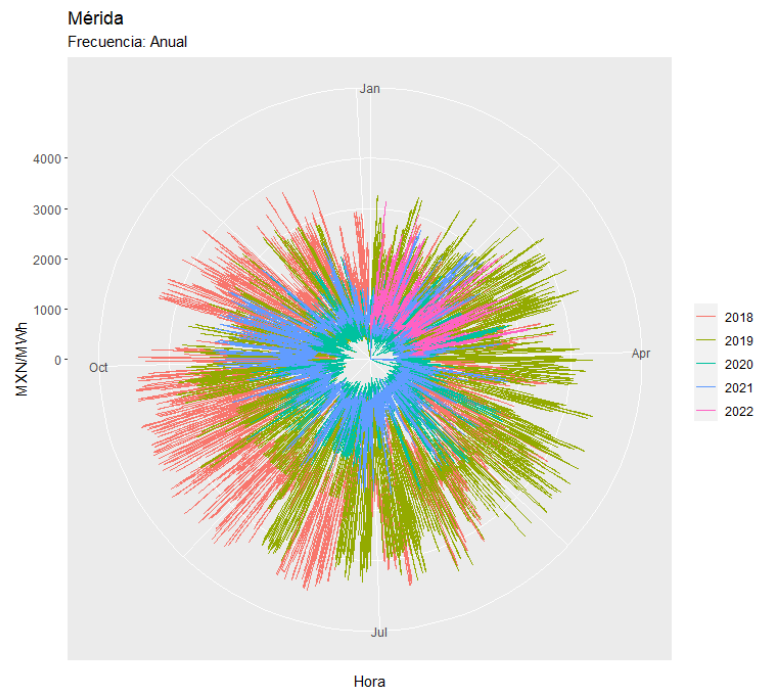


Figura 3.51: Distribución de precio por mes del año región Mérida

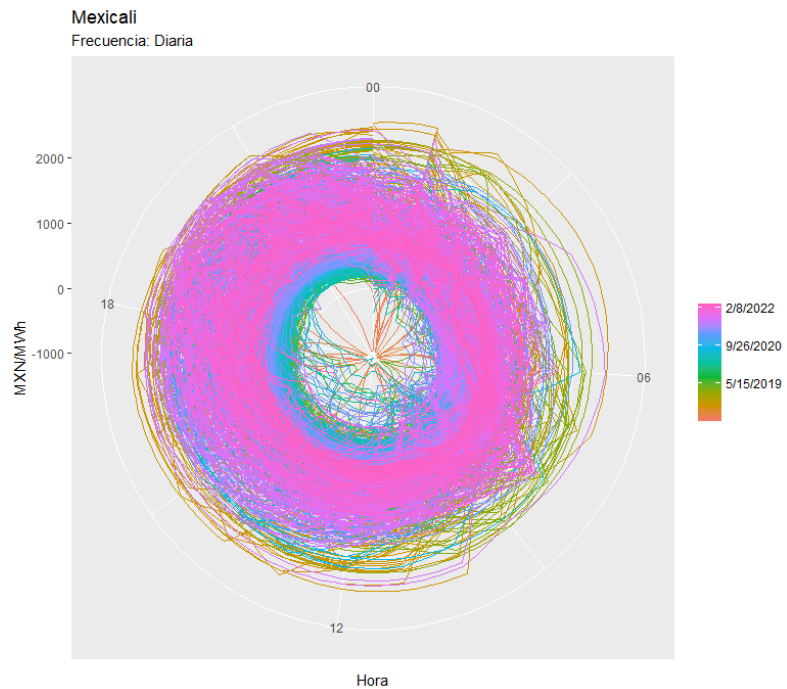


Figura 3.52: Distribución de precio por hora del día región Mexicali

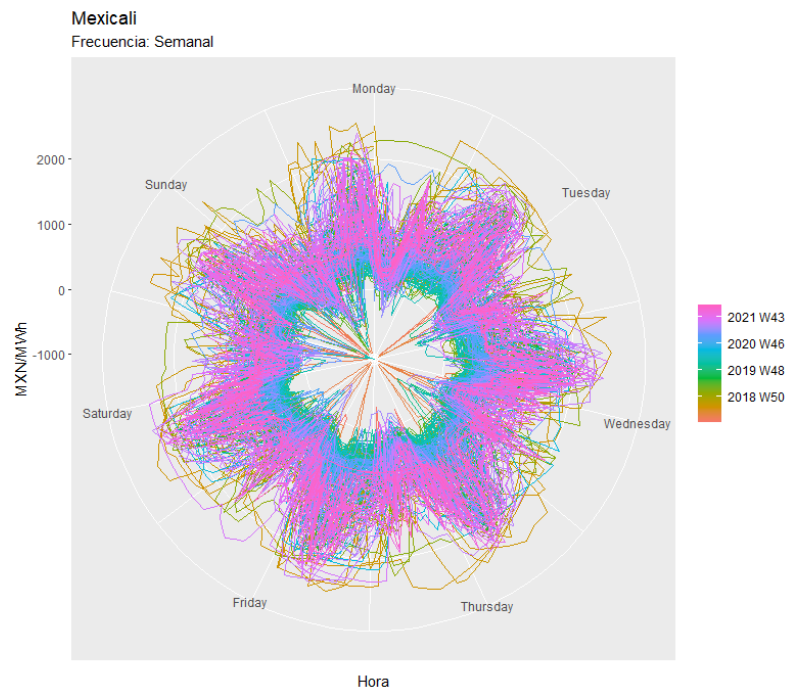


Figura 3.53: Distribución de precio por día de la semana región Mexicali

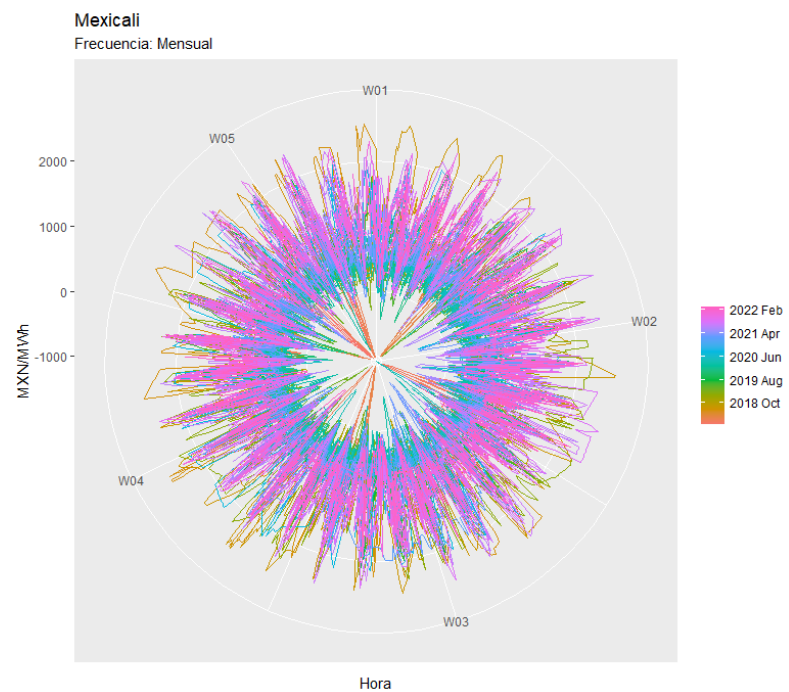


Figura 3.54: Distribución de precio por semana del mes región Mexicali

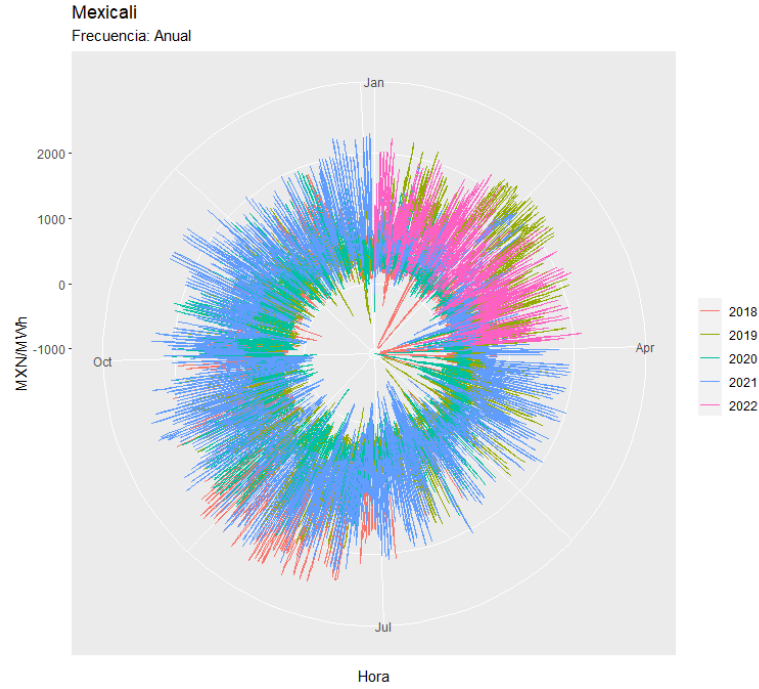


Figura 3.55: Distribución de precio por mes del año región Mexicali

- Reescribir la ecuación reemplando t con $T + h$.
- En el lado derecho de la ecuación, reemplazar las observaciones con la predicción, el error con cero, y los errores pasados con los residuos correspondientes.

Todo esto se debe hacer para $h = 1$, estos pasos deben ser luego repetidos para $h = 2, 3, \dots$ hasta que todas las predicciones sean calculadas.

Utilizando el ejemplo de [27], con un modelo ARIMA(3,1,1), que se puede escribir como sigue,

$$(1 - \hat{\phi}_1 B - \hat{\phi}_2 B^2 - \hat{\phi}_3 B^3)(1 - B)x_t = (1 + \hat{\theta}_1 B)\epsilon_t$$

Primero, se expande el término de la igualdad,

$$\left[1 - (1 + \hat{\phi}_1)B + (\hat{\phi}_1 - \hat{\phi}_2)B^2 + (\hat{\phi}_2 - \hat{\phi}_3)B^3 + \hat{\phi}_3 B^4 \right] x_t = (1 + \hat{\theta}_1 B)\epsilon_t$$

Ahora, aplicando el operador de retroceso (\mathbf{B}) da como resultado,

$$x_t - (1 + \hat{\phi}_1)x_{t-1} + (\hat{\phi}_1 - \hat{\phi}_2)x_{t-2} + (\hat{\phi}_2 - \hat{\phi}_3)x_{t-3} + \hat{\phi}_3 x_{t-4} = \epsilon_t + \hat{\theta}_1 \epsilon_{t-1}$$

Separando términos, para completar el primer paso

$$x_t = (1 + \hat{\phi}_1)x_{t-1} - (\hat{\phi}_1 - \hat{\phi}_2)x_{t-2} - (\hat{\phi}_2 - \hat{\phi}_3)x_{t-3} - \hat{\phi}_3x_{t-4} + \epsilon_t + \hat{\theta}_1\epsilon_{t-1}$$

Ahora, la ecuación parece un ARIMA(4,0,1), pero sigue siendo el modelo ARIMA (3,1,1), sin embargo no puede ser un ARIMA (4,0,1) porque no cumple las condiciones de estacionalidad. Reemplazando, t con $T + 1$,

$$x_{T+1} = (1 + \hat{\phi}_1)x_T - (\hat{\phi}_1 - \hat{\phi}_2)x_{t-1} - (\hat{\phi}_2 - \hat{\phi}_3)x_{t-2} - \hat{\phi}_3x_{t-3} + \epsilon_{T+1} + \hat{\theta}_1\epsilon_T$$

Asumiendo, que tenemos observaciones hasta el tiempo T , todos los términos al lado derecho de la ecuación son conocidos a excepción de ϵ_{T+1} , el cual se reemplaza por cero, y ϵ_T que se reemplaza con el último valor de los residuos observado para ϵ_T ,

$$\hat{x}_{T+1|T} = (1 + \hat{\phi}_1)x_T - (\hat{\phi}_1 - \hat{\phi}_2)x_{t-1} - (\hat{\phi}_2 - \hat{\phi}_3)x_{t-2} - \hat{\phi}_3x_{t-3} + \hat{\theta}_1\epsilon_T$$

La predicción para x_{T+2} , se obtiene reemplazando t por $T + 2$. Todos los valores a la derecha son conocidos al tiempo T , excepto x_{T+1} , el cual es reemplazado por $\hat{x}_{T+1|T}$ y ϵ_{T+2} , ϵ_{T+1} que se reemplazan por cero,

$$\hat{x}_{t2|T} = (1 + \hat{\phi}_1)\hat{x}_{tT+1|T} - (\hat{\phi}_1 - \hat{\phi}_2)x_T - (\hat{\phi}_2 - \hat{\phi}_3)x_{T-1} - \hat{\phi}_3x_{T-2}$$

Y este proceso se continua, por el periodo que queramos considerar. Es importante resaltar, que las predicciones se basan en el supuesto de que los residuos están descorrelacionados y normalmente distribuidos. Si alguno de estos supuestos no se cumple, la predicción puede ser incorrecta. Se aconseja graficar la función de autocorrelación (ACF) y el histograma de los residuos para verificar que se cumplan estas dos hipótesis.

3.8 CRONOGRAMA

La figura 3.56 presenta el cronograma de desarrollo del presente trabajo de investigación.

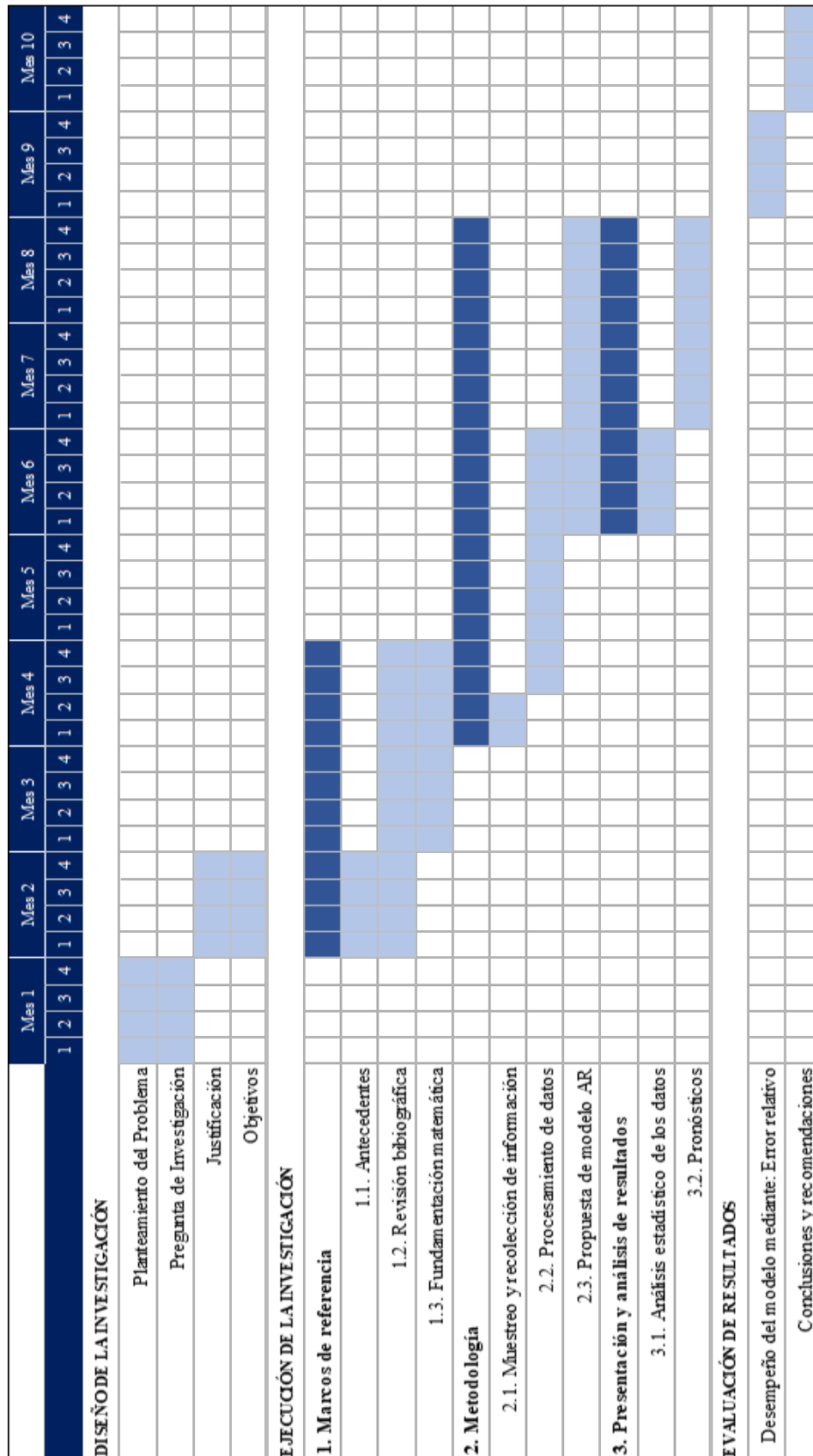


Figura 3.56: Cronograma de investigación

CAPÍTULO 4

PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

4.1 COMPARACIÓN ENTRE UNA Y MÚLTIPLES PERIODICIDADES

Al analizar las gráficas generadas a partir del modelo ETS, se observa una clara periodicidad en los datos, indicando la presencia de un patrón estacional. Sin embargo, en los casos donde los datos exhiben múltiples periodicidades, el modelo ETS puede mostrar limitaciones en la captura de efectos de tendencia más complejos. Esto puede deberse a su enfoque predominante en la estimación de las componentes de nivel y estacionalidad, lo que podría resultar menos efectivo en reflejar cambios de tendencia más sutiles o fluctuaciones irregulares. Ver figura 4.1.

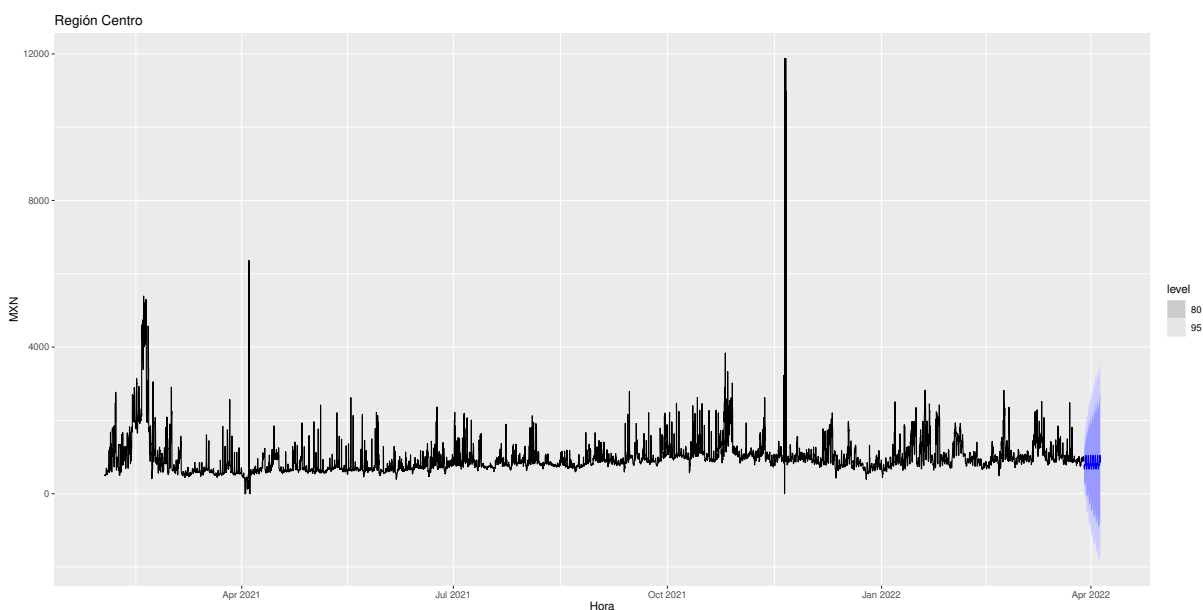


Figura 4.1: Predicción con una periodicidad.

En contraste, las gráficas generadas por un modelo SARIMA (ver figura 4.2), especialmente cuando se enfrentan a datos con múltiples periodicidades, tienden a ofrecer una representación más precisa de los efectos de la tendencia. Al incorporar la autoregresión y el promedio móvil en su formulación, el modelo SARIMA es capaz de capturar patrones más complejos y variados en los datos, incluyendo aquellos que presentan múltiples periodicidades superpuestas. Esta versatilidad permite una adaptación más eficiente a las fluctuaciones de tendencia en los datos, lo que es crucial para la generación de pronósticos más precisos y una comprensión más profunda de la dinámica subyacente en series temporales complejas.

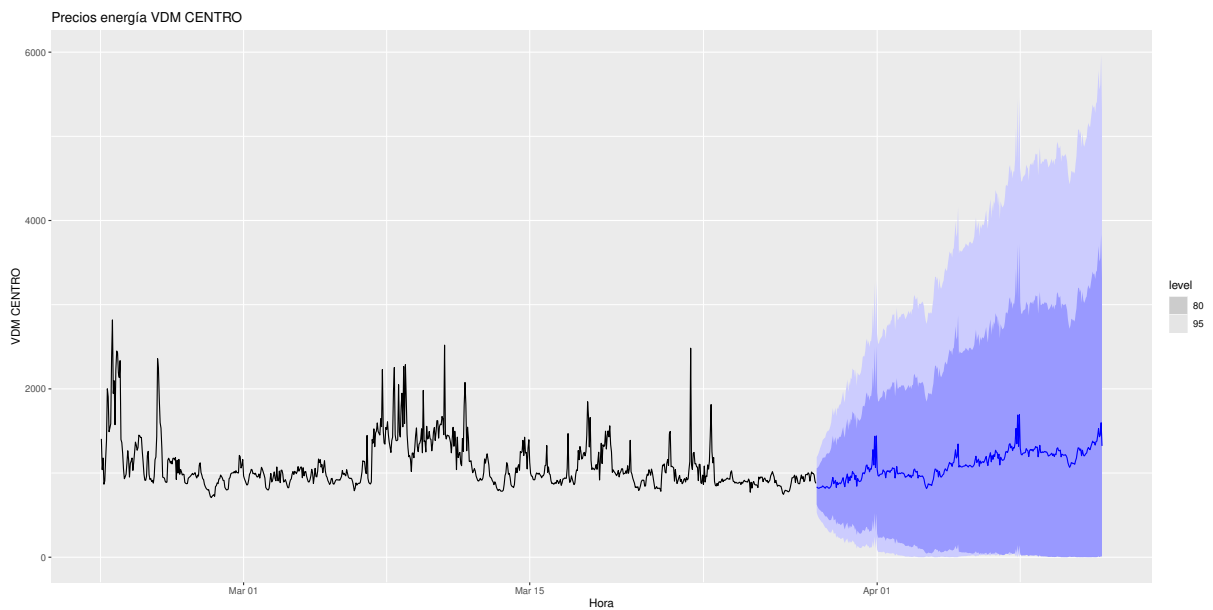


Figura 4.2: Predicción con múltiples periodicidades.

4.2 RESULTADOS DE LA PRUEBA PILOTO PARA LA SELECCIÓN DEL MODELO AUTORREGRESIVO

Dado que los resultados del modelo piloto ETS (ver figura 3.30) indicaron que únicamente capturaba de manera efectiva el componente estacional de los datos, se tomó la decisión de emprender una exploración más profunda en busca de una alternativa más robusta. Esta decisión condujo a la consideración del modelo SARIMA (Seasonal Autoregressive Integrated Moving Average), que es conocido por su capacidad para capturar tanto los componentes estacionales como los cambios en la tendencia de una serie temporal.

La elección de considerar el modelo SARIMA se basó en la necesidad de abordar los eventos de tendencia que el modelo ETS no logró reflejar con precisión. Al incorporar la autoregresión y el promedio móvil en su formulación, el modelo SARIMA permite capturar patrones de tendencia y estacionalidad más complejos, lo que puede ser especialmente valioso en situaciones donde ambas componentes son relevantes. La transición hacia un modelo SARIMA se realiza con el objetivo de mejorar la calidad de las predicciones y garantizar una representación más completa de la dinámica subyacente en los datos temporales.

4.3 ANÁLISIS DE CORRELACIÓN DE LARGO ALCANCE (EXPONENTE DE HURST)

Los resultados que se presentan en la tabla 4.1 muestran el exponente de Hurst para cada región en el conjunto de datos. El exponente de Hurst es una medida de la memoria a largo plazo de una serie temporal, donde un valor entre 0.5 y 1 indica un comportamiento persistente, y un valor entre 0 y 0.5 indica un comportamiento anti-persistente.

A partir de los resultados, podemos ver que todas las regiones tienen exponentes de Hurst mayores que 0.5, lo que indica que los datos de la serie temporal de consumo de energía en todas las regiones muestran un comportamiento persistente. Esto sugiere que el consumo de energía en estas regiones probablemente se ve afectado por tendencias e influencias a largo plazo que persisten en el tiempo.

Además, podemos observar que algunas regiones tienen exponentes de Hurst más altos que otras. Por ejemplo, las regiones de Monterrey (MTY) y Puebla (PUEBLA) tienen los exponentes de Hurst más altos, lo que indica un grado más fuerte de persistencia en los patrones de consumo de energía en comparación con otras regiones. Mientras tanto, la región de Valle de México Sur (VDM SUR) tiene el exponente de Hurst más bajo, lo que indica un grado más débil de persistencia en los patrones de consumo de energía en comparación con otras regiones.

Estos resultados pueden proporcionar información sobre el comportamiento del consumo de energía en diferentes regiones y pueden ser útiles para la planificación energética y la toma de decisiones en políticas energéticas.

La elección del exponente de Hurst a utilizar depende del contexto específico y los objetivos del análisis:

Tabla 4.1: Coeficiente de Hurst en diferentes regiones de México

Región	Simple R/S	Corregido R/S	Empírico	Empírico corregido	Teórico
VDM Centro	0.887	0.946	0.903	0.893	0.521
Monterrey	0.883	0.975	0.953	0.942	0.521
Guadalajara	0.892	0.956	0.912	0.902	0.521
Querétaro	0.889	0.954	0.910	0.900	0.521
Puebla	0.885	0.943	0.900	0.890	0.521
VDMSur	0.887	0.946	0.903	0.893	0.521
Mérida	0.891	0.950	0.909	0.898	0.521
Mexicali	0.813	0.919	0.975	0.962	0.521

- La estimación Simple R/S Hurst es el método más básico y proporciona una estimación simple del exponente de Hurst. Se asume que los datos son de largo alcance y siguen un patrón de caminata aleatoria.
- El exponente de Hurst corregido R/S es un método mejorado que corrige los sesgos en el método Simple R/S. Es una buena opción cuando se sospecha que los datos no son estrictamente de largo alcance o no siguen un patrón de caminata aleatoria.
- El exponente de Hurst empírico es un método no paramétrico que estima el exponente de Hurst basado en los datos de muestra. Es útil cuando no se quiere hacer ninguna suposición sobre el proceso subyacente que genera los datos.
- El exponente de Hurst empírico corregido es una modificación del exponente de Hurst empírico que corrige los sesgos. Es una buena opción cuando se sospecha que los datos no son estrictamente de largo alcance o no siguen un patrón de caminata aleatoria, y se quiere estimar el exponente de Hurst basado en los datos de muestra.
- El exponente de Hurst teórico es el exponente de Hurst para un proceso aleatorio estrictamente de largo alcance. Es útil como valor de referencia para comparar el exponente de Hurst empírico, pero no se espera que los datos del mundo real tengan un exponente de Hurst exactamente igual a este valor.

Con los datos de la tabla y teniendo en cuenta el tipo de análisis para calcular el coeficiente de Hurst, a continuación se presenta cómo se interpretan estos resultados en cada región:

- **VDM Centro:** Todos los métodos estiman que el exponente de Hurst está alrededor de 0.89-0.95, lo que sugiere un nivel moderado de dependencia a largo plazo en los

precios de la energía. Esta región tiene un exponente de Hurst más alto que el valor teórico de 0.521, lo que indica que los precios de la energía pueden exhibir más persistencia o tendencia que un proceso puramente aleatorio.

- **Monterrey:** Todos los métodos estiman que el exponente de Hurst está alrededor de 0.88-0.97, lo que sugiere un nivel moderado a alto de dependencia a largo plazo en los precios de la energía. Esta región también tiene un exponente de Hurst más alto que el valor teórico de 0.521, lo que indica que los precios de la energía pueden exhibir más persistencia o tendencia que un proceso puramente aleatorio.
- **Guadalajara:** Todos los métodos estiman que el exponente de Hurst está alrededor de 0.89-0.96, lo que sugiere un nivel moderado de dependencia a largo plazo en los precios de la energía. Esta región tiene un exponente de Hurst más alto que el valor teórico de 0.521, lo que indica que los precios de la energía pueden exhibir más persistencia o tendencia que un proceso puramente aleatorio.
- **Querétaro:** Todos los métodos estiman que el exponente de Hurst está alrededor de 0.89-0.95, lo que sugiere un nivel moderado de dependencia a largo plazo en los precios de la energía. Esta región tiene un exponente de Hurst más alto que el valor teórico de 0.521, lo que indica que los precios de la energía pueden exhibir más persistencia o tendencia que un proceso puramente aleatorio.
- **Puebla:** Todos los métodos estiman que el exponente de Hurst está alrededor de 0.88-0.94, lo que sugiere un nivel moderado de dependencia a largo plazo en los precios de la energía. Esta región tiene un exponente de Hurst más alto que el valor teórico de 0.521, lo que indica que los precios de la energía pueden exhibir más persistencia o tendencia que un proceso puramente aleatorio.
- **VDM Sur:** Todos los métodos estiman que el exponente de Hurst está alrededor de 0.89-0.95, lo que sugiere un nivel moderado de dependencia a largo plazo en los precios de la energía. Esta región tiene un exponente de Hurst más alto que el valor teórico de 0.521, lo que indica que los precios de la energía pueden exhibir más persistencia o tendencia que un proceso puramente aleatorio.
- **Mérida:** Todos los métodos estiman que el exponente de Hurst está alrededor de 0.89-0.95, lo que sugiere un nivel moderado de dependencia a largo plazo en los precios de la energía. Esta región tiene un exponente de Hurst más alto que el valor teórico de 0.521, lo que indica que los precios de la energía pueden exhibir más persistencia o tendencia que un proceso puramente aleatorio.
- **Mexicali:** Todos los métodos estiman que el exponente de Hurst está alrededor de 0.81-0.97, lo que sugiere un nivel de dependencia a largo plazo bajo a alto en

los precios de la energía. Esta región tiene un exponente de Hurst más alto que el valor teórico de 0.521, lo que indica que los precios de la energía pueden mostrar más persistencia o tendencia que un proceso puramente aleatorio. El método Simple R/S estima un exponente de Hurst más bajo que los otros métodos, lo que puede deberse a sesgos en el método cuando los datos no son estrictamente dependientes a largo plazo o no siguen un patrón de caminata aleatoria.

4.4 APLICACIÓN DEL MODELO AUTORREGRESIVO

La ecuación general de un modelo SARIMA se ve de la siguiente manera:

$$\text{ARIMA}(p, d, q)(P, D, Q)[s]$$

Donde:

p : Orden del componente AR. d : Orden de diferenciación no estacional. q : Orden del componente MA. P : Orden del componente AR estacional. D : Orden de diferenciación estacional. Q : Orden del componente MA estacional. s : Periodicidad estacional (número de observaciones en una temporada completa).

Donde $Y(t)$ es el valor en el tiempo t y $\varepsilon(t)$ es el término de error en el tiempo t .

Esta ecuaciones proporcionan la forma explícita para la serie de tiempo y_t como función de sus valores pasados y los términos de error. Podemos usar esta ecuación para simular la serie de tiempo, predecir valores futuros o analizar su comportamiento.

- σ representa la varianza de los residuos (errores) del modelo. En este contexto los residuos son las diferencias entre los valores observados y los valores predichos por el modelo.
- log-likelihood (logaritmos de la función de verosimilitud) es una medida que indica cuán bien se ajusta el modelo a los datos observados. En general, un valor mayor de log-likelihood indica un mejor ajuste del modelo a los datos.
- AIC. El criterios de información de Akaike es una medida de calidad del ajuste del modelo que penaliza los modelos más complejos. Un modelo con valor de AIC más bajo se considera preferible.

- AICc El AIC corregido es una versión ajustada del AIC que se utiliza para muestras pequeñas. El hecho de que el AICc tenga el mismo valor que el AIC indica que la muestra podría ser lo suficientemente grande como para no necesitar la corrección.
- BIC El criterio de información bayesiano es una medida de la calidad del ajuste del modelo que penaliza los modelos más complejos. Al igual que con el AIC, un valor más bajo del BIC indica un mejor ajuste del modelo.

En resumen, estos valores son métricas utilizadas para evaluar el ajuste del modelo a los datos y para comparar diferentes modelos. Un modelo con valores más bajos de AIC o BIC y una mayor log-likelihood se considera generalmente preferible, ya que indica un mejor ajuste a los datos observados.

4.5 MODELOS SARIMA

4.5.1 REGIÓN CENTRO AÑO 2022

A continuación se presenta la ecuación que representa el ajuste de un modelo SARIMA para los datos correspondientes a la región centro en 2022.

$$y(t) = 1068.5489 + 0.9248y_{t-1} - 0.2484\epsilon_{t-1} + 0.1547y_{t-24} + 0.1777y_{t-48} + \epsilon_t \quad (4.1)$$

- 1068.5489 es la media estimada de la serie.
- 0.9248 es el coeficiente autorregresivo para el rezago 1, lo que indica que la serie está positivamente correlacionada con su propio valor del rezago 1.
- -0.2484 es el coeficiente de media móvil para el rezago 1, lo que indica que la serie está negativamente correlacionada con el término de error del rezago 1.
- 0.1547 y 0.1777 son los coeficientes autorregresivos estacionales para los rezagos 24 y 48, respectivamente, lo que indica que la serie tiene un patrón estacional con un período de 24 horas.
- $\epsilon[t]$ es el término de error en el tiempo t , que se asume que sigue una distribución normal con una media de 0 y una varianza constante de 23188.

El modelo $SARIMA(1, 0, 1)(2, 0, 0)_{24}$ con media no nula se estima de la siguiente manera:

Coefficientes:

- Autoregresión de orden 1 (ar1): 0.9248
- Media móvil de orden 1 (ma1): -0.2484
- Autoregresión estacional de orden 1 (sar1): 0.1547
- Autoregresión estacional de orden 2 (sar2): 0.1777
- Media no nula (mean): 1068.5489

Errores estándar (s.e.):

- ar1: 0.0093
- ma1: 0.0235
- sar1: 0.0221
- sar2: 0.0216
- mean: 49.1381
- Varianza del error (σ^2) = 23188
- Log-verosimilitud = -13462.3

Criterios de información

- AIC (Criterio de Información de Akaike) = 26936.6
- AICc (AIC corregido) = 26936.64
- BIC (Criterio de Información Bayesiano) = 26970.47

Medidas de error en el conjunto de entrenamiento:

- Error medio (ME): 0.005907881
- Error cuadrático medio (RMSE): 152.0934

- Error absoluto medio (MAE): 87.39373
- Error porcentual medio (MPE): -1.405116
- Error porcentual absoluto medio (MAPE): 7.470733
- Error medio absoluto escalado (MASE): 0.4026968
- Autocorrelación del primer orden (ACF1): -0.01457401

4.5.2 REGIÓN CENTRO AÑO 2019

A continuación se utilizó una serie de la misma región pero con más datos, y estos fueron los resultados obtenidos.

$$\begin{aligned} (1 - 0.6705B - 0.135B^2 + 0.0375B^3 - 0.0347B^4 - 0.0226B^5)(1 - B)(Y_t - Y_{t-24}) \\ = (1 + 0.9829B)(e_t + 0.2055e_{t-24} + 0.1e_{t-48}) \end{aligned} \quad (4.2)$$

Expandiendo el lado izquierdo de la ecuación 4.2, y simplificando, se obtiene:

$$\begin{aligned} Y_t - Y_{t-1} - 0.6705(Y_{t-1} - Y_{t-2}) - 0.1357(Y_{t-2} - Y_{t-3}) + 0.0375(Y_{t-3} - Y_{t-4}) \\ - 0.0347(Y_{t-4} - Y_{t-5}) - 0.0226(Y_{t-5} - Y_{t-6}) - (Y_t - Y_{t-24}) + Y_{t-24} - Y_{t-25} + 0.6705(Y_{t-1} - Y_{t-25}) \\ + 0.1357(Y_{t-2} - Y_{t-26}) - 0.0375(Y_{t-3} - Y_{t-27}) + 0.0347(Y_{t-4} - Y_{t-28}) + 0.0226(Y_{t-5} - Y_{t-29}) \\ = e_t + 0.9829e_{t-1} + 0.2055e_{t-24} + 0.13e_{t-48} \end{aligned} \quad (4.3)$$

Si se quiere expresar la ecuación 4.2 Y_t de forma explícita, se tiene la ecuación 4.4

$$\begin{aligned} Y_t = Y_{t-1} + 0.6705(Y_{t-1} - Y_{t-2}) + 0.1357(Y_{t-2} - Y_{t-3}) - 0.0375(Y_{t-3} - Y_{t-4}) \\ + 0.0347(Y_{t-4} - Y_{t-5}) + 0.0226(Y_{t-5} - Y_{t-6}) + (Y_t - Y_{t-24}) - Y_{t-24} + Y_{t-25} - 0.6705(Y_{t-1} - Y_{t-25}) \\ - 0.1357(Y_{t-2} - Y_{t-26}) + 0.0375(Y_{t-3} - Y_{t-27}) - 0.0347(Y_{t-4} - Y_{t-28}) - 0.0226(Y_{t-5} - Y_{t-29}) \\ - e_t - 0.9829e_{t-1} - 0.2055e_{t-24} - 0.13e_{t-48} \end{aligned} \quad (4.4)$$

Si se quiere seguir la notación que se encuentra en la literatura,

$$ARIMA(5, 1, 1)(2, 0, 0)_{24}$$

Tabla 4.2: Coeficientes

	ar1	ar2	ar3	ar4	ar5	ma1	sar1	sar2
Coefficiente	0.6705	0.1357	-0.0375	0.0347	0.0226	-0.9829	0.2055	0.1300
s.e.	0.0253	0.0378	0.0237	0.0322	0.0110	0.0041	0.0115	0.0106

Tabla 4.3: Medidas de Error en el Conjunto de Entrenamiento (Parte 1)

	ME	RMSE	MAE	MPE
Conjunto de entrenamiento	0.2648804	144.5113	74.68036	-Inf

Tabla 4.4: Medidas de Error en el Conjunto de Entrenamiento (Parte 2)

	MAPE	MASE	ACF1
Conjunto de entrenamiento	Inf	0.4416624	-0.0001691691

De donde,

$$\sigma^2 = 20905, \quad \text{Varianza de los residuos del modelo}$$

Log-likelihood = -55992.06 , Medida del ajuste del modelo a los datos observados

$AIC = 112002.1$, Criterio de información de Akaike (más bajo es mejor)

$AICc = 112002.1$, Criterio de información de Akaike corregido

$BIC = 112065.8$, Criterio de información bayesiano (más bajo es mejor)

Un modelo con valores más bajos de AIC o BIC y una mayor log-likelihood se considera generalmente preferible, ya que indica un mejor ajuste a los datos observados.

4.5.3 REGIÓN CENTRO AÑO 2020

Este año fue particular pues coincido con la pandemia, y modifíco los parámetros de producción-consumo.

$$(1 - 0.8134B)(1 - B)(Y_t - Y_{t-24}) = (1 + 0.7092B - 0.0525B^2)(e_t + 0.2186e_{t-24} + 0.16e_{t-48})$$

donde B es el operador de retraso (**backshift operator**).

Para aislar Y_t en el lado izquierdo de la ecuación, podemos reorganizar la forma explícita del modelo $ARIMA(1, 1, 2)(2, 0, 0)_{24}$ de la siguiente manera:

$$Y_t = Y_{t-1} + Y_{t-24} - Y_{t-25} - (-0.8134(Y_{t-1} - Y_{t-2}) + e_t + 0.7092e_{t-1} - 0.0525e_{t-2} + 0.2186e_{t-24} + 0.16e_{t-48}) \quad (4.5)$$

Simplificando la ecuación 4.5, se obtiene la ecuación 4.6

$$Y_t = 0.1866Y_{t-1} + Y_{t-2} - 0.1866Y_{t-3} + e_t - 0.7092e_{t-1} + 0.0525e_{t-2} - 0.2186e_{t-24} - 0.16e_{t-48} + Y_{t-24} + Y_{t-25} \quad (4.6)$$

Esta ecuación muestra que el valor actual de la serie de tiempo Y_t es una función lineal de sus valores pasados Y_{t-1} , Y_{t-2} , Y_{t-3} y los términos de error, con coeficientes y retardos especificados por el modelo SARIMA. La ecuación también incluye los valores de la serie de tiempo de hace 24 y 25 periodos Y_{t-24} , Y_{t-25} en el lado derecho, reflejando el componente estacional de los datos.

Si se quiere seguir la notación que se encuentra en la literatura,

$$ARIMA(1, 1, 2)(2, 0, 0)_{24}$$

Este modelo sugiere que la serie de tiempo ha sido diferenciada estacionalmente una vez ($d=1$) con un retraso de 24 periodos. También incluye un término autorregresivo (AR) de orden 1, y términos de media móvil (MA) de orden 1 y 2. Además, el modelo incluye términos AR estacionales de hasta orden 2.

Tabla 4.5: Coeficientes

	ar1	ma1	ma2	sar1	sar2
Coefficient	-0.8134	0.7092	-0.0525	0.2186	0.1600
s.e.	0.0601	0.0611	0.0151	0.0109	0.0108

Los coeficientes de los términos AR y MA tienen signos opuestos, lo que indica que la serie de tiempo tiene una autocorrelación negativa en el retraso 1 y una autocorrelación positiva en los retrasos 1 y 2. Los coeficientes AR estacionales son positivos, lo que indica que la serie de tiempo presenta patrones estacionales en los retrasos de 24 y 48 periodos.

De donde,

Tabla 4.6: Medidas de Error en el Conjunto de Entrenamiento (Parte 1)

	ME	RMSE	MAE	MPE
Conjunto de entrenamiento	0.00903	87.81968	46.79632	-0.69937

Tabla 4.7: Medidas de Error en el Conjunto de Entrenamiento (Parte 2)

	MAPE	MASE	ACF1
Conjunto de entrenamiento	6.8566	0.35465	0.0002698

$\sigma^2 = 7718$, Varianza de los residuos del modelo

Log-likelihood = -51770.95 , Medida del ajuste del modelo a los datos observados

$AIC = 103553.9.1$, Criterio de información de Akaike (más bajo es mejor)

$AICc = 103553.9$, Criterio de información de Akaike corregido

$BIC = 103596.4$ Criterio de información bayesiano (más bajo es mejor)

El valor de log-verosimilitud y los valores de AIC/BIC indican que el modelo proporciona un ajuste razonable a los datos. Las medidas de error en el conjunto de entrenamiento también muestran que el modelo captura bien la variabilidad en los datos, con un bajo error cuadrático medio (RMSE) y error absoluto medio (MAE) en relación con el valor medio de la serie de tiempo.

4.5.4 RESUMEN DE LOS RESULTADOS

En el presente capítulo, se introduce un análisis exhaustivo de la variabilidad de los precios de la energía en distintas regiones a lo largo de los años, a través de la utilización de modelos SARIMA. Estos modelos constituyen una herramienta sólida y establecida en el análisis de series de tiempo, permitiendo examinar de manera precisa cómo las condiciones pasadas y estacionales influyen en los valores futuros de los precios energéticos. A lo largo de este capítulo, se presentan las ecuaciones matemáticas resultantes de los modelos SARIMA ajustados para cinco regiones específicas: VDM CENTRO, MTY, GDL, MERIDA y MEXICALI. Dichas ecuaciones brindan una comprensión detallada de las interrelaciones inherentes en los datos de series temporales, cimentando una base sólida para la predicción y el análisis de la variabilidad energética en cada una de estas regiones particulares. Además, se exponen los resúmenes de los resultados de los modelos, así como los coeficientes asociados a cada componente, contribuyendo así a un conocimiento más

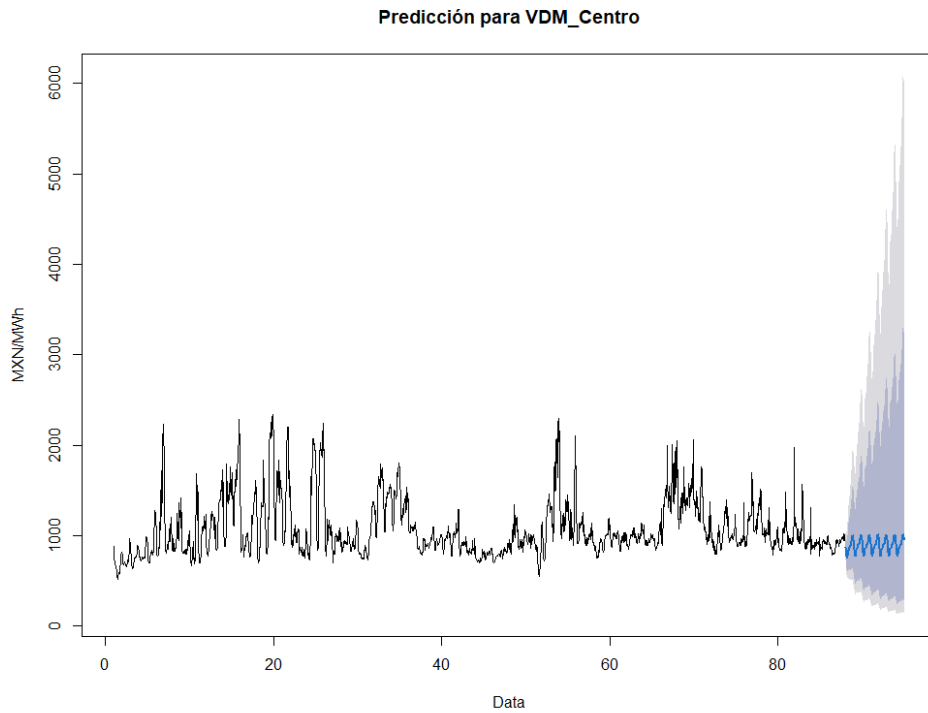


Figura 4.3: Modelo Región VDM Centro

profundo de los factores que moldean los patrones de precios energéticos en contextos regionales precisos.

■ **VDM Centro:**

$$\begin{aligned}
 Y(t) = & 1068.5489 + 0.9248 \cdot Y(t - 1) - 0.2484 \cdot \varepsilon(t - 1) \\
 & + 0.1547 \cdot \varepsilon(t - 24) + 0.1777 \cdot \varepsilon(t - 48) \\
 & + \varepsilon(t)
 \end{aligned}$$

Esto significa que el valor actual de la serie en la región VDM Centro depende fuertemente de su propio valor en el período anterior, así como de la influencia de valores pasados en los lags 24 y 48. Además,

1. El coeficiente asociado a $\varepsilon(t - 1)$ es negativo (-0.2484). Esto indica que el modelo está teniendo en cuenta los errores de predicción pasados para corregir las predicciones actuales. Si hay una gran discrepancia entre el valor real y la predicción anterior, el modelo ajustará la predicción actual en consecuencia.

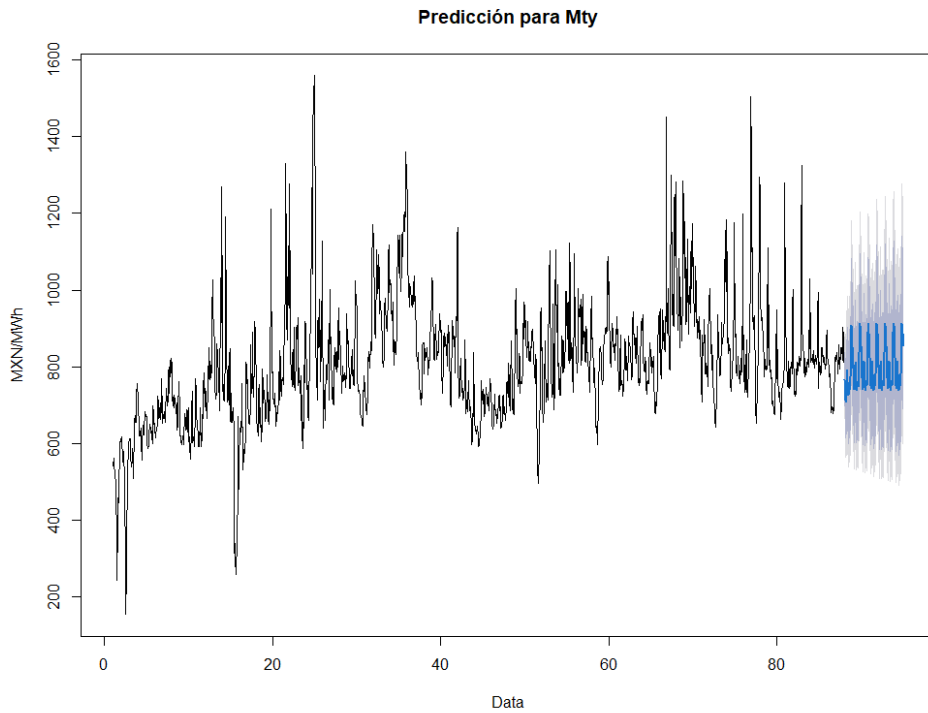


Figura 4.4: Modelo Región Monterrey

2. el término independiente 1068.5489 y el término de error $\varepsilon(t)$ (representa la variabilidad no explicada por los componentes del modelo) también contribuyen a la predicción.

■ **Monterrey (MTY):**

$$\begin{aligned}
 Y(t) = & -0.2942 \cdot Y(t-1) - 0.1186 \cdot Y(t-2) - 0.1009 \cdot Y(t-3) \\
 & - 0.1036 \cdot Y(t-4) - 0.0780 \cdot Y(t-5) \\
 & + 0.2098 \cdot \varepsilon(t-24) + 0.1815 \cdot \varepsilon(t-48) + \varepsilon(t)
 \end{aligned}$$

En este modelo sugiere que los precios de la electricidad en esta región están fuertemente influenciados por los valores pasados y los errores anteriores, con una reversión a la media pronunciada. Además, hay patrones estacionales que están relacionados con los errores en períodos específicos.

■ **Guadalajara (GDL):**

$$Y(t) = 1003.0087 + 0.9142 \cdot Y(t-1) - 0.2313 \cdot \varepsilon(t-1)$$

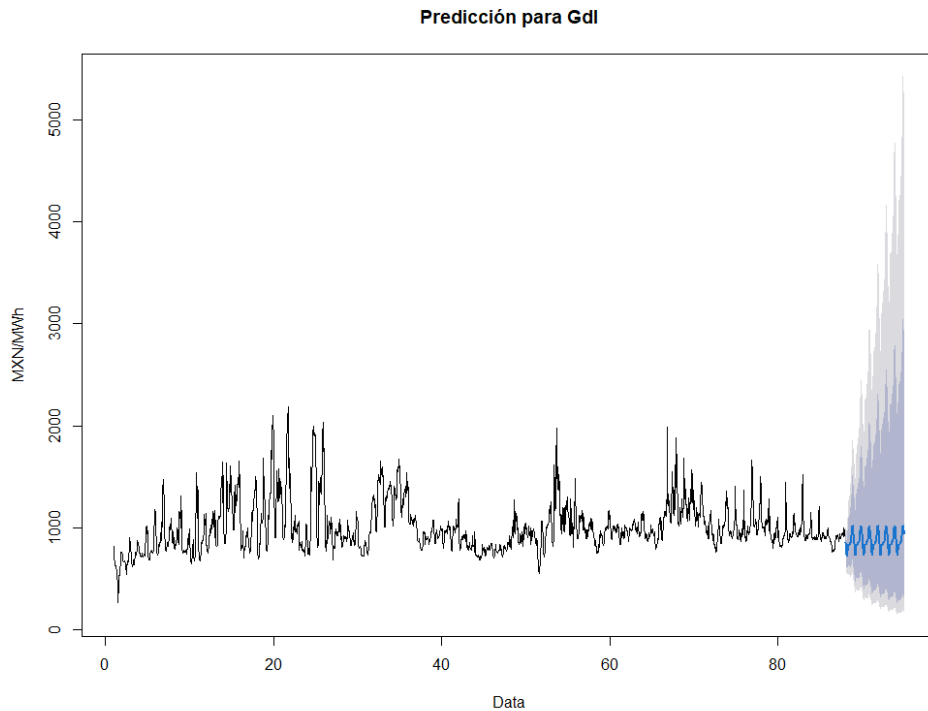


Figura 4.5: Modelo Región Guadalajara

$$+ 0.1639 \cdot \varepsilon(t - 24) + 0.2073 \cdot \varepsilon(t - 48) + \varepsilon(t)$$

1. El coeficiente 0.9142 asociado a $Y(t - 1)$ indica una correlación positiva fuerte entre el precio actual y el precio anterior. Esto sugiere que los precios tienden a mantener cierta tendencia o inercia de un período a otro.
2. El coeficiente -0.2313 asociado a $\varepsilon(t - 1)$ indica que el error en el período anterior tiene un efecto negativo en el precio actual. Esto podría indicar cierta corrección o ajuste que ocurre en los precios después de un error.
3. Los coeficientes positivos 0.1639 y 0.2073 asociados a los errores retardados en 24 y 48 períodos, respectivamente, indican que los errores pasados en esos momentos tienen un efecto positivo en el precio actual. Esto sugiere que hay patrones estacionales en los precios.
4. La presencia de términos $\varepsilon(t - 24)$ y $\varepsilon(t - 48)$ indica una clara componente estacional en los precios, con un ciclo de estacionalidad de 24 períodos.

Este modelo sugiere que los precios de la electricidad en esta región están influenciados por tendencias pasadas, errores anteriores y patrones estacionales, con una fuerte correlación positiva entre los precios actuales y los precios inmediatamente anteriores.

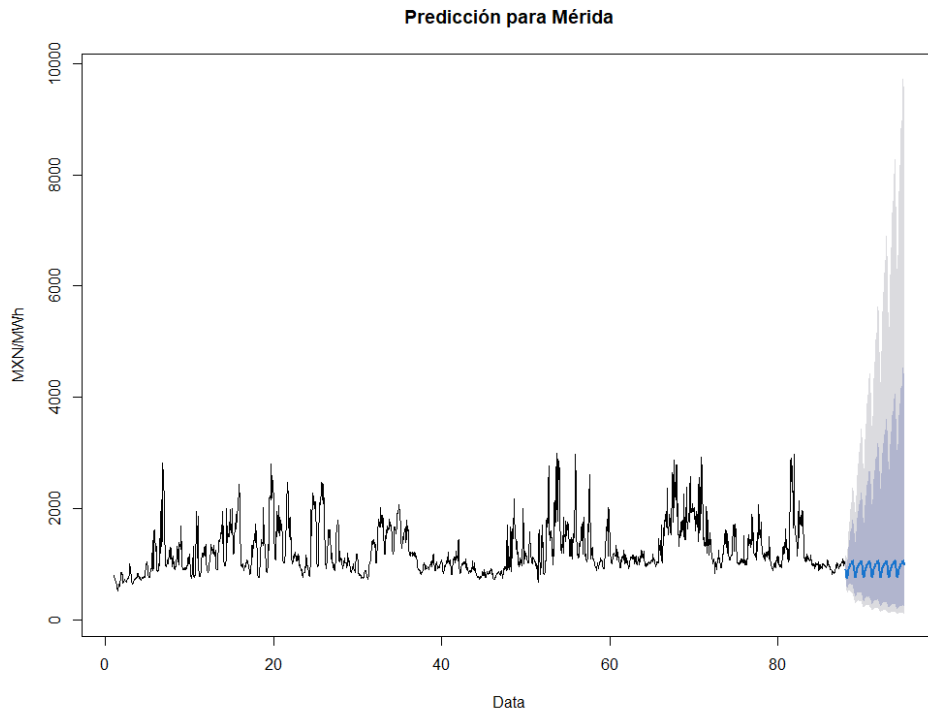


Figura 4.6: Modelo Región Merida

■ Mérida (MER):

$$Y(t) = 0.8017 \cdot Y(t - 1) - 0.9666 \cdot \varepsilon(t - 1) + 0.1041 \cdot \varepsilon(t - 24) + 0.0438 \cdot \varepsilon(t - 48) + \varepsilon(t)$$

1. El coeficiente AR1 es positivo (0.8017), lo que indica una fuerte correlación positiva entre los precios de MWh en momentos consecutivos. Esto sugiere que si el precio de MWh aumenta en un periodo, es probable que el siguiente periodo también experimente un aumento.
2. El coeficiente asociado a $\varepsilon(t - 1)$ es negativo (-0.9666). Esto indica que el modelo está tomando en cuenta los errores de predicción pasados para corregir las predicciones actuales. Si hay una gran discrepancia entre el valor real y la predicción anterior, el modelo ajustará la predicción actual en consecuencia.
3. Los coeficientes de los términos $\varepsilon(t - 24)$ y $\varepsilon(t - 48)$ son positivos (0.1041 y 0.0438, respectivamente). Esto sugiere que hay patrones estacionales en los precios de MWh, y estos patrones son influenciados por los errores de predicción pasados en los mismos periodos estacionales.

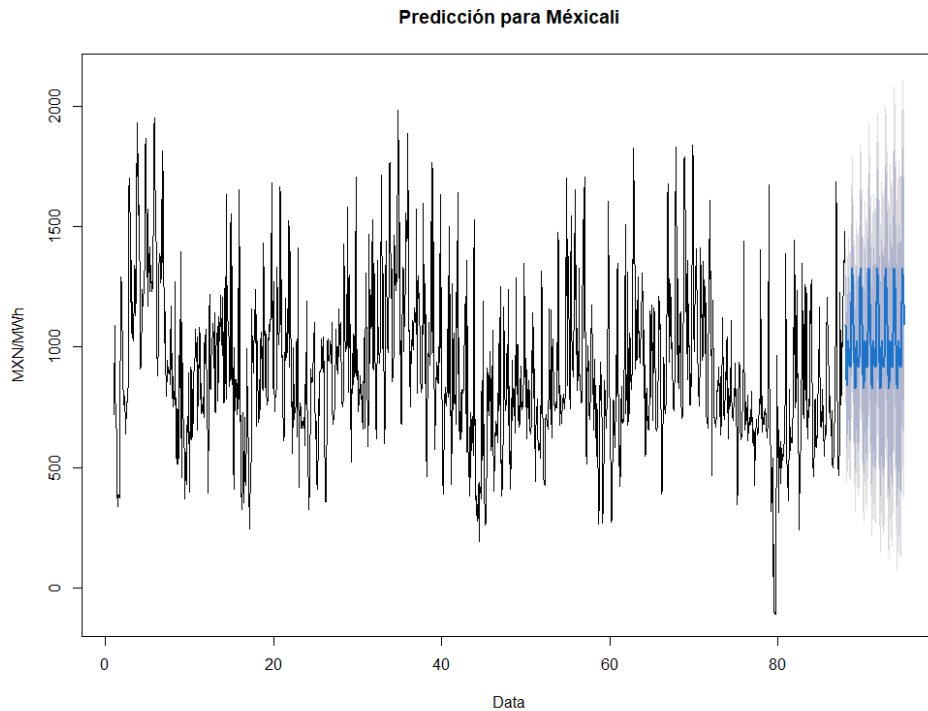


Figura 4.7: Modelo Región Mexicali

4. La presencia de términos $\varepsilon(t - 24)$ y $\varepsilon(t - 48)$ indica una clara componente estacional en los precios, con un ciclo de estacionalidad de 24 períodos.

Puede haber una correlación entre los precios de MWh y los precios de los insumos o combustibles utilizados en las termoeléctricas, donde e inferiendo de las gráfica 4.6 la variabilidad en los precios de MWh puede estar sujeta a patrones estacionales, posiblemente relacionados con la demanda de energía en diferentes momentos del año.

■ **Mexicali (MEX):**

$$Y(t) = 0.6047 \cdot Y(t - 1) - 0.9822 \cdot \varepsilon(t - 1) + 0.0694 \cdot Y(t - 2) + 0.1739 \cdot \varepsilon(t - 24) + 0.1516 \cdot \varepsilon(t - 48) + \varepsilon(t)$$

1. El coeficiente **AR1** es positivo (0.6047), lo que indica una correlación positiva entre los precios de MWh en momentos consecutivos. Esto sugiere que si el precio de MWh aumenta en un periodo, es probable que el siguiente periodo también experimente un aumento.

2. El coeficiente ma1 asociado a $\text{varepsilon}(t - 1)$ es negativo (-0.9822). Esto indica que el modelo está tomando en cuenta los errores de predicción pasados para corregir las predicciones actuales. Si hay una gran discrepancia entre el valor real y la predicción anterior, el modelo ajustará la predicción actual en consecuencia.
3. Hay un término con un rezago de 2 periodos de tiempo $Y(t - 2)$ con un coeficiente positivo (0.0694). Esto indica una correlación positiva entre los precios de MWh separados por dos periodos de tiempo.
4. Los coeficientes de los términos $\varepsilon(t - 24)$ y $\varepsilon(t - 48)$ son positivos (0.1739 y 0.1516, respectivamente). Esto sugiere que hay patrones estacionales en los precios de MWh, y estos patrones son influenciados por los errores de predicción pasados en los mismos periodos estacionales.

Estas ecuaciones matemáticas revelan cómo las condiciones pasadas y los componentes estacionales influyen en los valores futuros de los precios energéticos en cada región específica. Cada coeficiente representa la fuerza y dirección de la relación entre las observaciones anteriores y los valores futuros. Además, los términos de error $\varepsilon(t)$ capturan las fluctuaciones no explicadas por el modelo y desempeñan un papel esencial en la precisión de las predicciones. A medida que avanzamos en este estudio, estas ecuaciones establecen la base para una comprensión más profunda de las dinámicas energéticas regionales y servirán como herramientas fundamentales para futuros análisis y pronósticos en el sector energético.

CONCLUSIONES

1. Mediante la aplicación de modelos SARIMA, este estudio ha logrado evidenciar las interconexiones fundamentales entre condiciones históricas, patrones estacionales y los valores energéticos futuros en diversas regiones. Los resultados, expresados a través de ecuaciones matemáticas con coeficientes representativos, han ilustrado la influencia de componentes autoregresivos y de promedio móvil en la formación de los precios de energía, al mismo tiempo que han abordado la inherente incertidumbre a través de términos de error. Estos descubrimientos no solo enriquecen la comprensión de los patrones energéticos, sino también establecen una base sólida para decisiones estratégicas y pronósticos informados en el ámbito energético regional, contribuyendo así a una gestión más eficaz y sostenible de los recursos energéticos en diferentes localidades.
2. Al examinar las tendencias históricas de los precios horarios de la energía eléctrica en el mercado mexicano, se llega a la conclusión de que existe un patrón periódico en su comportamiento, especialmente cuando se excluyen los datos atípicos. Esta observación resalta la presencia de ciclos recurrentes en los precios y sugiere la influencia de factores estacionales o temporales en su variabilidad. La identificación de este comportamiento periódico es esencial para una comprensión más profunda de la dinámica del mercado energético y puede ofrecer información valiosa para la toma de decisiones estratégicas. No obstante, es importante tener en cuenta que el análisis y la mitigación adecuada de los datos atípicos son elementos críticos para respaldar la validez de estas conclusiones. En conjunto, este estudio subraya la importancia de considerar tanto los patrones periódicos como la presencia de valores extremos al examinar los precios horarios de la energía eléctrica en el mercado mexicano.
3. Se ha observado de manera consistente que existen períodos específicos a lo largo del período de análisis de 4 años en los cuales los precios de la energía eléctrica presentan valores más elevados en comparación con la media anual. Estos picos de precios se repiten en meses determinados, sugiriendo una pauta recurrente en la variación de precios. La identificación de estos meses con precios más altos en relación

con la media a lo largo de múltiples años destaca la importancia de las estacionalidades en la dinámica de los precios energéticos. Estos hallazgos son esenciales para la comprensión y la planificación estratégica en el sector energético, ya que indican momentos específicos en los cuales los costos de energía tienden a ser más pronunciados y pueden influir en la toma de decisiones informadas.

4. Las posibles perturbaciones ambientales o sociopolíticas demostraron ser factores exógenos de gran relevancia que desencadenaron un aumento considerable en los precios horarios de la energía eléctrica, tal como se evidenció en este análisis. Tomando como ejemplo la ola de frío que afectó el sur de Estados Unidos y el norte de México al comienzo del año 2021, así como las elecciones presidenciales de julio de 2018, se constata cómo estos eventos tuvieron un impacto significativo en los precios. Durante ambas situaciones, se observó un incremento de hasta cinco veces la media de los precios horarios en comparación con el resto del año. Estos ejemplos subrayan cómo las perturbaciones externas, ya sean de carácter climático o sociopolítico, pueden ejercer una influencia determinante en la dinámica de los precios energéticos, aportando una dimensión adicional a la comprensión de los factores que impulsan la variabilidad de los costos de energía.
5. En el contexto de los precios de la energía para diferentes regiones en México, el hecho de que el pico sea más intenso a 12 horas en comparación con 24 horas, o que estén casi iguales, o que sea lo contrario, está relacionado con los patrones de consumo eléctrico y las fluctuaciones en la demanda a lo largo del día.
 - **Pico más intenso a 12 horas que a 24 horas:** (Región Centro, Guadalajara y Mérida) Pico más intenso a 12 horas que a 24 horas: Esto podría indicar que en esas regiones, hay una mayor demanda de energía en las horas cercanas al mediodía (12 horas) en comparación con las horas de la tarde y la noche (24 horas). Este patrón podría estar influenciado por actividades industriales, comerciales y domésticas que tienden a aumentar durante las horas laborales y disminuir en la noche.
 - **Picos de intensidad similar a 12 y 24 horas:**(Región Monterrey) Esto podría indicar que la demanda de energía es relativamente constante a lo largo del día, sin una variación significativa entre las horas cercanas al mediodía y las horas de la tarde y la noche. Podría ser característico de una región con un equilibrio en la actividad económica y la demanda de energía durante esas horas.
 - **Pico más intenso a 24 horas que a 12 horas:** (Región Mexicali) Pico más intenso a 24 horas que a 12 horas: Esto podría sugerir que en esas regiones,

la demanda de energía es más alta durante las horas nocturnas (24 horas) en comparación con las horas cercanas al mediodía (12 horas). Este patrón podría estar relacionado con el consumo residencial y la iluminación en la noche, así como con la reducción de la demanda en actividades comerciales e industriales.

Estos hallazgos ofrecen una comprensión más profunda de las complejidades que influyen en los precios de la energía en diferentes regiones. Asimismo, subrayan la importancia de considerar factores económicos, ambientales e industriales específicos de cada región al analizar los datos de precios de energía. La capacidad de identificar estos patrones y tendencias en los precios puede ser valiosa para la toma de decisiones informada en la gestión de la oferta y demanda de energía, así como para la formulación de estrategias que mitiguen la volatilidad de los precios en el mercado energético.

6. En regiones como Monterrey y en menor medida Mérida se observan armónicos en los picos y en los componentes que les siguen, esto sugiere que ciertas fluctuaciones en los precios de la energía están siguiendo un patrón que es una múltiple de la frecuencia principal. En otras palabras, hay variaciones en los precios que se repiten en intervalos regulares, como múltiplos de las frecuencias dominantes en la serie.

La presencia de armónicos puede ser indicativa de patrones de consumo repetitivos y regulares en la demanda de energía en esas regiones. Esto puede deberse a diversos factores, como horarios de trabajo, ciclos económicos y comportamientos de consumo típicos en esas áreas. La identificación de armónicos puede proporcionar información valiosa sobre la estructura subyacente de la serie de tiempo y cómo los factores externos pueden estar influyendo en los precios de la energía.

7. La conclusión obtenida a partir de la predicción generada por el modelo ETS revela su destreza en la captura precisa de patrones estacionales presentes en los datos. Sin embargo, es importante destacar que su capacidad para anticipar eventos de tendencia no muestra la misma robustez. Esta limitación puede ser atribuida a la naturaleza inherente del modelo ETS, que tiende a priorizar las variaciones estacionales en lugar de las tendencias de largo plazo.
8. El modelo SARIMA se ha establecido como la elección más adecuada para el mercado eléctrico mexicano, debido a su capacidad para manejar datos con periodicidades diarias y semanales. En efecto, las pruebas piloto realizadas utilizando un modelo ARIMA muestran de manera visual que no es la opción que mejor se ajusta a los datos disponibles. La elección del modelo SARIMA resalta su capacidad para capturar las variaciones cíclicas y estacionales presentes en los precios de la energía eléctrica

en el mercado mexicano, lo que mejora significativamente la capacidad predictiva y la comprensión de la dinámica subyacente en los precios.

9. El grado máximo del polinomio resultante es 2. Un polinomio de mayor grado puede capturar con mayor precisión detalles complejos y variaciones en los datos, pero existe el riesgo de sobreajuste, donde el polinomio se adapta excesivamente a ruido o valores atípicos en los datos de muestra. Por otro lado, un polinomio de grado inferior puede ser más estable pero puede subestimar la complejidad real de los datos.
10. La elección del horizonte de predicción tiene un impacto de gran relevancia en la incertidumbre inherente al modelo y en la precisión de las predicciones generadas. A medida que se extiende el período de pronóstico, la incertidumbre asociada aumenta de manera proporcional.
11. La región de Mexicali se distingue por su comportamiento único, atribuido a su estrecha conexión eléctrica con Estados Unidos y la apertura al intercambio comercial resultante de esta situación.
12. Dado su fuerte dependencia en combustibles fósiles, Mérida afronta costos más elevados. Sin embargo, esta dependencia también le otorga un patrón más predecible en contraste con otras regiones, lo que se traduce en una menor frecuencia de observaciones atípicas en su comportamiento.

RECOMENDACIONES

- Ajustar un modelo que analice de manera individual los datos atípicos de la serie temporal. Esto reducirá la incertidumbre en las predicciones al mitigar el efecto de valores extremos, mejorando la precisión y confiabilidad del pronóstico.
- Considerar métodos más sólidos, como el enfoque TBATS, para la predicción de series de tiempo con múltiples periodicidades. El método TBATS ha demostrado ser eficaz en la modelización de patrones complejos y estacionales en los datos. Su implementación podría mejorar significativamente la precisión de las predicciones al capturar de manera más efectiva las diferentes periodicidades presentes en la serie. Esta aproximación más robusta podría enriquecer el análisis y pronóstico de series temporales, como los precios de la energía.
- Expandir el análisis a modelos que integren variables exógenas, tales como registros históricos de temperaturas, precios de materias primas, efectos climáticos, etc., con el fin de evaluar posibles correlaciones y su influencia en el análisis. La incorporación de estas variables externas puede proporcionar una comprensión más completa y precisa de los factores que impactan en la serie de tiempo en consideración. Explorar las relaciones y tendencias entre estas variables externas y la serie de interés puede enriquecer significativamente el modelo predictivo, permitiendo una mejor interpretación y pronóstico de las fluctuaciones en los precios. Esta metodología ampliada abre la puerta a un análisis más integral y fundamentado, con potencial para identificar patrones ocultos y relaciones causales que podrían contribuir a la toma de decisiones informada en el ámbito energético.
- Explorar a predicción de costos de energía mediante el empleo de la Transformada Rápida de Fourier (FFT). Esta técnica ofrece la capacidad de descomponer la serie temporal de costos en sus componentes de frecuencia, permitiendo la identificación de patrones cíclicos y estacionales en los datos. Al considerar la FFT para la predicción, se puede capturar eficazmente la estructura periódica de los costos de energía y utilizarla para generar pronósticos más informados. Este enfoque puede

enriquecer el análisis predictivo al incorporar las propiedades de frecuencia de los datos, brindando insights valiosos para la toma de decisiones en la gestión de costos energéticos.

- Utilizar modelo de Espacio de Estado con componentes de frecuencia, La esencia de estos modelos radica en la descomposición de la serie en componentes observadas y no observadas, lo que permite incorporar características temporales complejas. Es recomendable explorar la incorporación de componentes de frecuencia y armónicos en los estados no observados del modelo de espacio de estado. Este enfoque permitirá capturar con precisión las variaciones a lo largo del tiempo en diferentes frecuencias y proporcionar pronósticos más certeros para los precios de energía.

APÉNDICE A

SOLUCIÓN DEL MODELO
AUTORREGRESIVO AR DE PRIMER
ORDEN

A continuación se presenta el desarrollo de la solución del modelo autorregresivo AR de primer orden de manera exacta:

$$x_t = \phi x_{t-1} + \epsilon_t; \quad \epsilon_t \sim N(0, \sigma^2); \quad \forall t \geq 0; \quad (\text{A.1})$$

$$x_1 = \phi x_0 + \epsilon_1$$

$$x_2 = \phi x_1 + \epsilon_2 = \phi^2 x_0 + \phi \epsilon_1 + \epsilon_2$$

⋮

$$x_n = \phi^n x_0 + \phi^{n-1} \epsilon_1 + \phi^{n-2} \epsilon_2 + \dots + \phi \epsilon_{n-1} + \epsilon_n$$

$$x_n = \phi^n x_0 + \sum_{k=1}^n \phi^{n-k} \epsilon_k \quad (\text{A.2})$$

Comentario! A.1 y A.2

$$E[x_n] \equiv \langle x_n \rangle = \phi^n x_0 \quad (\text{A.3})$$

$$\phi = \left[\frac{\langle x_n \rangle}{x_0} \right]^{1/n} \quad (\text{A.4})$$

$$Var(x_n) = E[x_n^2] - E^2[x_n] \quad (\text{A.5})$$

Comentario!

$$E^2[x_n] = \left(\phi^n x_0 + \sum_{k=1}^n \phi^{n-k} \epsilon_k \right)^2 \quad (\text{A.6})$$

$$\begin{aligned} &= \left(\phi^n x_0 + \sum_{k=1}^n \phi^{n-k} \epsilon_k \right) \left(\phi^n x_0 + \sum_{l=1}^n \phi^{n-l} \epsilon_l \right) \\ &= \phi^{2n} x_0^2 + 2\phi^n x_0 \sum_{l=1}^n \phi^{n-l} \epsilon_l + \phi^{2n} \sum_{l=1}^n \sum_{k=1}^n \phi^{-(l+k)} \epsilon_k \epsilon_l \\ &= \phi^{2n} x_0^2 + \phi^{2n} \sum_{l=1}^n \sum_{k=1}^n \phi^{-(l+k)} \sigma^2 \delta_{kl} \end{aligned} \quad (\text{A.7})$$

Donde δ_{kl} corresponde a la función delta de Kronecker y que se define en la ecuación A.8, así:

$$\delta_{kl} = \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases} \quad (\text{A.8})$$

De este modo:

$$\text{Var}(x_n) = \phi^{2n} x_0^2 + \phi^{2n} \sigma^2 \sum_{j=1}^n \phi^{-2j} - \phi^{2n} x_0^2 \quad (\text{A.9})$$

$$= \sigma^2 \phi^{2n} \sum_{j=1}^n \phi^{-2j} \quad (\text{A.10})$$

Utilizando la transformación A.11:

$$\frac{1}{\phi^2} \equiv \alpha \quad (\text{A.11})$$

$$\sum_{j=1}^n \alpha^j = \alpha + \alpha^2 + \dots + \alpha^n \quad (\text{A.12})$$

Por lo tanto, la sumatoria se puede escribir como la expresión A.13

$$\begin{aligned}
 S &= \alpha + \alpha^2 + \cdots + \alpha^n \\
 \alpha S &= \alpha^2 + \alpha^3 \cdots + \alpha^{n+1} \\
 \alpha S - S &= -\alpha + \alpha^{n+1} \\
 S(\alpha - 1) &= \alpha^{n+1} - \alpha \\
 S &= \sum_{j=1}^n \alpha^j = \frac{\alpha(\alpha^n - 1)}{\alpha - 1}
 \end{aligned} \tag{A.13}$$

Reemplazando α de la ecuación A.11 en A.13, se tiene la expresión en términos de ϕ , así:

$$\begin{aligned}
 \sum_{j=1}^n \alpha^j &= \frac{\frac{1}{\phi^2} \left(\frac{1}{\phi^{2n}} - 1 \right)}{\frac{1}{\phi^2} - 1} \\
 &= \frac{\frac{1}{\phi^2} \left(\frac{1 - \phi^{2n}}{\phi^{2n}} \right)}{\frac{1 - \phi^2}{\phi^2}} \\
 &= \frac{\frac{1 - \phi^{2n}}{\phi^{2n+2}}}{\frac{1 - \phi^2}{\phi^2}} \\
 &= \frac{\phi^2(1 - \phi^{2n})}{\phi^2 \phi^{2n} (1 - \phi^2)} \\
 &= \frac{1 - \phi^{2n}}{\phi^{2n} (1 - \phi^2)}
 \end{aligned} \tag{A.14}$$

Retomando la ecuación A.10, se obtiene que:

$$\begin{aligned}
 Var(x_n) &= \sigma^2 \phi^{2n} \left(\frac{1 - \phi^{2n}}{\phi^{2n} (1 - \phi^2)} \right) \\
 Var(x_n) &= \frac{\sigma^2 (1 - \phi^{2n})}{1 - \phi^2}, \quad \phi \neq 1
 \end{aligned} \tag{A.15}$$

Para la convergencia del modelo autorregresivo de orden 1 AR(1), se requiere que $|\phi| < 1$.

Además si,

$$\lim_{n \rightarrow \infty} \text{Var}(x_n) = \frac{\sigma^2}{1 - \phi^2} \quad (\text{A.16})$$

Se obtiene la expresión general A.16 de la varianza de un modelo autorregresivo de orden 1.

A continuación se presenta el cálculo de la **covarianza** para este mismo caso:

$$\text{Cov}(x_m, x_n) = E[x_m X_n] - E[x_m]E[x_n], \quad m \leq n \quad (\text{A.17})$$

$$= E \left[\left(\phi^m x_0 + \sum_{k=1}^m \phi^{m-k} \epsilon_k \right) \left(\phi^n x_0 + \sum_{l=1}^n \phi^{n-l} \epsilon_l \right) \right] - E[x_m]E[x_n] \quad (\text{A.18})$$

$$= E \left[\phi^{m+n} x_0^2 + \phi^m x_0 \sum_{l=1}^n \phi^{n-l} \epsilon_l + \phi^n x_0 \sum_{k=1}^m \phi^{m-k} \epsilon_k + \sum_{k=1}^m \sum_{l=1}^n \phi^{m+n-(l+k)} \epsilon_k \epsilon_l \right] - \phi^m x_0 \phi^n x_0 \quad (\text{A.19})$$

$$= \phi^{m+n} x_0^2 + \sum_{k=1}^m \sum_{l=1}^n \phi^{m+n-(l+k)} E[\epsilon_l \epsilon_k] - \phi^{m+n} x_0^2 = \sum_{k=1}^m \sum_{l=1}^n \phi^{m+n-(l+k)} \sigma^2 \delta_{kl} \quad (\text{A.20})$$

$$= \sigma^2 \phi^{m+n} \sum_{k=1}^m \sum_{l=1}^n \phi^{-(l+k)} \quad (\text{A.21})$$

Cambio de índice por $\min\{m, n\}$

$$\text{Cov}(x_m, x_n) = \sigma^2 \phi^{m+n} \sum_{j=1}^m \phi^{-2j} \quad (\text{A.22})$$

Utilizando A.11, A.12 y A.13:

$$\begin{aligned}
 \sum_{j=1}^m \alpha^j &= \frac{\alpha(\alpha^m - 1)}{\alpha - 1} \\
 &= \frac{1 - \phi^{2m}}{\phi^{2m}(1 - \phi^2)} = \frac{1 - \phi^{2m}}{\phi^{2m} - \phi^{2m+2}} \\
 Cov(x_m, x_n) &= \frac{\sigma^2 \phi^{m+n}(1 - \phi^{2m})}{\phi^{2m}(1 - \phi^2)} \\
 &= \sigma^2 \phi^{n-m} \frac{1 - \phi^{2m}}{1 - \phi^2}
 \end{aligned} \tag{A.23}$$

Si se define el paso de la serie como $h = (n - m)$ y m es lo suficientemente grande $m \rightarrow \infty$ por tanto $n \rightarrow \infty$; entonces:

$$Cov(x_m, x_n) = \frac{\sigma^2 \phi^h}{1 - \phi^2} \tag{A.24}$$

Finalmente, con la ecuación A.15y A.23, se calcula la correlación para este proceso.

$$\begin{aligned}
 \rho &= \frac{\frac{\sigma^2 \phi^{n-m}}{1 - \phi^2}}{\sqrt{\frac{\sigma^2(1 - \phi^{2n})}{1 - \phi^2}} \sqrt{\frac{\sigma^2(1 - \phi^{2m})}{1 - \phi^2}}} \\
 &= \frac{\sigma^2 \phi^{n-m}(1 - \phi^{2m})}{\sigma^2(1 - \phi^{2n})^{1/2}(1 - \phi^{2m})^{1/2}} \\
 &= \frac{\phi^{n-m}(1 - \phi^{2m})^{1/2}}{(1 - \phi^{2n})^{1/2}} \\
 \rho &= \phi^{n-m} \sqrt{\frac{1 - \phi^{2m}}{1 - \phi^{2n}}}
 \end{aligned} \tag{A.25}$$

Si m es lo suficientemente grande $m \rightarrow \infty$ y por consiguiente $n \rightarrow \infty$, ($m \leq n$); entonces la ecuación A.25 se transforma en la ecuación A.26.

$$\rho_x = \phi^{n-m} \tag{A.26}$$

APÉNDICE B

FRECUENCIA DE NYQUIST

La frecuencia de Nyquist es la frecuencia de corte asociada a una tasa de muestreo y corresponde a un medio de la frecuencia de muestreo. Cuando muestrea una señal continua, cualquier frecuencia mayor a la frecuencia de Nyquist sera irreconocible, generando un distorsión llamada “alias”.

De acuerdo con [48] un intervalo de muestreo Δ y su correspondiente frecuencia de muestreo $1/\Delta$ muestras por segundo. Una onda sinusoidal con frecuencia de $1/\Delta$ ciclos por segundo generado por el radio de la figura B.1, que rota en sentido antihorario, a una tasa de 1 revolución por muestra Δ , y esto conduce a que no pueda ser detectado cuando se muestrea a esta tasa. De la misma forma una onda sinusoidal con una frecuencia $-1/\Delta$ ciclos por segundo generado por el radio de la figura B.1, que rota en sentido horario a una tasa de 1 revolución por muestra Δ es también indistinguible. Ahora si se considera una onda sinusoidal con frecuencia en el intervalo $[-1/(2\Delta), 1/(2\Delta)]$; esta onda no se podrá identificar para cualquier frecuencia generada por el radio que complete alguna vuelta entera, ya sea horario o antihorario en el intervalo de muestra.

Esto quiere decir, que la frecuencia f , será indistinguible para,

$$f \pm k\Delta$$

donde k es un número entero. En la figura B.2 se muestra una función seno con una frecuencia de 1 Hz, $\sin(2\pi t)$ muestreada a una tasa de 0.2 segundos, el fenómeno de “alias” se presenta cuando $k = -1$. Esta frecuencia es $1 - 1/0.2$, que es igual a -4 Hz. Físicamente una frecuencia de -4 Hz es igual a 4 Hz siendo solamente diferente cuando la diferencia en la fase es de medio ciclo ($\sin(-\theta) = -\sin(\theta) = \sin(\theta - \pi)$).

Entonces, la frecuencia de Nyquist Q esta relacionada con el intervalo de muestreo Δ , de la forma,

$$Q = \frac{1}{2\Delta}$$

y Q debe ser mayor que cualquiera de las frecuencias que componen la señal original.

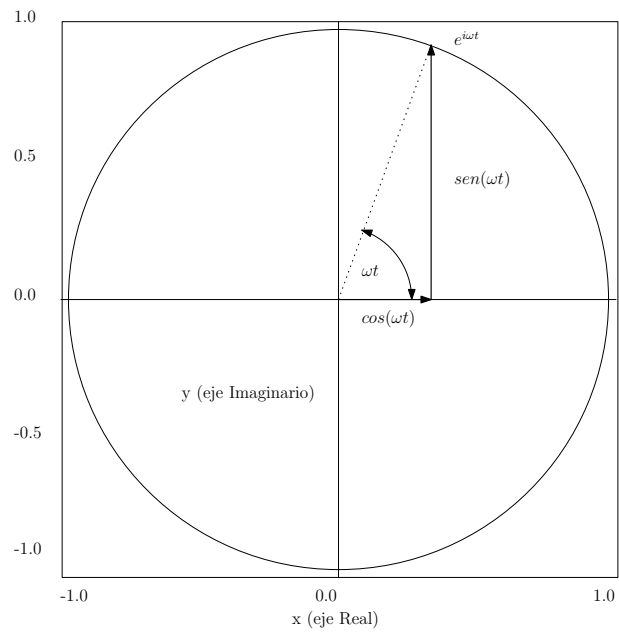


Figura B.1: En ángulo en radianes de longitud ωt . La proyección del radio sobre los ejes x y y es $\cos(\omega t)$ y $\text{sen}(\omega t)$ respectivamente. Tomado de [48]

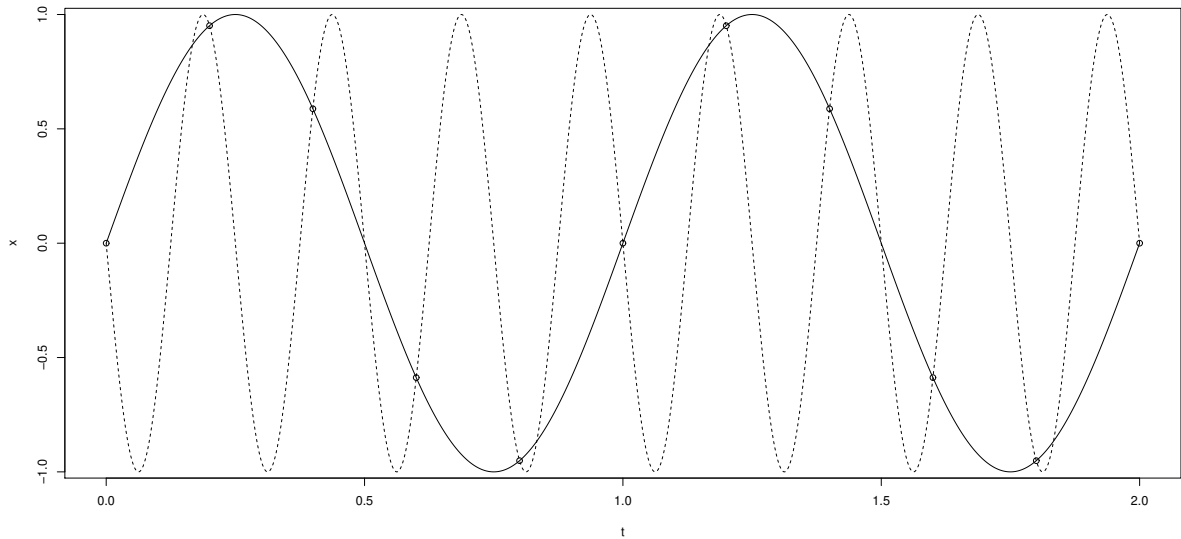


Figura B.2: Sobre posición de frecuencias: 1 Hz y 4 Hz con $\Delta = 0.2$ segundos Tomado de [48]

Ahora, si se supone que una serie de tiempo tiene una longitud n , el número de muestras a utilizar, T , es $n\Delta$. La frecuencia fundamental es $1/T$ Hz, y esta será la distancia entre cada uno de los picos en el espectro de Fourier. Si queremos distinguir las frecuencias separadas por ϵ Hz debemos tratar de tener estimados independientes centrados en esas frecuencias. Esto obliga al ancho de banda a ser al menos igual a ϵ . Si se toma una media móvil de L picos en el espectro de Fourier, tendremos la siguiente relación,

$$\frac{2L}{n\Delta} = \frac{2L}{T} \leq \epsilon$$

Por ejemplo, si se quiere distinguir frecuencias separadas por 1 Hz. Y $\Delta = 0.000001$. Tomando $L = 100$, el número de picos en el espectrograma, utilizando la ecuación anterior n deberá ser igual a $200 * 10^6$.

PRUEBAS ESTADÍSTICAS

C.1 BOX-PIERCE

Es un test para determinar si una serie de tiempo consiste en observaciones aleatorias (ruido blanco).

Estadístico de prueba

$$Q^* = N \sum_{k=1}^h r_k^2 \quad (\text{C.1})$$

Donde, N es el número de observaciones y h es el número máximo de corrimientos.

H_0 : Las observaciones se distribuyen de forma independiente (es decir, las correlaciones en la población son 0, cualquier correlación es producto de la aleatoriedad del proceso de muestreo).

H_1 : Las observaciones tienen correlación.

C.2 LJUNG-BOX

Se le conoce también como el estadístico modificado de Box-Pierce, se utiliza para determinar si una serie es o no un proceso de ruido blanco.

Estadístico de prueba

$$Q^* = N(N+2) \sum_{k=1}^h (N-k)^{-1} \rho_k^2 \quad (\text{C.2})$$

Donde, N es el número de observaciones, ρ es la auto correlación de la muestra en el retraso k y h es el numero máximo de corrimientos.

H_0 : Las observaciones se distribuyen de forma independiente (es decir, las correlaciones en la población son 0, cualquier correlación es producto de la aleatoriedad del proceso de muestreo).

H_1 : Las observaciones no son independientes.

Si la muestra es ruido blanco, Q^* tiene una distribución χ^2 con $(h - K)$ grados de libertad, donde K es el número de parámetros en el modelo.

C.3 DICKEY FULLER

Busca determinar la existencia o no de raíces unitarias en una serie de tiempo $[X_t]$.

Estadístico de prueba

$$\Delta X_i = \beta x_{i-1} + \epsilon_i$$

Donde, $\Delta X_i = X_i - x_{i-1}$ y $\beta = \phi - 1$

H_0 : Existe una raíz unitaria en la serie de tiempo (X_t) . Por lo tanto, $\beta = 0$ es equivalente a tener $\phi = 1$. En conclusión el proceso no es estacionario.

H_1 : El proceso es estacionario. Se tendrá $\beta < 0$ que es equivalente a tener $\phi < 1$.

BIBLIOGRAFÍA

- [1] Bloomberg 2022. «Day ahead electricity price in Mexico 01/01/2018 to 29/03/2022 - Retrieved from Bloomberg database». En: ().
- [2] Dimitri Afanasyev y Elena Fedorova. «On the impact of outlier filtering on the electricity price forecasting accuracy». En: *Applied Energy* (2019), págs. 196-210. DOI: <https://doi.org/10.1016/j.apenergy.2018.11.076>.
- [3] International Energy Agency. «OECD - Net electrical capacity». En: (2016). DOI: <https://doi.org/https://doi.org/10.1787/data-00460-en>. URL: <https://www.oecd-ilibrary.org/content/data/data-00460-en>.
- [4] The World Bank. *Access to electricity (% of population)*. "<https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?locations=MX&view=map>", Visitado el 06/11/2022. 2022.
- [5] Francisco Barnés-Regueiro, Matthew Leach y Matthias Ruth. «The Mexican energy sector: Integrated dynamic analysis of the natural gas/refining system». En: *Energy Policy* 30 (2002), págs. 767-779. ISSN: 03014215. DOI: 10.1016/S0301-4215(01)00137-9.
- [6] Anna Gloria Billé et al. «Forecasting electricity prices with expert, linear, and nonlinear models». En: *International Journal of Forecasting* (2022), In Press. DOI: <https://doi.org/10.1016/j.ijforecast.2022.01.003>.
- [7] Bloomberg. «Servicio Bloomberg Profesional». En: (2022). URL: <https://www.bloomberg.com/>.
- [8] Omar Castrejon-Campos, Lu Aye y Felix Kin Peng Hui. «Competition, coordination, or institutional change? A multi-perspective analysis of historical electricity transitions in Mexico». En: *Energy Research & Social Science* 84 (2022), pág. 102362. ISSN: 2214-6296. DOI: <https://doi.org/10.1016/j.erss.2021.102362>. URL: <https://www.sciencedirect.com/science/article/pii/S2214629621004527>.

- [9] CENACE. *¿Sabes qué es el Mercado Eléctrico Mayorista?* 2018. URL: <https://www.gob.mx/cenace/articulos/sabes-que-es-el-mercado-electrico-mayorista?idiom=es>.
- [10] CENACE. «¿Sabes qué es el Sistema Eléctrico Nacional?» En: (2018). URL: <https://www.gob.mx/cenace/articulos/sabes-que-es-el-sistema-electrico-nacional?idiom=es>.
- [11] Joshua Chan y Angelia Grant. «Modeling energy price dynamics: GARCH versus stochastic volatility». En: *Energy Economics* (2016), págs. 18-189. DOI: <http://dx.doi.org/10.1016/j.eneco.2015.12.003>.
- [12] Atilla Cifter. «Forecasting electricity price volatility with the Markov-switching GARCH model: Evidence from the Nordic electric power market». En: *Electric Power Systems Research* (2013). DOI: <http://dx.doi.org/10.1016/j.epsr.2013.04.007>.
- [13] Antonio J Conejo et al. «Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models». En: *Transaction on Power Systems* (2005). DOI: [doi:10.1109/TPWRS.2005.846054](https://doi.org/10.1109/TPWRS.2005.846054).
- [14] Javier Contreras et al. «ARIMA model to predict Next-Day electricity prices». En: *IEEE Explore* (2003).
- [15] E Cruz May et al. «An empirical comparison of alternate regime-switching models for electricity spot prices». En: *International Journal of Electrical Power and Energy Systems* (2022). DOI: <https://doi.org/10.1016/j.ijepes.2021.107505>.
- [16] ENEL. *Conoce el Mercado Eléctrico Mayorista*. 2022. URL: <https://www.enel.mx/es/productos/enel-energia/conoce-mem>.
- [17] Bruno Escotto. «¿Qué es Bloomberg?» En: (2017). URL: <https://www.integrity-la.com/que-es-bloomberg>.
- [18] Nicolas Fulghum. «Global Electricity Review 2021: Mexico». En: (2021). URL: <https://ember-climate.org/app/uploads/2022/02/Global-Electricity-Review-2021-Mexico-Translated.pdf>.
- [19] Kwoklun Lo Gao y Fulin Fan. «Comparison of ARIMA and ANN Models Used in Electricity Price Forecasting for Power Market». En: *Energy and Power Engineering* (2017), págs. 120-126. DOI: <https://doi.org/10.4236/epe.2017.94B015>.
- [20] Reinaldo García et al. «A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices». En: *IEEE Transactions on Power Systems* (2005). DOI: [10.1109/TPWRS.2005.846044](https://doi.org/10.1109/TPWRS.2005.846044).

- [21] John Geweke y Susan Porter-Hudak. «the Estimation and Application of Long Memory Time Series Models». En: *Journal of Time Series Analysis* 4 (1983), págs. 221-238. ISSN: 14679892. DOI: 10.1111/j.1467-9892.1983.tb00371.x. URL: [http://dx.doi.org/10.1016/S0378-4371\(02\)00961-5](http://dx.doi.org/10.1016/S0378-4371(02)00961-5).
- [22] G.P Girish. «Spot electricity price forecasting in Indian electricity market using autoregressive-GARCH models». En: *Energy Strategy Reviews* (2016), págs. 52-57. DOI: <http://dx.doi.org/10.1016/j.esr.2016.06.005>.
- [23] González-López et al. «Integrated analysis of the Mexican electricity sector: Changes during the Covid-19 pandemic». En: *The Electricity Journal* (2022). DOI: <https://doi.org/10.1016/j.tej.2022.107142>.
- [24] Azize Hayfavi e Irem Talasli. «Stochastic multifactor modeling of spot electricity prices». En: *Journal of Computational and Applied Mathematics* 259 (2014), págs. 434-442. ISSN: 0377-0427. DOI: <https://doi.org/10.1016/j.cam.2013.10.008>. URL: <https://www.sciencedirect.com/science/article/pii/S0377042713005529>.
- [25] Emily Hickey, David Loomis y Mohammadi Hassan. «Forecasting hourly electricity prices using ARMAX-GARCH models: An application to MISO hubs». En: *Energy Economics* (2012), págs. 307-315. DOI: [doi:10.1016/j.eneco.2011.11.011](https://doi.org/10.1016/j.eneco.2011.11.011).
- [26] Ronald Huisman, Christian Huurman y Ronald Mahieu. «Hourly electricity prices in day-ahead markets». En: *Energy Economics* (2007), págs. 240-248. DOI: [doi:10.1016/j.eneco.2006.08.005](https://doi.org/10.1016/j.eneco.2006.08.005).
- [27] Rob J Hyndman. *Detecting time series outliers*. 2021. URL: <https://robjhyndman.com/hyndsight/tsoutliers/>.
- [28] Law Insider. *Mercado del Día en Adelanto definition*. 2014. URL: <https://www.lawinsider.com/es/dictionary/mercado-del-dia-en-adelanto>.
- [29] Joanna Janczura y Rafal Weron. «An empirical comparison of alternate regime-switching models for electricity spot prices». En: *Energy Economics* (2010), págs. 1059-1073. DOI: [doi:10.1016/j.eneco.2010.05.008](https://doi.org/10.1016/j.eneco.2010.05.008).
- [30] Marco A. Jano-Ito y Douglas Crawford-Brown. «Socio-technical analysis of the electricity sector of Mexico: Its historical evolution and implications for a transition towards low-carbon development». En: *Renewable and Sustainable Energy Reviews* 55 (2016), págs. 567-590. ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2015.10.153>. URL: <https://www.sciencedirect.com/science/article/pii/S1364032115012320>.
- [31] Ladislav Kristoufek y Petra Lunackova. «Long-term memory in electricity prices: Czech market evidence». En: (2013).

- [32] Joao Lagarto et al. «Price forecasting in the day-ahead Iberian electricity market using a conjectural variations ARIMA model». En: *IEEE Explore* (2012).
- [33] Jesús Lago et al. «Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark». En: *Applied Energy* (2021). DOI: <https://doi.org/10.1016/j.apenergy.2021.116983>.
- [34] Lifeder. *Transformada de Fourier: propiedades, aplicaciones, ejemplos*. 2023. URL: <https://www.lifeder.com/transformada-de-fourier/>.
- [35] Francesco Lisi y Fany Nan. «Component estimation for electricity prices: Procedures and comparisons». En: *Energy Economics* (2014), págs. 143-159. DOI: <https://doi.org/10.1016/j.eneco.2018.05.027>.
- [36] Francesco Lisi y Matteo M. Pelagatti. «Component estimation for electricity market data: Deterministic or stochastic?». En: *Energy Economics* (2018), págs. 13-37. DOI: <https://doi.org/10.1016/j.eneco.2018.05.027>.
- [37] Benitez Luis. *¿Qué es un modelo autorregresivo?* 2022. URL: <https://statologos.com/modelo-autorregresivo/>.
- [38] Catherine McHugh, Sonya Coleman y Dermot Kerr. «Hourly electricity price forecasting with NARMAX». En: *Machine Learning with Applications* (2022). DOI: <https://doi.org/10.1016/j.mlwa.2022.100383>.
- [39] Ministerio de Medio Ambiente de la República Argentina. *Tasa media anual de crecimiento poblacional*. 2016. URL: http://estadisticas.ambiente.gob.ar/archivos/web/Indicadores/file/multisitio/pdf/13_%5C%20Tasa%5C%20media%5C%20anual%5C%20de%5C%20crecimiento%5C%20poblacional.pdf.
- [40] Industria y Turismo de Colombia Ministro de Comercio. *Guía de contratación pública para micro y pequeñas empresas - MIPYME*. 2021. URL: <http://www.aplicaciones-mcit.gov.co/guiapymes/c8i9.html#:~:text=9.2.2.1.1%5C%20Subasta%5C%20inversa,precios%5C%20durante%5C%20un%5C%20tiempo%5C%20determinado..>
- [41] Francisco Nogales et al. «Forecasting Next-Day Electricity Prices by Time Series Models». En: *IEEE Transactions on Power Systems* (2002).
- [42] Jakub Nowotarski, Jakub Tomczyk y Rafal Weron. «Robust estimation and forecasting of the long-term seasonal component of electricity spot prices». En: *Energy Economics* (2013), págs. 13-27. DOI: <http://dx.doi.org/10.1016/j.eneco.2013.04.004>.
- [43] Jakub Nowotarski y Rafal Weron. «On the importance of the long-term seasonal component in day-ahead electricity price forecasting». En: *Energy Economics* (2016), págs. 228-235. DOI: <http://dx.doi.org/10.1016/j.eneco.2016.05.009>.

- [44] Jakub Nowotarski y Rafal Weron. «Recent advances in electricity prices forecasting: A review of probabilistic forecasting». En: *Renewable and Sustainable Energy Reviews* (2018), págs. 1548-1568. DOI: <http://dx.doi.org/10.1016/j.rser.2017.05.234>.
- [45] Jakub Nowotarski et al. «An empirical comparison of alternative schemes for combining electricity spot price forecasts». En: *Energy Economics* (2014), págs. 395-412. DOI: <http://dx.doi.org/10.1016/j.eneco.2014.07.014>.
- [46] Open.Ed. *Applied Time Series Analysis*. 2021. URL: <https://online.stat.psu.edu/stat510/lesson/4/4.1>.
- [47] R. Ortega Lamelín. *La evolución constitucional de la energía a partir de 1917. Colección INEHRM*. Universidad Nacional Autónoma de México - Instituto de Investigaciones Jurídicas, 2016. ISBN: 9786078507559. URL: <https://archivos.juridicas.unam.mx/www/bjv/libros/11/5275/9.pdf>.
- [48] Andrew V. Metcalfe Paul S.P. Cowpertwait. *Introductory Time Series with R*. Springer, 2009. ISBN: 9780387886985.
- [49] Steven Jorge Pedrosa. *Heterocedasticidad*. 2020. URL: <https://economipedia.com/definiciones/heterocedasticidad.html>.
- [50] Oscar Quintero y Jonathan Ruiz. *Estimación del componente de Hurst y la dimensión fractal de una superficie topográfica a través de la extracción de perfiles*. 2011. URL: <https://core.ac.uk/download/pdf/229157368.pdf>.
- [51] Preethi Rajan y Vijaya Chandrakala. «Statistical Model Approach of Electricity Price Forecasting for Indian Electricity Market». En: *IEEE Madras Section Conference (MASCON)* (2021). DOI: doi:10.1109/MASCON51689.2021.9563474.
- [52] Adriana María Ramírez-Camperos, Víctor Rodríguez-Padilla y Pedro Antonio Guido-Aldana. «the Mexican electricity sector: Policy analysis and reform (1992-2009)». En: *Energy Policy* 62 (2013), págs. 1092-1103. ISSN: 03014215. DOI: 10.1016/j.enpol.2013.06.063. URL: <http://dx.doi.org/10.1016/j.enpol.2013.06.063>.
- [53] M. de la P. Ramos Lara. *La compañía Mexican Light and Power Company Limited durante la Revolución Mexican*. Universidad de Barcelona, 2017. ISBN: 9788469768358. URL: <http://www.ub.edu/geocrit/Electr-y-territorio/PazRamos.pdf>.
- [54] RICEnergy. *Qué es y cómo funciona el Mercado Eléctrico Mayorista*. URL: <https://www.ric.mx/cultura/energia/mexico/que-es-y-como-funciona-el-mercado-electrico-mayorista/>.

- [55] Hyndman Rob y Athanasopoulos George. *Forecasting: Principles and Practice*. Texts, 2021. ISBN: 0987507133. URL: <https://otexts.com/fpp3/>.
- [56] Paula Rodó. *Modelo autorregresivo (AR)*. 2019. URL: <https://economipedia.com/definiciones/modelo-autorregresivo-ar.html>.
- [57] Daniela Rodríguez. «Investigación Aplicada: características, definición, ejemplos». En: (2020). URL: <https://www.lifeder.com/investigacion-aplicada/>.
- [58] Alejandro Rodríguez Puerta. «Investigación cuantitativa: características, técnicas, ejemplos». En: (2020). URL: <https://www.lifeder.com/investigacion-cuantitativa/>.
- [59] SENER. *Programa de desarrollo del Sistema Eléctrico Nacional 2020-2034*. 2020. URL: [https://base.energia.gob.mx/dgaic/DA/P/SubsecretariaElectricidad/ConjuntosProyectosInversion/SENER_07_ProgramaDesarrolloSistemaElectricoNacional2020-2034\(PRODESEN\).pdf](https://base.energia.gob.mx/dgaic/DA/P/SubsecretariaElectricidad/ConjuntosProyectosInversion/SENER_07_ProgramaDesarrolloSistemaElectricoNacional2020-2034(PRODESEN).pdf).
- [60] Significados. *Qué es el Ruido*. 2023. URL: <https://www.significados.com/ruido/>.
- [61] National Institute of Standards et al. *Engineering Statistics Handbook*. 2001.
- [62] Daniel A. Steck. *Rubidium 85 D Line Data*. available online at <http://steck.us/alkalidata>. (revision 2.2.1, 21 November 2019).
- [63] Rick Steinert y Florian Ziel. «Short- to Mid-term Day-Ahead Electricity Price Forecasting Using Futures». En: *The Energy Journal* (2019), págs. 105-127. DOI: <https://doi.org/10.5547/01956574.40.1.rste>.
- [64] S Vijayalakshmi et al. «Is Indian Spot Electricity Price Stationary?» En: *Energy Procedia* (2017), págs. 1067-1072. DOI: 10.1016/j.egypro.2017.10.114.
- [65] Rafal Weron. «Electricity price forecasting: A review of the state-of-the-art with a look into the future». En: *International Journal of Forecasting* (2014), págs. 1030-1081. DOI: <http://dx.doi.org/10.1016/j.ijforecast.2014.08.008>.
- [66] Rafał Weron. «Estimating long-range dependence: finite sample properties and confidence intervals». En: *Physica A: Statistical Mechanics and its Applications* 312 (2001), págs. 285-299. ISSN: 03784371. DOI: 10.1016/S0378-4371(02)00961-5. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0378437102009615>.
- [67] Guillermo Westreicher. «Muestreo no probabilístico». En: (2022). URL: <https://economipedia.com/definiciones/muestreo-no-probabilistico.html>.
- [68] Mengchen Xie, Claes Zhu Sandels y Lars Kun Nordstrom. «A seasonal ARIMA model with exogenous variables for elspot Electricity prices in Sweden». En: *IEEE Explore* (2013).

-
- [69] Jorge Zarco. *Entendiendo el Mercado Eléctrico: ¿Cómo funciona?* 2019. URL: <https://www.pv-magazine-mexico.com/2019/07/23/entendiendo-el-mercado-electrico-como-funciona/#:~:text=El%5C%20Mercado%5C%20de%5C%20Balance%5C%20de%5C%20Potencia%5C%20es%5C%20un, los%5C%20periodos%5C%20de%5C%20mayor%5C%20demanda%5C%20en%5C%20el%5C%20a%5C%C3%5C%B1o..>
- [70] Tan Zhongfu et al. «Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models». En: *Applied Energy* (2010), págs. 3606-3610. DOI: doi:10.1016/j.apenergy.2010.05.012.

ÍNDICE DE FIGURAS

2.1. Regiones del Sistema Eléctrico Nacional (SEN) ([59])	28
2.2. Sistema de suministro eléctrico - Elaborado por SENER con información de CENACE ([59])	29
2.3. Capacidad de las interconexiones internacionales en 2019 del SEN - Elaborado por SENER con información de CENACE ([59])	30
2.4. Principales Centrales Eléctricas de la CFE y PIE al 31 de diciembre de 2019 ([59])	30
2.5. Principales Centrales Eléctricas privadas al 31 de diciembre de 2019 ([59]) .	32
2.6. Pronóstico del Consumo Bruto del SEN 2020 - 2034 en los escenarios de Planeación, Alto y Bajo - Elaborado por SENER con información de CENACE ([59])	33
2.7. Pronóstico del Consumo Bruto del SEN 2020 - 2034, contingencia sanitaria, en los escenarios de Planeación, Alto y Bajo - Elaborado por SENER con información de CENACE ([59])	34
2.8. Pronóstico de la Demanda Máxima Integrada del SEN 2020 - 2034, escenarios de Planeación, Alto y Bajo - Elaborado por SENER con información de CENACE ([59])	35
2.9. Mercados que componen el MEM ([9])	36
2.10. Mecanismo de subasta del Día de Adelanto: los participantes del MEM (compradores y vendedores) entregan sus necesidades de energía y sus ofertas antes del cierre en el día $d-1$ para la entrega de electricidad en el día d ([33])	37
2.11. Diferentes medias móviles aplicadas a los precios de le energía en la región Centro para el primer trimestre de 2018	41

2.12. Correlograma de los precios de la energía diarios a las 7 AM en el periodo 2018 al 2022.	44
2.13. Función definida por partes con periodo 10.	60
3.1. Región Valle de México Centro y Monterrey como una serie de tiempo . . .	66
3.2. Región Guadalajara - Querétaro como una serie de tiempo	67
3.3. Región Valle de México Sur y Puebla como una serie de tiempo	68
3.4. Región Mérida y Mexicali como una serie de tiempo	69
3.5. Diagrama de cajas para los datos analizados. Los puntos rojos indican los valores atípicos.	72
3.6. Región Valle de México Centro y Monterrey como serie de tiempo con tendencia y sin datos atípicos	74
3.7. Región Guadalajara y Querétaro como serie de tiempo con tendencia y sin datos atípicos	74
3.8. Región Valle de México Sur y Puebla como serie de tiempo con tendencia y sin datos atípicos	75
3.9. Región Valle de Mérida y Mexicali como serie de tiempo con tendencia y sin datos atípicos	75
3.10. Diagrama de cajas para los datos analizados eliminando valores extremos. Los puntos rojos indican los valores atípicos.	77
3.11. Valle de México Centro con valores atípicos y datos faltantes	78
3.12. Monterrey con valores atípicos y datos faltantes	79
3.13. Guadalajara con valores atípicos y datos faltantes	79
3.14. Queretaro con valores atípicos y datos faltantes	80
3.15. Puebla con valores atípicos y datos faltantes	80
3.16. Valle de México Sur con valores atípicos y datos faltantes	81
3.17. Mérida con valores atípicos y datos faltantes	81
3.18. Mexicali con valores atípicos y datos faltantes	82

3.19. Función de Autocorrelación para las series de tiempo. Se observa autocorrelación significativa en los datos con un lag de 1 semana	82
3.20. Función de Autocorrelación para las series de tiempo. Se observa autocorrelación significativa en los datos con un lag de 1 semana	83
3.22. Gráfico de correlación con histogramas, funciones de densidad, líneas de regresión y coeficientes de correlación con nivel de significancia del 1% para los precios horarios de la energía eléctrica en las 8 regiones del SIN . .	88
3.23. Gráfico de correlación con histogramas, funciones de densidad, líneas de regresión y coeficientes de correlación con nivel de significancia del 1% para los precios horarios de la energía eléctrica en las 8 regiones del SIN, sin datos atípicos	89
3.24. Gráfica de Autocorrelación por horas para el nodo VDM Centro y a diferentes retrasos	91
3.25. Gráfica de Autocorrelación por horas para el nodo Monterrey y a diferentes retrasos	92
3.26. Gráfica de Autocorrelación por horas para el nodo Guadalajara y a diferentes retrasos	93
3.27. Gráfica de Autocorrelación por horas para el nodo Mérida y a diferentes retrasos	94
3.28. Gráfica de Autocorrelación por horas para el nodo Mexicali y a diferentes retrasos	95
3.29. Descomposición por medio de ETS	97
3.30. Predicción a una semana	100
3.31. Espectrograma de frecuencias para Valle de México Centro	102
3.32. Espectrograma de frecuencias para Monterrey	103
3.33. Espectrograma de frecuencias para Guadalajara	104
3.34. Espectrograma de frecuencias para Mérida	105
3.35. Espectrograma de frecuencias para Mexicali	106
3.36. Distribución de precio por hora del día región Valle de México Centro . . .	108

3.37. Distribución de precio por día de la semana región Valle de México Centro	109
3.38. Distribución de precio por semana del mes región Valle de México Centro .	109
3.39. Distribución de precio por mes del año región Valle de México Centro	110
3.40. Distribución de precio por hora del día región Monterrey	110
3.41. Distribución de precio por día de la semana región Monterrey	111
3.42. Distribución de precio por semana del mes región Monterrey	111
3.43. Distribución de precio por mes del año región Monterrey	112
3.44. Distribución de precio por hora del día región Guadalajara	112
3.45. Distribución de precio por día de la semana región Guadalajara	113
3.46. Distribución de precio por semana del mes región Guadalajara	113
3.47. Distribución de precio por mes del año región Guadalajara	114
3.48. Distribución de precio por hora del día región Mérida	114
3.49. Distribución de precio por día de la semana región Mérida	115
3.50. Distribución de precio por semana del mes región Mérida	115
3.51. Distribución de precio por mes del año región Mérida	116
3.52. Distribución de precio por hora del día región Mexicali	116
3.53. Distribución de precio por día de la semana región Mexicali	117
3.54. Distribución de precio por semana del mes región Mexicali	117
3.55. Distribución de precio por mes del año región Mexicali	118
3.56. Cronograma de investigación	120
4.1. Predicción con una periodicidad.	121
4.2. Predicción con múltiples periodicidades.	122
4.3. Modelo Región VDM Centro	133
4.4. Modelo Región Monterrey	134
4.5. Modelo Región Guadalajara	135

4.6. Modelo Región Merida	136
4.7. Modelo Región Mexicali	137
B.1. En ángulo en radianes de longitud ωt . La proyección del radio sobre los ejes x y y es $\cos(\omega t)$ y $\sin(\omega t)$ respectivamente. Tomado de [48]	151
B.2. Sobre posición de frecuencias: 1 Hz y 4 Hz con $\Delta = 0.2$ segundos Tomado de [48]	151

ÍNDICE DE TABLAS

2.1. Modelos multiagente	8
2.2. Métodos fundamentales	9
2.3. Modelos de forma reducida	10
2.4. Modelos de forma reducida	11
2.5. Técnicas de inteligencia computacional (Inteligencia artificial y estadística no lineal)	12
2.6. Revisión bibliográfica: Pronóstico de precios de la energía mediante modelos autorregresivos	13
2.7. Capacidad instalada de la CFE y del resto de permisionarios (MW) - Elaboración de la SENER con información de CENACE ([59])	31
3.1. Resumen de la base de datos utilizada	70
3.2. Fuerza de la periodicidad por región	71
4.1. Coeficiente de Hurst en diferentes regiones de México	124
4.2. Coeficientes	130
4.3. Medidas de Error en el Conjunto de Entrenamiento (Parte 1)	130
4.4. Medidas de Error en el Conjunto de Entrenamiento (Parte 2)	130
4.5. Coeficientes	131
4.6. Medidas de Error en el Conjunto de Entrenamiento (Parte 1)	132
4.7. Medidas de Error en el Conjunto de Entrenamiento (Parte 2)	132

GLOSARIO

Caminata Aleatoria es un buen modelo para los datos estocásticos que tienen alguna tendencia.. 47,

Coefficiente de Hurst 'El coeficiente o exponente de Hurst es una medida de independencia de las series de tiempo como elemento para distinguir series fractales. Hurst descubrió que muchos fenómenos naturales exhiben un comportamiento que puede ser caracterizado por un proceso aleatorio sesgado, en el cual existe "memoria de largo plazo" entre las observaciones, es decir, que los eventos de un periodo influyen en todos los siguientes"[50]. 38,

Heterocedasticidad en estadística, cuando los errores no son constantes a lo largo de toda la muestra. En otras palabras, en los modelos de regresión lineales se dice que hay heterocedasticidad cuando la varianza de los errores no es igual en todas las observaciones realizadas. Así, no se cumple uno de los requisitos básicos de las hipótesis de los modelos lineales. [49]. 11,

Mercado de Balance de Potencia 'es un mercado anual que tiene como objetivo reflejar la escasez o exceso de capacidad disponible de generación a través de señales de precios con el fin de que el Sistema Eléctrico Nacional (SEN) vaya adquiriendo la capacidad disponible para suministrar energía en los periodos de mayor demanda en el año."[69]. 37,

Mercado de Día de Adelanto 'significa el mercado de antelación cuyos participantes podrán presentar ofertas horarias de venta de energía y servicios conexos, así como las ofertas horarias de compra de energía, las cuales resultarán en compromisos financieramente vinculantes para la entrega o recepción de energía y servicios conexos en el día siguiente a la realización del Mercado del Día en Adelanto."[28]. 37,

Modelo autorregresivo 'Un modelo autorregresivo (AR) predice el comportamiento futuro en función del comportamiento pasado . Se utiliza para pronosticar cuando existe alguna correlación entre los valores de una serie temporal y los valores que los preceden y suceden."[37]. 3,

Modelos Autoregresivos 'Los modelos de autorregresión, también conocidos como modelos AR, se utilizan para realizar pronósticos sobre variables ex-post (observaciones que conocemos completamente su valor) en determinados momentos del tiempo normalmente ordenados cronológicamente."[56]. 48,

Ruido Blanco 'El ruido blanco o sonido blanco es un tipo de señal de carácter aleatorio y que no presenta correlación estadística entre sus valores en dos tiempos diferentes. Este tipo de ruido presenta todas las frecuencias y su potencia es constante (la potencia de densidad del espectro es plana)."[60]. 46,

Serie de tiempo Una serie de tiempo es una secuencia de datos ordenados, generalmente este ordenamiento se da en el tiempo se hace para entender el pasado y predecir el futuro. [48]. 38,

Subasta Inversa 'es el mecanismo por el cual los interesados en ofrecer sus bienes y servicios realizan una puja dinámica, mediante la reducción sucesiva de sus precios durante un tiempo determinado".[40]; esto, aplicado al sistema eléctrico mexicano, significa que los generadores de energía eléctrica envían sus ofertas del costo del MWh y el CENACE clasifica por precio y despacha del más económico al más costoso. 2,

Tasa media de crecimiento anual '(tcma) Es el ritmo al que la población aumenta o disminuye durante un período dado, debido al efecto de la natalidad, la mortalidad y las migraciones."[39]. 33,

Transformada de Fourier 'La transformada de Fourier es un método de adecuación analítica orientado a funciones integrables que pertenece a la familia de las transformadas integrales."[34]. 38,

SIGLAS

- AR** Autorregressive (Autorregresivo). 11,
- ARIMA** Autoregressive Integrated Moving Average (Autorregresivo Integrado de Media Móvil). 11,
- ARMA** (Autoregressive Moving Average (Autorregresivo de Media Móvil). 11,
- ARX** Autoregressive with Extra Input (Autorregresivo con variables exógenas). 11,
- CEL** Certificados de Energía Limpia. 37,
- CENACE** Centro Nacional de Control de Energía. 2,
- CFE** Comisión Federal de Electricidad. 21,
- CNDSP** Confederación Nacional de los Servicios Públicos. 21,
- CPP** Captive Power Plant (Co-generador). 26,
- CRE** Comisión Reguladora de Energía. 25,
- DFA** Detrended Fluctuation Analysis (análisis de fluctuación sin tendencia). 53,
- DOF** Diario Oficial de la Federación. 20, 21,
- DTF** Derechos Financieros de Transmisión. 38,
- ETS** Error, Trend, Seasonality (Error, Tendencia y Estacionalidad). 96,
- FFT** Fast Fourier Transform (Transformada Rápida de Fourier). 101,
- GATT** General Agreement on Tariffs and Trade (Acuerdo General sobre Aranceles Aduaneros y Comercio). 24,
- IPP** Independent Power Producers (Generadores Independientes Privados). 24,

- LGCC** Ley General de Cambio Climático. 26,
- LyFC** Compañía de Luz y Fuerza del Centro. 23,
- MDA** Mercado de Día de Adelanto. 2,
- MEM** Mercado Eléctrico Mayorista. 1,
- MLPC** Mexican Light and Power Company. 19,
- NAFTA** North American Free Trade Agreement (Tratado de libre comercio de Norte América. 24,
- OECD** Organisation for Economic Cooperation and Development (Organización para la Cooperación y el Desarrollo Económicos). 25,
- PEMEX** Petroleos Mexicanos. 27,
- PIB** Producto Interno Bruto. 26,
- PIE** Productor Independiente de Energía Eléctrica. 29,
- PML** Precio Marginal Local. 4,
- PPA** Power Purchase Agreement (Acuerdos de Compra de Potencia). 24,
- PRI** Partido Revolucionario Institucional. 25,
- PSS** Power Self Supply (Autogenerador). 26,
- RDG** Red de Distribución General. 28,
- RMSE** Root Mean Square Error (Error de raíz cuadrada media). 4,
- RNT** Red Nacional de Transmisión. 28,
- SARIMA** Seasonal Autoregressive Integrated Moving Average (Auto Regresivo Estacional de Media Móvil. 11,
- SE** Secretaria de Energía. 23,
- SEN** Sistema Eléctrico Nacional. 27,
- SENER** La Secretaria de Energía. 27,
- SIN** Sistema Interconectado Nacional. 28,

SME Sindicato Mexicano de Electricistas. 19,

SPP Small Power Producer (Generadores Pequeños de Potencia). 24,

tmca Tasa Media de Crecimiento Anual. 33,

