



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE CIUDAD MADERO

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

MAESTRÍA EN CIENCIAS EN INGENIERÍA



"POR MI PATRIA Y POR MI BIEN"

TESIS

MÉTODOS DE PREDICCIÓN DE LA TEMPERATURA COMO VARIABLE RELEVANTE EN EL CAMBIO CLIMÁTICO EN REGIONES DE MÉXICO.

Que para obtener el Grado de
Maestro en Ciencias en Ingeniería

Presenta

Ing. Rebeca Daniela Yáñez Díaz

G16071024

CVU 1169360

Director de Tesis

Dr. Juan Frausto Solís

CVU 31308

Co-director de Tesis

Juan Javier González Barbosa

Cd. Madero, Tamaulipas

abril 2024

Cd. Madero, Tam., a 14 de Diciembre de 2023

Dr. Marco Antonio Coronel García

Jefe de la División de Estudios de Posgrado e Investigación
Instituto Tecnológico de Ciudad Madero

Dr. Nelson Rangel Valdez

Coordinador del Programa en Ciencias de la Ingeniería
Instituto Tecnológico de Ciudad Madero

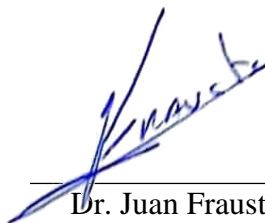
Presente

Por este conducto nos dirigimos a usted para comunicarle que, habiendo revisado el trabajo de investigación presentado por el **C. Ing. Rebeca Daniela Yáñez Díaz**, alumna de la Maestría en Ciencias de la Ingeniería, titulado:

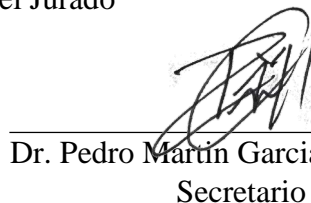
**“MÉTODOS DE PREDICCIÓN DE LA TEMPERATURA COMO
VARIABLE RELEVANTE EN EL CAMBIO CLIMÁTICO EN
REGIONES DE MÉXICO.”**

así mismo considerando que se han realizado todas las correcciones a las observaciones señaladas en dicho trabajo, nos permitimos solicitar la autorización para la impresión final de la tesis.

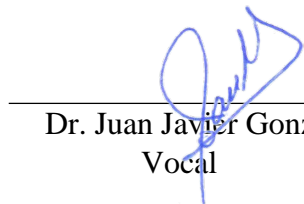
ATENTAMENTE
Los integrantes del Jurado



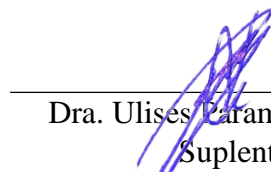
Dr. Juan Frausto Solís
Presidente



Dr. Pedro Martín García Vite
Secretario



Dr. Juan Javier González Barbosa
Vocal



Dra. Ulises Paramo Garcia
Suplente

**DECLARACIONES DE ORIGINALIDAD, PROPIEDAD
INTELECTUAL, CESION DE DERECHOS Y/O
CONFIDENCIALIDAD**

Por medio de la presente, yo Rebeca Daniela Yáñez Díaz declaro que la tesis titulada Métodos de Predicción de la Temperatura como Variable Relevante en el Cambio Climático en Regiones de México es un trabajo original e inédito de mi autoría, que no ha sido publicado ni presentado en otro lugar para obtener algún grado académico o reconocimiento.

Cedo los derechos patrimoniales de autor de mi tesis al Tecnológico Nacional de México.

Atentamente



Ing. Rebeca Daniela Yáñez Díaz

AGRADECIMIENTOS

Agradezco ampliamente a la Secretaría de Educación Pública, al Tecnológico Nacional de México, al Instituto Tecnológico de Ciudad Madero y al CONACYT por su apoyo con una beca que me permite continuar con mi formación académica, a los profesores que me apoyaron para la escritura de esta tesis, por su amplio conocimiento y experiencia, a los doctores Juan Frausto Solís y Juan Javier González Barbosa, directores de esta tesis.

Métodos de Predicción de la Temperatura como Variable Relevante en el Cambio Climático en Regiones de México.

Ing. Rebeca Daniela Yáñez Díaz

Resumen

Se estima que la temperatura es la variable relevante para el Cambio Climático (CC), que puede aumentar más de 1.5 °C entre 2030 y 2052. Se han obtenido los mejores resultados de pronósticos con suavizamiento exponencial (SE) y sus combinaciones en la competencia M4. Sin embargo, para CC solo ha sido fácil y eficaz para corto plazo. Por ello, es importante realizar una predicción a mediano plazo de la temperatura en una ciudad pequeña con registros históricos de varias décadas. Este trabajo preparó información de 30 años para Tampico, Tamaulipas y Xalapa, Veracruz ciudades costeras en crecimiento. Esto ayudaría a tomar acciones concretas contra este grave problema. Este trabajo comprueba que el mejor método para la ciudad de Xalapa es el Suavizamiento Exponencial Triple Aditivo; la implementación adoptada obtuvo un excelente resultado con una métrica MAPE de 2.96% de estimación mensual difícil de superar incluso con el método clásico denominado Autorregresivo Integrado de Media Móvil Estacional. Para la ciudad de Tampico también se obtuvo un buen resultado con el modelo con una métrica de 3.7% y también se utilizó SARIMA con una métrica de 3.64%. Esta diferencia de error es muy poca y sigue demostrando que SET es muy buen modelo. Se hizo también una estimación de la temperatura en ambas ciudades hacia los años 2024 a 2027 y se hace una comparación de estos contra el promedio de temperatura cercanos a los años 2000 de ambas ciudades. Finalmente, en ambas se puede notar un incremento de la temperatura promedio conforme pasa el tiempo.

Temperature Prediction Methods as a Relevant Variable in Climate Change in Regions of Mexico.

Ing. Rebeca Daniela Yáñez Díaz

Abstract

Temperature is estimated to be the relevant variable for Climate Change (CC), which may increase by more than 1.5 °C between 2030 and 2052. The best forecasting results have been obtained with exponential smoothing (ES) and its combinations in the M4 competition. However, for CC, forecasting for the short term has only been effortless and effective. Therefore, performing a medium-term temperature prediction in a small city with historical records of several decades is crucial. This work prepared 30-year data for Tampico, Tamaulipas, and Xalapa, Veracruz growing coastal cities. This work would help to take several concrete actions against this serious problem. The present research proves that the best method for the city of Xalapa is the Triple Additive Exponential Smoothing; the adopted implementation obtained an excellent result with a MAPE metric of 2.96% monthly estimation, which is difficult to overcome even with the classical method called Seasonal Moving Average Integrated Autoregressive. For the city of Tampico, a good result was obtained with the model, where this metric achieved 3.7%, and SARIMA achieved 3.64%. This tiny difference shows that SET is an excellent model. An estimation of the temperature in both cities for 2024 to 2027 was also made, and a comparison was made with the average temperature near 2000 for both cities. Finally, in both cities, an increase in the average temperature can be noted as time goes by.

Índice General

Resumen.....	V
Abstract.....	VI
Índice tablas.....	IX
Índice de figuras	X
1 Introducción.....	1
1.1 Planteamiento del problema.....	1
1.2 Objetivo General.....	4
1.3 Objetivos específicos	4
1.4 Hipótesis	5
1.5 Justificación del estudio.....	5
1.6 Alcances.....	6
1.7 Limitaciones.....	6
1.8 Organización de la tesis.	6
2 Marco teórico	7
2.1.1 Gases de efecto invernadero.....	7
2.1.2 Contaminantes atmosféricos y sus fuentes	8
2.1.3 Temperatura.....	8
2.1.4 Predicción de cambio climático.....	8
2.1.5 Notación en pronósticos	9
2.1.6 Autocorrelación	9
2.1.7 Suavizamiento exponencial	10
2.1.8 Modelos ARMA, ARIMA y SARIMA	15
2.1.9 Modelo autorregresivo integrado de media móvil (ARIMA).....	16
2.1.10 Medición del error	17

2.1.11	Varianza.....	18
2.1.12	Prueba de Shapiro-Wilk para normalidad	18
2.1.13	Prueba de Levene para homogeneidad de varianzas	19
3	Antecedentes	20
3.1	Trabajos de predicción de temperatura en cambio climático.....	20
3.2	Temperatura en el área del cambio climático	24
4	Metodología.....	31
5	Análisis y Resultados.....	36
5.1	Preselección de modelo de suavizamiento exponencial	36
5.2	Experimentación con Holt-Winters y SARIMA.....	39
5.3	Predicción de la temperatura en Tampico y Xalapa	42
5.3.1	Prueba Tampico.....	43
5.3.2	Prueba Xalapa.....	45
5.3.3	Predicción de temperatura de 2023 a 2027 en Tampico y Xalapa	47
5.4	Prueba de hipótesis	48
5.5	Comparación de predicción de temperatura Vs su promedio real	49
6	Conclusiones.....	51
6.1	Recomendaciones	52
	Bibliografía.....	53

Índice Tablas

Tabla 2.1. Peso asignado de acuerdo con el número de observación	11
Tabla 3.1. Resumen del estado del arte.....	27
Tabla 3.2. Comparación de métodos de pronóstico.....	30
Tabla 5.3. MAPE en etapa de validación de modelos de SE.....	38
Tabla 5.9. Comparación de MAPE y RMSE Tampico 1993-2014.....	41
Tabla 5.16. Comparación de MAPE y RMSE Tampico y Xalapa 1993-2023.....	47
Tabla 5.19. Comparación de MAPE y RMSE Tampico y Xalapa de los años 1997-2010 contra los estimados 2024-2026.....	49

Índice de Figuras

Figura 4.1. Entrenamiento, validación y prueba.....	34
Figura 4.2. Diagrama metodología.....	35
Figura 5.1. Datos de temperatura promedio Tampico C°.....	37
Figura 5.2. Gráfica de comparación de modelos en validación.....	38
Figura 5.4. Gráfica de etapa de prueba Holt Winters A.....	39
Figura 5.5. Gráfica datos de temperatura promedio mensuales.....	39
Figura 5.6. Prueba HW comparado con real.....	40
Figura 5.7. Prueba SARIMA comparado con real.....	41
Figura 5.8. Comparación de pronósticos.....	41
Figura 5.10. Datos históricos de Tampico 1993-2023.....	43
Figura 5.11. Datos históricos de Xalapa 1993-2023.....	43
Figura 5.12 Prueba de pronóstico Holt-Winters Vs real en Tampico	44
Figura 5.13. Prueba pronóstico SARIMA Vs real en Tampico 2019-2023.....	45
Figura 5.14. Prueba ronóstico Holt-Winters Vs real en Xalapa2019-2023	46
Figura 5.15. Prueba de pronóstico SARIMA Vs real en Xalapa 2019-2023.....	46
Figura 5.17. Pronóstico de temperatura mensual SARIMA Tampico 2023-2027.....	48
Figura 5.18. Pronóstico de temperatura mensual Holt-Winters Xalapa 2023-2027.....	48

1 Introducción

Los pronósticos, muy utilizados actualmente, son una herramienta que sirve para hacer estimaciones de lo que sucederá en el futuro a partir de los datos del pasado. Estos son muy utilizados en varios campos que pueden ser predecibles y generalmente con un promedio de error muy bajo como en el caso de acciones en la bolsa, pero algo muy diferente es la naturaleza, en la que cambios tan inciertos como una erupción volcánica, en el viento o actividades realizadas por el hombre como las actividades de construcción, entre otras, pueden implicar cambios en las condiciones ambientales locales o incluso a escala global. Una variable muy importante dentro del cambio climático es la temperatura, ya que es un indicador clave de las alteraciones en el sistema climático de la Tierra. El aumento del efecto invernadero causado por actividades humanas eleva la temperatura promedio del planeta, desencadenando una serie de impactos significativos. Por esto, el pronóstico de la temperatura como variable relevante en el cambio climático es muy incierto a cualquier escala, ya que existen gran cantidad de datos atípicos que pueden afectar en su precisión [1].

De acuerdo a la literatura, es muy probable que el calentamiento global llegue a aumentar 1.5 °C entre el 2030 y 2052 si continúa con el ritmo actual [2][3]. Por esto es importante realizar en este trabajo la predicción de la temperatura como variable relevante en el cambio climático en Tampico, Tamaulipas y Xalapa, Veracruz en México. Se toman estas ciudades ya que son zonas en las que se tienen registros históricos de temperatura máxima, mínima y media de hace más de 50 años, y son ciudades que va en crecimiento. Además, que son área de interés

para los ciudadanos. Para resolver el Problema de Métodos de Predicción de Temperatura como variable relevante del Cambio Climático en regiones de México, mismo al que nos referimos aquí como PTCC, hay muchos métodos, pero existen indicios de cual pudiera ser el mejor para el área de cambio climático, pues se ha publicado en la popular competencia de pronóstico mundial conocida como M4, que uno de los métodos ganadores mundialmente es el de suavizamiento exponencial SES [4]. Otro método muy utilizado que también es incluido entre los clásicos es el Modelo Autorregresivo Integrado de Media Móvil ARIMA.

Para este trabajo se utiliza el método de Holt-Winters aditivo, ya que se realizó una prueba con los otros métodos mencionados con una serie de tiempo más pequeña que la utilizada aquí y fue el que mejor desempeño mostró para el problema que se aborda. Debido a que es simple, robusto y con buenas predicciones a corto plazo, el método de Suavizamiento Exponencial sigue vigente en la actualidad y ha sido utilizado como parte de los métodos de pronóstico más exitosos [4]. Este método se compara con SARIMA que es el más conveniente para la serie de tiempo con estacionalidad que se usa, siendo también un método con buen desempeño. Ambos son métodos clásicos que han demostrado ser muy exitosos [5]. Se espera obtener un promedio de error muy bajo, y demostrar que estos métodos pueden ayudar para el problema mencionado de este trabajo. Por esto se va seleccionar y aplicar el método para pronóstico de temperatura como variable relevante del cambio climático a corto plazo en regiones de México.

1.1 Planteamiento del problema

En la naturaleza todo está interrelacionado, es por esto que, hasta pequeños cambios en el viento, las erupciones volcánicas o la construcción de nuevas represas pueden significar cambios en las condiciones ambientales locales, regionales o incluso a escala global, como la temperatura. Esto se debe a los procesos naturales que ocurren en la tierra, los océanos, la atmósfera y los otros componentes del “sistema climático”. Esto no debe preocuparnos porque se conoce que la tierra es un sistema dinámico que cambia constantemente como lo ha venido haciendo desde hace millones de años. Sin embargo, hay ciertos cambios que están empezando a afectar seriamente a los sistemas naturales y a los medios de vida de millones de personas. Estos son provenientes de las causas antropogénicas [1].

Por esto es necesario evaluar los impactos del calentamiento global en México, para conocer lo más aproximado posible su comportamiento a futuro y encontrar una forma de prevenir o mitigar esto, pero sigue siendo una tarea incierta debido a que las proyecciones de los diferentes modelos varían ampliamente, y debido al mal desempeño de los modelos, en la reproducción del clima observado de México [6].

El cambio climático está afectando a los seres humanos, sociedades [7] y ecosistemas naturales y seguirá haciéndolo durante el siglo 21 debido a las causas mencionadas. Dentro de esto, es muy probable que el calentamiento global aumente 1.5 °C entre 2030 y 2052 si continúa con el ritmo actual [2][3]. Por lo anterior es importante resolver PTCC (Métodos de Predicción de Temperatura como variable relevante del Cambio Climático en regiones de México) abordado para las ciudades de Tampico, Tamaulipas y Xalapa, Veracruz de México.

Aquí se propone estudiar el problema PTCC como se menciona en el título del trabajo, para las regiones mexicanas mencionadas y para ello es necesario hacer una predicción precisa de cómo se va a comportar a futuro la temperatura, la cual es muy afectada por diversos factores tales como la concentración de gases de efecto invernadero.

Para resolver por completo PTCC se necesitan predicciones precisas a escala mundial y más importante, a escala regional y local, pero en este trabajo se iniciará en escala regional [8][9]. La preselección de un conjunto de predictores, permite aumentar la precisión del pronóstico de manera significativa [10].

En este estudio se realizó la búsqueda de los datos de temperatura en bases de datos conocidas como CLICOM (CLImate COMputing Project) y Weather Underground para varias regiones en México que sean útiles y en gran cantidad, para después buscar el mejor método de predicción para este problema [11].

Para lo mencionado, existen muchos métodos de predicción. Los estudios de la literatura han demostrado que las técnicas de aprendizaje automático lograron un mejor rendimiento que los métodos estadísticos tradicionales para la predicción de variables climáticas [12]. Se toman en cuenta los estudios realizados en México y el mundo para conocer cuáles son los mejores métodos que se han utilizado de pronóstico, ya que todos los métodos tienen sus

variaciones [13][14]. Los métodos más usados de pronóstico son los clásicos, los cuáles por las herramientas necesarias de aprendizaje computacional es conveniente implementarlos en el lenguaje de programación Python; por otra parte, en el área de pronóstico es conocido que entre los mejores métodos clásicos se encuentran:

- Auto regresión (AR por sus siglas en inglés Autoregression) y media móvil (MA por sus siglas en inglés Moving Average).
- Modelo autorregresivo de media móvil (ARMA por sus siglas en inglés Autoregressive Moving Average), y modelo autorregresivo integrado de media móvil (ARIMA por sus siglas en inglés Autoregressive Moving Average).
- Suavizamiento exponencial simple (SES por sus siglas en inglés Simple Exponential Smoothing).
- Suavizamiento exponencial de Holt Winters (HWES por sus siglas en inglés de Holt Winter's Exponential Smoothing) [15].

Hay que hacer notar que estos métodos de pronóstico, aun cuando es común citarlos como métodos individuales, en realidad conforman familias de métodos. Por otra parte, cada año aparecen nuevos métodos que están basados en los citados anteriormente.

En la literatura especializada en el área de pronóstico se encuentran publicaciones donde se utilizan varios métodos para pronosticar variables de afectación del cambio climático, que utilizan métodos clásicos, arriba mencionados, y los de minería de datos o de aprendizaje (ML, por sus siglas en inglés Machine Learning) que utilizan análisis matemático para deducir los patrones y tendencias que existen en los datos, entre otros [16]. Por otra parte, se ha publicado a partir de la popular competencia internacional de pronóstico conocida como M4, que uno de los métodos ganadores mundialmente es el método de suavizamiento exponencial SES arriba mencionado, también es conocido que los mejores métodos de Machine Learning son los que utilizan alguno de dichos métodos clásicos. Además, al momento de escribir este documento, no se ha encontrado cuál es el mejor método clásico para este problema. No obstante, existen indicios de cual pudiera ser el mejor para el área de cambio climático, pues como se ha hecho notar anteriormente, se ha publicado que a partir de la popular competencia de pronóstico mundial conocida como M4, uno de los métodos

ganadores mundialmente es el método de suavizamiento exponencial SES arriba mencionado [4].

Se debe recalcar que hay una cantidad incontable de métodos de suavizamiento exponencial, pues utilizan una ponderación de observaciones y de pronósticos de periodos pasados y hay una diversidad de métodos de pronóstico y formas de ponderarlos. Además, la aplicabilidad de SES en combinación con otros de los métodos arriba mencionados fue presentada en una tesis doctoral en el año 2021, en dicho trabajo se verifica que los mejores métodos de ML combinan algún método clásico tipo SES o ARIMA [17]. Es de hacer notar que no existen evidencias de que suavizamiento exponencial o ARIMA haya sido analizado y adaptado para el caso de pronóstico para temperatura en cambio climático en Tampico o Xalapa.

Por esto se recalca que en esta tesis se busca encontrar el mejor método para la predicción de temperatura en México, y para obtener esto, primero es necesario conocer y analizar los métodos comunes seleccionados para el trabajo, suavizamiento exponencial y ARIMA, ya que cuando se conozcan estos van a ser probados durante la investigación propuesta, para el caso PTCC. Para probarlos se utilizará el método de software adecuado y características involucradas que se analizarán en este estudio.

1.2 Objetivo General

Seleccionar y aplicar métodos de predicción en la temperatura como variable relevante del cambio climático a corto plazo en Tampico, Tamaulipas y Xalapa, Veracruz en México.

1.3 Objetivos específicos

1. Estudiar métodos clásicos de pronóstico para determinar su aplicabilidad en variables del cambio climático.
2. Seleccionar y emplear el mejor método de predicción para el problema de cambio climático.
3. Obtener el promedio de error del método de pronóstico.

1.4 Hipótesis

Los métodos clásicos en modelos de pronóstico para temperatura a corto plazo, como los de suavizamiento exponencial, en Tampico, Tamaulipas y Xalapa, Veracruz, México, pueden lograr un promedio de error bajo, comparado a los del estado del arte.

1.5 Justificación del estudio

El cambio climático sucede a escala mundial y afecta seriamente a la humanidad y al planeta de forma muy rápida mientras estos cambios en su mayoría son irreversibles. Es por eso que es urgente aportar estudios a esta temática, como la predicción de la temperatura, ya que los cambios drásticos en las temperaturas máximas y mínimas ambientales que han sucedido en los últimos años son un reflejo de la contaminación. A partir de esto, se pueden buscar formas de reducir su impacto y aplicar medidas de mitigación. Como se menciona en el Acuerdo de París, se debe reducir las emisiones de gases de efecto invernadero ya que se propuso limitar el aumento de la temperatura global en este siglo a 2 C y esforzarse para limitar este aumento incluso a más de 1.5 C. Se tiene que encontrar un método o varios de alta eficiencia para aplicarlos a escala regional (Tampico, Tamaulipas y Xalapa, Veracruz, México) dentro de los métodos de suavizamiento exponencial o ARIMA, que han demostrado ser muy precisos en los modelos de predicción. Aún no existe este trabajo para las variables que afectan en el cambio climático como la temperatura en estas regiones de México, ya que es un proceso largo. Encontrar este método supondría otro gran avance, que puede ser utilizado en otras regiones y futuras investigaciones de este tema. Aunque es un problema complicado, sería un gran aporte a la investigación y por esto es necesario encontrar los métodos de predicción de temperatura en las regiones mencionadas en México.

1.6 Alcances

- Los métodos utilizados clásicos pueden proporcionar predicciones con promedio de error muy bajo a corto y mediano plazo.
- Los métodos clásicos de pronóstico a utilizar son de los grupos ARIMA y Suavizamiento Exponencial.
- Los métodos a utilizar son adecuados para capturar y modelar patrones estacionales en las temperaturas.

1.7 Limitaciones

- La calidad de las predicciones puede depender significativamente de la disponibilidad y calidad de los datos históricos.
- La predicción para años en los que aún no se tienen datos puede ser incierta ya que pueden ocurrir cambios inesperados o eventos extremos que afecten el cambio climático.

1.8 Organización de la tesis.

En esta sección se describe la organización de esta tesis, se describe brevemente cada sección. Empezando por el capítulo 1, se muestra una introducción para adentrarse y comprender el tema de interés en esta tesis, se plantea la razón por la que es importante y la necesidad de obtener resultados. Continúa con el capítulo 2 donde se muestra una investigación de marco teórico relevante para realizar este trabajo, yendo desde el concepto de temperatura hasta los métodos clásicos de pronóstico. En el capítulo 3 se encuentra el estado del arte, dentro de esto, en esta revisión efectuada no se encontraron trabajos sobre la predicción de la temperatura en el cambio climático en las ciudades de Xalapa y Tampico en las que se utilicen métodos clásicos entre los que destacan los de suavizamiento exponencial y ARIMA. Se muestra una tabla de comparación para estos antecedentes. En cuanto a la metodología para el pronóstico que se realiza, se encuentra en el capítulo 4. Continuando, los resultados experimentales pueden ser consultados en el capítulo 5. Finalmente, las conclusiones se encuentran en el capítulo 6.

2 Marco teórico

El clima está cambiando a través de los años, debido a procesos que ocurren dentro o fuera del sistema climático, estos cambios son los llamados antropogénicos, que incluyen las emisiones de gases provocadas por las actividades humanas. La gravedad del cambio climático actual es que se está produciendo a un ritmo tan acelerado, a comparación de los cambios históricos, que los ecosistemas naturales y las personas no estamos preparados para poder afrontarlo. Este cambio está llevando a una variación del clima promedio a mediano y largo plazo, que puede durar decenios de años o períodos más largos. Este marco teórico también explora las contribuciones cruciales que han impulsado la evolución de la predicción para el cambio climático.

2.1.1 Gases de efecto invernadero

Los Gases de Efecto Invernadero (GEI) son componentes gaseosos en la atmósfera, de origen natural o antropogénico, que absorben y emiten la radiación proveniente de la superficie terrestre, por la propia atmósfera y por las nubes. Al efecto de retener la radiación dentro de la atmósfera se conoce como Efecto Invernadero y tiene la función de mantener la temperatura en un rango en el que la vida en la tierra sea viable. Cuando existe una concentración de GEI demasiado elevada, la radiación solar se concentra en la capa baja de la atmósfera, la temperatura incrementa y la tierra tiende a calentarse; esta situación da lugar al llamado calentamiento global. Los principales gases de efecto invernadero, además del

vapor de agua (H_2O), son el dióxido de carbono (CO_2), el óxido nitroso (N_2O), el metano (CH_4), el monóxido de carbono (CO), los hidrofluorocarbonos (HFC) y otros.

2.1.2 Contaminantes atmosféricos y sus fuentes

Los contaminantes atmosféricos, que se miden normalmente en la atmósfera urbana, provienen de fuentes vehiculares y de fuentes fijas de combustión (industrias, usos residenciales climatización, y procesos de eliminación de residuos). Se clasifican en contaminantes primarios y secundarios. Los primarios son los que proceden directamente de la fuente de emisión. Los contaminantes secundarios se producen como consecuencia de las transformaciones y reacciones químicas y físicas que sufren los contaminantes primarios en el seno de la atmósfera, distinguiéndose, sobre todo, la contaminación fotoquímica y la acidificación del medio [18].

2.1.3 Temperatura

La temperatura es una magnitud física que indica la energía interna de un cuerpo, de un objeto o del medio ambiente en general.

2.1.4 Predicción de cambio climático

En cuanto a la predicción o pronóstico, los modelos climáticos tratan de simplificar la realidad y usan registros climáticos históricos (datos de años atrás) y tendencias globales de emisiones (en función de cómo evolucione la sociedad y de cómo nos comportemos) para proyectar lo que puede pasar al futuro. Por eso, los modelos pueden generar diferentes resultados que hay que interpretar con precaución. El calentamiento del clima es inequívoco, desde los años 1950 se han observado cambios sin precedentes [1].

Podemos esperar tres tipos de cambios fundamentales del clima.

Uno de ellos son los cambios en el promedio, que significa que se cambia el valor medio de una variable haciendo que el valor aumente o disminuya, otro cambio es el incremento de

variabilidad o amplitud, en el que el valor promedio se mantiene, pero el rango será más amplio y variable. El otro es el que conlleva ambos cambios (promedio + variabilidad) [19].

2.1.5 Notación en pronósticos

Cuando se trata el tema de pronósticos, la notación básica estándar utilizada es la siguiente:

Y_t = valor de una serie de tiempo en el periodo t

\hat{Y}_t = valor pronosticado de Y_t

$e_t = Y_t - \hat{Y}_t$ = residuo o error del pronóstico

2.1.6 Autocorrelación

Al medir una variable a lo largo del tiempo, las observaciones en diferentes periodos generalmente están relacionadas o correlacionadas. Esta correlación se mide usando el coeficiente de autocorrelación, se define como autocorrelación a la correlación que existe entre una variable retrasada uno o más periodos consigo misma. La fórmula (2.1) del coeficiente de autocorrelación se define de la siguiente manera:

$$(2.1) \quad r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad k=0, 1, 2, \dots$$

donde

r_k = coeficiente de autocorrelación para un retraso de k periodos

\bar{Y}_t = media de los valores de la serie

Y_t = observación en el periodo t

Y_{t-k} = observación k periodos anteriores o durante un periodo $t-k$

Un correlograma o función de autocorrelación es una gráfica de las autocorrelaciones para varios retrasos de una serie de tiempo. Los patrones en un correlograma se usan para analizar las características clave de los datos.

Con el coeficiente de autocorrelación, es posible determinar el patrón de la serie (aleatoria, tendencia, estacional, estacionario).

Aleatoriedad: Las autocorrelaciones entre Y_t y Y_{t-k} para cualquier retraso de tiempo k son cercanas a cero. Los valores sucesivos de una serie de tiempo no están relacionados entre sí.

Tendencia: Las observaciones sucesivas están altamente correlacionadas; los coeficientes de correlación son diferentes de cero para los primeros retrasos de tiempo, y de forma gradual tienden a cero, conforme se incrementa el número de retrasos.

Estacional: Se presentará un coeficiente de autocorrelación significativo en el retraso de tiempo estacional o en los múltiplos del retraso estacional. El retraso estacional se considera de tamaño 4 para datos trimestrales y tamaño 12 para datos mensuales.

Estacionaria: Es una serie de tiempo donde la media y la varianza, permanecen constantes en el tiempo. Una serie que contiene una tendencia es no estacionaria. Los coeficientes de autocorrelación de una serie estacionaria decrecen hacia cero bastante rápidamente, por lo común después del segundo o tercer retraso de tiempo.

Las autocorrelaciones de serie no estacionaria permanecen muy grandes durante varios periodos. Para analizar una serie no estacionaria, se debe de eliminar la tendencia antes de aplicar cualquier modelo de pronóstico [20].

2.1.7 Suavizamiento exponencial

El método de Suavizamiento Exponencial tuvo sus orígenes entre los años 1950 a 1960 con el trabajo de Brown [21], Holt [22], y Winters [5].

El método utiliza los promedios históricos de una variable en un período para intentar predecir su comportamiento futuro.

Se analiza continuamente un estimado dando mayor importancia a las experiencias más recientes. Este método se basa en promediar (suavizar) valores pasados de una serie dando

un menor peso a la importancia de los valores de la serie, de manera exponencialmente decreciente, de los valores más recientes hacia el pasado.

La Tabla 2.1 muestra la asignación del valor de α que corresponde al peso que se asigna a cada una de las observaciones, de acuerdo a su tiempo t .

Tabla 2.1. Peso asignado de acuerdo con el número de observación

Observación	Peso asignado
Y_t	$\alpha(\text{donde } 0 < \alpha < 1)$
Y_{t-1}	$\alpha(1-\alpha)$
Y_{t-2}	$\alpha(1-\alpha)^2$

La ecuación (2.2) de Suavizamiento Exponencial es la siguiente:

$$(2.2) \quad \hat{Y}_{t-1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t$$

Puede quedar de la siguiente manera:

$$(2.3) \quad \hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t = \alpha Y_t + \hat{Y}_t - \alpha \hat{Y}_t$$

$$(2.4) \quad \hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t)$$

Donde:

\hat{Y}_{t+1} = nuevo valor suavizado o el valor del pronóstico para el siguiente periodo

α = constante de suavización ($0 < \alpha < 1$)

Y_t = nueva observación o el valor real de la serie en el periodo t

\hat{Y}_t = último valor suavizado o el pronóstico del periodo t

Los dos modelos más sencillos que relacionan los valores observados (Y_t) de una serie de tiempo con los componentes de la tendencia (T_t), estacionales (S_t) e irregulares o aleatorios

(I_t) son el modelo de componentes aditivo y el modelo de componente multiplicativo, los cuales se describen a continuación:

2.1.7.1 Modelo de Componente Aditivo (Suavizamiento Exponencial).

El modelo de componentes aditivo funciona mejor cuando la serie de tiempo sometida a análisis tiene aproximadamente la misma variabilidad a lo largo de toda la serie. Es decir, que todos los valores de la serie caen dentro de una banda de anchura constante centrada en la tendencia. Esto se representa con la ecuación 2.5.

$$(2.5) \quad \bar{Y}_t = T_t + S_t + I_t$$

2.1.7.2 Modelo de Componente Multiplicativo.

El modelo de componentes multiplicativo, representada en la ecuación 2.6 funciona mejor cuando la variabilidad de la serie de tiempo aumenta con el nivel, es decir, los valores de la serie se dispersan conforme la tendencia aumenta, y el conjunto de observaciones tiene la apariencia de un megáfono o embudo.

$$(2.6) \quad Y_t = T_t + S_t + I_t$$

Existen 15 métodos de suavizamiento exponencial, de los que destacan:

- SES – Simple Exponencial Smoothing (no tendencia, no estacionariedad)
- Holt's Linear – Suavizamiento Exponencial Lineal de Holt (tendencia aditiva)
- Holt-Winters additive method – Suavizamiento Exponencial de Holt-Winters (tendencia aditiva, estacionalidad aditiva)
- Holt-Winters multiplicative method – Suavizamiento Exponencial de Holt-Winters (tendencia aditiva, estacionalidad multiplicativa) [5].

Debido a que es simple, robusto y con buenas predicciones a corto plazo, el método de Suavizamiento Exponencial sigue vigente en la actualidad y ha sido utilizado como parte de los métodos de pronóstico más exitosos [5] [21-23].

2.1.7.2.1 Suavizamiento exponencial doble – Holt

A diferencia del SES, el modelo de Holt logra prever la tendencia del pronóstico. En este método, aparte de la constante de suavización alfa (α), se agrega otra constante de suavización beta (β) que su valor puede estar entre 0 y 1 para la tendencia. Una beta elevada responde más rápido a los cambios en la tendencia, y una beta más bajo tiende a suavizar la tendencia actual, dando menos peso a los datos recientes. A continuación, en 2.7 a 2.11 se muestran las ecuaciones de este modelo.

$$(2.7) \quad S_t = \alpha X_t + (1 - \alpha)S_{t-1}$$

$$(2.8) \quad S'_t = \alpha S_t + (1 - \alpha)S'_{t-1}$$

$$(2.9) \quad a_t = 2S_t - S'_t$$

$$(2.10) \quad b_t = \frac{\alpha(S_t - S'_t)}{1 - \alpha}$$

$$(2.11) \quad F_{t+m} = a_t + b_t m$$

Donde

S_t = Suavizamiento simple

α = Alpha (sintonizable)

X_t = Observación real del periodo

S_{t-1} = Estimación de suavizamiento simple anterior inmediata

S'_t = Suavizamiento doble (estimación prima)

a_t = Ajuste de la serie

b_t = Ajuste de la tendencia

m = Periodos a pronosticar

F_{t+m} = Pronóstico para el periodo $t + m$

Alpha pondera la importancia de:

S_t -> Valor real actual contra valor suavizamiento simple anterior

S'_t -> Valor de suavizamiento simple contra suavizamiento doble anterior

2.1.7.2.2 Suavizamiento exponencial triple – Holt-Winters.

En este modelo se le agrega un ajuste a la estacionalidad. Incluye 3 constantes de suavizamiento. Tiene enfoque aditivo y multiplicativo.

Datos aditivos estacionales: Datos con estacionalidad constante (que no aumenta a lo largo del tiempo).

Datos multiplicativos estacionales: Datos cuya estacionalidad aumenta o disminuye a lo largo del tiempo.

A continuación, en 2.12 a 2.15 se muestran las ecuaciones para el modelo de Suavizamiento exponencial Holt-Winters Multiplicativo (Estacionalidad creciente).

$$(2.12) \quad A_t = \alpha \frac{Y_t}{S_{t-L}} + (1 - \alpha)(A_{t-1} + T_{t-1})$$

$$(2.13) \quad T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1}$$

$$(2.14) \quad S_t = \gamma \frac{Y_t}{A_t} + (1 - \gamma)S_{t-L}$$

$$(2.15) \quad Y_{t+p}' = (A_t + pT_t)S_{t-L+p}$$

En las ecuaciones 2.16 a se muestra el modelo de Suavizamiento exponencial Holt-Winters Aditivo (Estacionalidad constante)

$$(2.16) \quad A_t = \alpha(Y_t - S_{t-L}) + (1 - \alpha)(A_{t-1} + T_{t-1})$$

$$(2.17) \quad T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1}$$

$$(2.18) \quad S_t = \gamma(Y_t - A_t) + (1 - \gamma)S_{t-L}$$

$$(2.19) \quad Y'_{t+p} = A_t + pT_t + S_{t-L+p}$$

Donde

A_t = valor atenuado del periodo t

T_t = estimación de la tendencia en el periodo t

S_t = estimación de la estacionalidad en el periodo t

S_{t-L} = longitud de la estacionalidad

Y'_{t+p} = pronóstico

α = Atenuación media de los datos (0,1)

β = Atenuación estimación de tendencia (0,1)

γ = Atenuación de estacionalidad (0,1)

p = Periodos a pronosticar [24].

2.1.8 Modelos ARMA, ARIMA y SARIMA

Una serie es estacionaria si las propiedades estadísticas como la media y varianza son constantes a través del tiempo. Por otro lado, una serie no estacionaria es muy común, en la que la media y varianza varían a través del tiempo y eso lo hace impredecible y no se puede modelar.

2.1.9 Modelo autorregresivo integrado de media móvil (ARIMA)

El modelo ARIMA desarrollado por Box, ha sido utilizado para pronósticos de series de tiempo estacionarias. El modelo de pronóstico de Box-Jenkins (ARMA) es muy popular por su alta eficiencia en predicciones de series de tiempo estacionarias. Es una clase de modelos que explica o predice una serie de tiempo determinada en función de sus propios valores pasados, es decir, sus propios retrasos y los errores de pronóstico retrasados, de modo que se puede utilizar la ecuación para pronosticar valores futuros. En la familia ARIMA existen miles de métodos, cada uno identificado como ARIMA (p, d, q), donde se identifican los siguientes parámetros:

p: Orden de retraso o Número de observaciones de retraso incluidas en el modelo.

d: Número de veces que se diferencian las observaciones o grado de diferenciación.

q: Orden del promedio móvil o tamaño de la ventana de promedio móvil.

El modelo de AR (p) es el de la ecuación 2.20.

$$(2.20) \quad AR(p) : y_t = \mu + \sum_{i=1}^p (\phi_i y_{t-i}) + \omega_t$$

Por otra parte, el polinomial del método MA(q) consiste en tres partes: La primera parte es la media de las series, la segunda es la sumatoria de la multiplicación de un número finito de coeficientes MA y los residuales del modelo, y la tercera parte es el ruido blanco. La ecuación de MA(q) es la 2.21.

$$(2.21) \quad MA(q) : y_t = \mu + \sum_{i=1}^q (\theta_i \omega_{t-i}) + \omega_t$$

El modelo ARMA (p, q) está compuesto por dos ecuaciones, AR(p) y MA(q). Esta está representada en la ecuación 2.22. O también como la 2.23.

$$(2.22) \quad y_t = \mu + \sum_{i=1}^p (\phi_i y_{t-i}) + \sum_{j=1}^q (\theta_j \omega_{t-j}) + \omega_t$$

$$(2.23) \quad \phi(B)y_t = \mu + \theta(B)\omega_t$$

La notación para este método es la siguiente:

y_t = es el valor predecido o variable dependiente en el periodo t

p = es el orden polinomial de AR o el número de observaciones de retraso incluidas en el modelo

q = es el orden del modelo MA o el número de retrasos de promedios móviles

ϕ_i = los coeficientes AR(p) que se necesitan estimar (i = 1, 2, ..., p)

θ_j = los coeficientes MA(q) (parámetros) que se necesitan estimar (j= 0, 1, 2, ..., q)

μ = el valor medio de los datos de la serie de tiempo

d = representa el número de diferencias y es calculado con la siguiente ecuación

$$\Delta y_t = y_t - y_{t-1}$$

ω_t = representa el ruido blanco de la serie de tiempo en el periodo t [25].

2.1.10 Medición del error

Para medir el error del pronóstico se pueden emplear diversos métodos como la raíz del error cuadrático medio o RMSE (Root Mean Squared Error), el error absoluto o MAE (Mean Absolute Error), el error porcentual absoluto medio o MAPE (Mean Absolute Percentage Error) y el nivel de concordancia o IA (Index of Agreement) [26]. En 2.20 se muestra la

fórmula de MAPE, que es el más utilizado en el estado del arte donde se obtiene el promedio de error. Y en 2.21 de RMSE, una métrica con la que se complementa donde se obtiene la raíz del error medio al cuadrado, dándole más penalización a los valores más alejados del real.

$$(2.20) \quad MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{Real - Estimado}{Real} \right|$$

$$(2.21) \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Real - Estimado)^2}$$

2.1.11 Varianza

Es una medida de dispersión que representa la variabilidad de una serie de datos con respecto a su media. Puede calcularse como la desviación estándar al cuadrado o con la ecuación 2.22.

$$(2.22) \quad Var(X) = \frac{\sum_{i=1}^n (X_i - X')(X_i - X')}{(n-1)}$$

Donde

X' = Media aritmética

$X - X'$ = Desviación de la observación individual de la media aritmética,

n = Número de observaciones.

2.1.12 Prueba de Shapiro-Wilk para normalidad

Esta prueba es para determinar si una muestra aleatoria de la serie de tiempo proviene de una distribución normal. Se basa en la ecuación 2.23. La hipótesis nula es que los datos provienen de una población normalmente distribuida. En esta si el valor de W es pequeño, se rechaza la hipótesis nula y se concluye que los datos no son normales [27].

$$(2.23) \quad W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Donde

x_i = valores de muestra aleatorios ordenados,

a_i = constantes generadas a partir de las covarianzas, varianzas y medias de la muestra (tamaño n) de una muestra normalmente distribuida

2.1.13 Prueba de Levene para homogeneidad de varianzas

La prueba de Levene es un método estadístico en la que se evalúa si las varianzas de diferentes poblaciones o muestras son iguales. Esta prueba compara la hipótesis nula de que las varianzas son homogéneas con la hipótesis alternativa de que hay alguna diferencia significativa entre ellas [28].

3 Antecedentes

3.1 Trabajos de predicción de temperatura en cambio climático

La predicción de la temperatura se ha convertido en un área de investigación muy importante en el contexto del cambio climático. Los métodos clásicos de predicción se han destacado como herramientas fundamentales para abordar esta problemática. Este apartado incluye un estado del arte sobre la predicción de la temperatura como variable importante del cambio climático, donde se busca identificar las fortalezas y limitaciones de los métodos usados.

En múltiples estudios se ha hecho la predicción de cómo cambian las variables dependientes respecto al tiempo de muchos casos, dando información muy valiosa a partir de conocer de la forma más precisa cómo se comportarán a futuro. A continuación, se muestran trabajos realizados en los últimos años relacionados con la predicción en el campo del cambio climático y/o contaminación, que es lo que se aborda principalmente en este trabajo, incluyendo la temperatura.

Shahid *et al.* [29], realizaron un estudio en el que se propone, lograr pronosticar el tráfico utilizando datos de contaminación del aire (monóxido de carbono CO, dióxido de carbono CO₂, compuestos orgánicos volátiles VOCs, hidrocarburos HCs, óxidos de nitrógeno NO_x

y partículas en suspensión PM), ya que, como ellos mencionan, estos tienen una gran relación, por las emisiones de ciertos gases se puede saber la concentración de coches en un determinado lugar. En este estudio realizaron un análisis comparativo de 7 diferentes modelos de regresión para encontrar cual modelo da una mejor precisión. De estos, se encontró que el que dio un mejor resultado es el modelo Multi-Layer Perceptron. Pero agregado a esto, los autores propusieron un modelo de regresión que da mejores resultados que los 7 usados anteriormente. Este es un ensamble entre MLP y SVR (por sus siglas en inglés Support Vector Regression). Los resultados experimentales muestran su efectividad señalan que el error de pronóstico con este nuevo modelo se reduce 2.47%.

Por otra parte, para comparar modelos de pronóstico, y en particular para evaluar estos métodos para cambio climático, se utilizan varios enfoques entre los más comunes se encuentran, el conocido como AIC (por sus siglas en inglés, Akaike Information Criterion) y BIC (por sus siglas en inglés, Bayesian Information Criterion) utilizan medidas de que tan probable es que un método obtenga mejores predicciones que otro con el cual se compara; para ello se utilizan métricas conocidas como AUC (area under the curve) o MaxEnt (Max entropy) [30]. Aún no se sabe cuál de los numerosos enfoques es el mejor criterio para evaluar la combinación de varios modelos de pronóstico, y muchos menos para cambio climático. Sin embargo, una evaluación reciente comprobó que para el caso de la competencia M4 el enfoque BIC está entre los mejores enfoques para tal comparación; además las métricas de ese enfoque pueden ser utilizadas para obtener una combinación de métodos de pronóstico y para una selección de la combinación de máxima calidad [31].

Otro ejemplo de investigación es el de Ibarra y Huerta [32], cuyo objetivo fue construir un modelo espacial para generar un escenario de predicción de incendios para el año 2050 en el Bosque La Primavera, para ello se tomaron como referencia los datos de las proyecciones climáticas para el occidente de México y las ocurrencias históricas de incendios de los últimos años. Este modelo fue tomado de un subconjunto de los mejores 10 modelos con menos de 10 % de error por omisión. El modelo espacial en este trabajo tuvo una AUC (área bajo la curva, es la salida gráfica en la que se observa la capacidad de discriminación de una

ocurrencia “sensibilidad” contra la capacidad de discriminación de una ausencia “especificidad”) de 0.81, lo que indica que la robustez para clasificar las presencias de los incendios en el APFFLP fue buena. En conclusión, el modelo espacial desarrollado con MaxEnt [33], bajo el escenario del cambio climático demostró ser una herramienta confiable para predecir del APFFLP para el año 2050. Usando también MaxEnt, está el trabajo de Qin, *et al.* en Truja, China [34]. las áreas por incendiarse

En otra ocasión, K. Krishna Rani Samal, *et al.* [35], realizaron un estudio de investigación que explora un nuevo modelo de pronóstico de contaminantes denominado autocodificador de red convolucional temporal de salida múltiple (MO-TCNA por sus siglas en inglés Multi-output temporal convolutional network autoencoder). La red MO-TCNA sirve tanto al pronóstico de contaminantes PM_{2.5} como PM₁₀ para varias ubicaciones, en lugar de realizar pronósticos de contaminantes de salida única y específicos del sitio. En consecuencia, los resultados experimentales muestran que la red MO-TCNA ahorra tiempo y tiene un mejor rendimiento que los modelos tradicionales de pronóstico específicos del sitio. También ha mostrado sus resultados satisfactorios de predicción a largo plazo para la contaminación por PM_{2.5} y PM₁₀ en múltiples sitios. Se obtuvo un error absoluto (MAE) de 32.

Continuando con el área de cambio climático Behzadi *et al.* [36], han investigado diferentes aspectos como la tendencia anual y estacional de la temperatura. Esto ha sido realizado con datos de estaciones meteorológicas ubicadas en la Ciudad de México y se ha comparado con conjunto de datos y proyecciones obtenidos con diversos modelos como GCM (por sus siglas en inglés, General Circulation Model). Este análisis se realizó con las series temporales utilizando la prueba no paramétrica de Mann-Kendall y el estimador de pendientes de Sen. Los resultados muestran tendencias crecientes estadísticamente significativas en la temperatura media de invierno y la precipitación de verano por tasas promedio de 0.1 ° C / década y 17.8 mm / década, respectivamente. Luego, se proyecta que, dentro de los próximos 30 años, la temperatura continuará aumentando entre 1 ° C y 3 ° C con una señal poco clara en el cambio de precipitación, pero posiblemente un aumento en las precipitaciones

monzónicas de finales de temporada. Obteniendo un nivel de significancia del 5% de forma general para la precipitación y temperatura.

Acerca de la contaminación del aire, Liu *et al.* [37], realizaron un estudio del pronóstico de esta concentración de NO₂ basado en modelos combinados de ponderación de predicción usando la transformada de ondas, que es un tipo especial de transformada matemática que representa una señal en términos de versiones trasladadas y dilatadas de una onda finita. Se demuestra que el modelo combinado construido en este artículo que está basado en métodos de la descomposición discreta de ondícula y LSTM, GRU y Bi-LSTM puede mejorar la precisión de la predicción combinando las ventajas de cada modelo de predicción y contrastándolo con cualquier modelo de predicción por sí sólo, el modelo combinado es más rentable para la predicción de la concentración de NO₂ en Beijing.

Es de llamar la atención el método de Wavelets (o transformación de ondas) en la manera implementada por Wang Wang *et al.* [38], el cual es aplicado con éxito en otros campos diferentes al cambio climático. En particular, este método se aplicó para la predicción de precios de las acciones o índices (el Shanghai Composite Index) de la bolsa de valores de Shanghai con resultados de mejor precisión que otros métodos de la literatura. En dicho método se aplica una transformación de ondas, común en las áreas de ingeniería eléctrica, electrónica y otras áreas de la física, para descomponer la serie de tiempo en 6 capas de baja y alta frecuencia buscando eliminar al máximo el ruido existente. En la referencia anterior se muestra que este método es superior al popular método de pronóstico conocido como BPNN (Back Propagation Neural Network) y se mostró que la capa 4, resultado de la descomposición de la serie de tiempo, obtiene errores más pequeños. Sobre esto, también existe una revisión de Liu *et al.* [39], realizada en 2020 que consiste en la descomposición de estrategias para la predicción de la contaminación del aire, entre ellas se revisan; Empirical Mode Decomposition o EMD [40,41], variational mode decomposition (VMD), wavelet decomposition (WD) y secondary decomposition (SD).

Otros métodos de predicción muy precisos aplicados en áreas diferentes al del cambio climático son los publicados por Ponce y Frausto, *et al.* [42] y el modelo FCTA (Forecasting Combined Methodology with a Threshold Accepting Approach) de Frausto, Rodríguez, González *et al.* [31].

Se realizó también un estudio por Gao *et al.* [43], en el que se analiza con un método de machine learning (ML) para la predicción espacial de la contaminación del suelo, utiliza un método llamado Rationale of two-point machine learning (TP-ML) para mejorar la exactitud de la predicción espacial de las concentraciones de SHM. Es un nuevo método de ML, que agrega un parámetro en comparación con algún otro método de ML adoptado. Al aplicar TP-ML, el bosque aleatorio puede ser reemplazado por otros modelos de aprendizaje supervisado de acuerdo con situaciones específicas. Debe ser capaz de integrar información espacial y no espacial. Dentro de la predicción espacial, en este otro estudio se toma en cuenta la contaminación del aire [44].

3.2 Temperatura en el área del cambio climático

En cuanto a la predicción de la temperatura como variable relevante en el cambio climático, se encontraron los trabajos mencionados a continuación. Empezando por mencionar el realizado por Lin *et al.* [45], en el que realiza el pronóstico de la temperatura máxima en el área del cambio climático en la región metropolitana de la ciudad de Tapei en Taiwan, utilizando un método híbrido de un ensamble multi dimensional complementario y una red neuronal, denominado en inglés como Multi-dimensional Complementary Ensemble Empirical Mode Decomposition y Radial Basis Function Neural Network (MCEEMD-RBFNN). Con esto se logra pronosticar la temperatura máxima diaria de los próximos 7 días. En resultados de este trabajo se obtuvo que su promedio de error de acuerdo a pruebas MAPE fue 2–4.5%. Cabe recalcar que se logra un error bajo porque es una predicción a muy corto plazo.

Por otra parte, Morid *et al.* [46], realizó un estudio donde hace un framework integrado para la predicción del impacto del cambio climático de un río simulado, tomando en cuenta la temperatura del agua y otros parámetros hidrológicos. La configuración de este modelado está basada en una combinación de dos modelos denominados Soil and Water Assessment Tool (SWAT) y el international river interface cooperative (iRIC). El módulo de la temperatura del agua en SWAT fue modificado para observar la eficacia del método. Con este método se obtuvo que la predicción de esta temperatura obtuvo un error cuadrático medio RMSE aproximadamente entre 2 y 3 °C.

Continuando con la temperatura en el agua, Mercado-Bettín, Clayer *et al.* [47], realizaron un pronóstico de la temperatura del agua en lagos y presas tomando en cuenta las estaciones climáticas. Usaron el sistema de pronóstico de estaciones climáticas (seasonal climate forecast system por sus siglas en inglés SEAS5) y por su nombre en inglés reanalysis (ERA5) of the European Centre for Medium Range Weather Forecasts (ECMWF). De acuerdo a los resultados no se logró un avance significativo en el pronóstico.

Al Sayah, *et al.* [48], realizaron un aporte para la evaluación del cambio climático en una cuenca hidrológica en el Mediterráneo obteniendo información a distancia y usando un modelo de pronóstico ARIMA para la temperatura promedio, mínima y máxima en la zona. Para el pronóstico de la temperatura promedio, se obtuvo que la tendencia de la temperatura aumentará más de 0.9 °C para el 2030, con el modelo ajustado SARIMA (2,0,0) (0,1,1)¹². De este, se logró un R^2 en el modelo de 0.823 y una bondad de ajuste R^2 de 0.825.

Otro estudio en el que se utilizó el modelo SARIMA es el de Farsi M, *et al.* [26] donde se utilizó un algoritmo genético para optimizarlo. Este fue usado para predecir como prueba datos de temperatura media en la India del 2000 al 2017 y se obtuvo un RMSE de 1.16 en el mejor de los casos.

Dentro de la predicción por métodos clásicos, Su, Y, *et al.* [49] realizó un pronóstico de la huella ecológica urbana del agua con suavizamiento exponencial doble, donde usaron un ajuste de $\alpha=0.3$ logrando una suma cuadrática de residuos (SSE) de 0.005.

Indriani R., *et al.* [50] realizaron también un pronóstico con suavizamiento exponencial, pero usaron el de Holt-Winters de tipo aditivo para predecir la temperatura máxima y mínima del aire. Para la temperatura máxima emplearon un $\alpha = 0.4$, $\delta = 0.1$, y un $\gamma = 0.1$, obteniendo un MAPE = 1.62917%. En cambio, para pronosticar la temperatura mínima emplearon $\alpha = 0.7$, $\delta = 0.1$, y $\gamma = 0.1$, obteniendo un MAPE = 1.92473%.

Para predecir la temperatura y precipitación ambiental mensual, Papacharalampous, G, *et al.* [51] utilizaron varios métodos automáticos de predicción, entre ellos ARFIMA, método Naive, ARMA, BATS, y un método de suavizamiento exponencial, quedando este con el error RMSE más bajo globalmente de 1.68.

Otro estudio que utiliza suavizamiento exponencial es el de Mahajan, S., *et al.* [52] pero en este modelo se le agrega un drift (ESD) para pronosticar a corto plazo el PM 2.5. Obtuvieron un error medio de 0.16, el más bajo comparado contra ARIMA que fue de 11.47 y NNAR de 1.19. Ganando ESD también en el menor tiempo de computación.

En resumen, existen muchos métodos de predicción y ensambles entre estos para disminuir el error, y numerosos estudios, pero no se ha encontrado alguno para temperatura como variable relevante del cambio climático en las regiones de Tampico, Tamaulipas y Xalapa, Veracruz en México. Se busca aplicar los métodos clásicos que se proponen en este trabajo como SARIMA y suavizamiento exponencial usados también por Liu, *et al.* [53].

A continuación, en la tabla 3.1 se muestra un resumen de los artículos del estado del arte más importantes que se encontraron en esta investigación, donde muestra el título de estos, la variable que se pronosticó, el método o modelo de predicción que se utilizó, y la precisión que se obtuvo.

Tabla 3.1. Resumen del estado del arte.

Título	Variable	Método	Precisión
Dynamic assessment and forecast of urban water ecological footprint based on exponential smoothing analysis.	Huella ecológica urbana del agua.	Suavizamiento exponencial doble [49].	Suma residual de cuadrados (SSE) = 0.0053
Forecasting of Air Temperature in Cilacap Regency with Triple Exponential Smoothing (Holt-Winters) Method.	Temperatura máxima del aire en el cambio climático.	Triple Exponential Smoothing (Holt-Winters) with additive type [50].	Temperatura máxima. $\alpha = 0.4, \delta = 0.1,$ and $\gamma = 0.1$ MAPE=1.629% Temperatura mínima. $\alpha = 0.7, \delta = 0.1,$ and $\gamma = 0.1$ MAPE=1.924%
Predictability of monthly temperature and precipitation using automatic time series forecasting methods.	Temperatura y precipitación ambiental.	Método naïve; AutoRegressive, Fractionally Integrated Moving Average (ARFIMA), Exponential smoothing state-space model with Box-Cox transformation; ARMA; Trend and Seasonal components (BATS); Simple exponential smoothing, Theta y Prophet methods [51].	Temperatura SES RMSE= 1.68

Short-Term PM2.5 Forecasting Using Exponential Smoothing Method: A Comparative Analysis.	Partículas en suspensión (PM 2.5)	Suavizamiento exponencial con drift (ESD) [52].	Mean error= 0.16
Daily maximum temperature forecasting in changing climate using a hybrid of Multi-dimensional Complementary Ensemble Empirical Mode Decomposition and Radial Basis Function Neural Network.	Temperatura máxima en el cambio climático.	Método híbrido de un ensamble multi dimensional complementario y una red neuronal, denominado en inglés como Multi-dimensional Complementary Ensemble Empirical Mode Decomposition y Radial Basis Function Neural Network (MCEEMD-RBFNN) [45].	MAPE=2–4.5%
Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model.	Contaminación del aire. (Concentración de NO ₂)	Wavelet transform - weighting forecasting model [37].	MAPE = 5.02%
Analysis of the climate change signal in Mexico City given disagreeing data sources and scattered projections.	Precipitación y temperatura ambiental.	Software de análisis para el cambio de climático en la Ciudad de México corrigiendo otras fuentes publicadas. (CMIP5 y GCM) [36].	Significancia = 5%
Towards greener smart cities and road traffic forecasting using air pollution data.	Monóxido de carbono, dióxido de carbono, compuestos orgánicos volátiles, hidrocarburos, óxidos de nitrógeno y partículas en suspensión	Ensamble entre MLP y SVR [29].	Error promedio = 2.47%
Climate change and prediction of fires for the year 2050 at the La Primavera Forest, Jalisco,	Incendios forestales.	Modelo espacial [32].	Error promedio < 10%

Maxent modeling for predicting impacts of climate change on the potential distribution of <i>Thuja sutchuenensis</i> Franch., an extremely endangered conifer from southwestern China.	Impactos del cambio climático en una conífera.	Modelo espacial desarrollado con MaxEnt (Máxima entropía) [33].	Nivel “excelente” y “preciso” con niveles AUC altos.
Multi-output TCN autoencoder for long-term pollution forecasting for multiple sites.	PM 2.5 y PM 10, partículas en suspensión con un diámetro aerodinámico de hasta 2.5 y 10 μm respectivamente	Multi-output temporal convolutional network autoencoder (MO-TCNA) [35].	MAE (error absoluto) = 32.
A framework for climate change assessment in Mediterranean data-sparse watersheds using remote sensing and ARIMA modeling.	Temperatura superficial promedio.	SARIMA (2, 0, 0) (0, 1, 1) 12 [48].	Bondad de ajuste $R^2 = 0.825$
Parallel genetic algorithms for optimizing the SARIMA model for better forecasting of the NCDC weather data.	Temperatura ambiental promedio.	SARIMA (0,0,0) (1, 0, 1, 12) [26]	RMSE = 1.16

Entre los principales métodos desarrollados relacionados a este proyecto se encuentra en la tabla 3.2 una comparación entre estos. En esta se compara en cuales métodos de los artículos seleccionados se ha hecho un ensamble, si se usa machine learning, SES, ARIMA, o si se predice la temperatura. Proporciona una visualización clara de los diferentes métodos que se han utilizado en el estado del arte. También es de importancia que no se encontró un artículo donde se haga el estudio en Tampico o Xalapa.

Tabla 3.2. Comparación de métodos de pronóstico.

Método	Ensambl e	Machin e Lear- ning	Predice temperatur a	SE S	ARIM A	Se aplica a regione s de México	Se aplica: Tampico & Xalapa
SE doble [49]	X	X	X	✓	X	X	X
SE triple aditivo [50].	X	X	✓	✓	X	X	X
SED [52].	✓	✓	X	✓	X	X	X
Ensembl e Neural -Network [45]	✓	✓	✓	X	X	X	X
Ensembl e Wavelet transform [37]	✓	✓	X	✓	X	X	X
CMIP5 y GCM [36]	✓	X	✓	✓	X	X	X
MLP-SVR [29].	✓	✓	X	X	X	X	X
MaxEnt [32-34]	✓	✓	✓	X	X	✓	X
SES-Hurst [42]	✓	✓	X	✓	X	X	X
FCTA [31]	✓	✓	X	✓	X	X	X
SARIMA [48]	X	✓	✓	X	✓	X	X
SARIMA [26]	X	✓	✓	X	✓	X	X

4 Metodología

En este apartado se especifica la metodología de este trabajo. Contiene las actividades a seguir para el cumplimiento del objetivo general.

El proceso de pronóstico generalmente está compuesto por siete pasos, que se describen a continuación.

1. Definición del problema.

En este paso se decide qué es lo que se desea pronosticar, en este caso la temperatura promedio mensual en Tampico y Xalapa y qué expectativas se tiene. Se deben resolver las preguntas de cuántos pasos a futuro se desean pronosticar, como que la predicción sea de corto, mediano o largo plazo, el intervalo de tiempo t , y también estimar el nivel de precisión del pronóstico que se busca, de acuerdo a los antecedentes a problemas parecidos al de investigación. En donde el marco teórico señala que las métricas de error más usadas son el MAPE y RMSE, que son dos métricas que se complementan y son las usadas en este trabajo.

2. Colección de datos

Es donde se obtienen los datos históricos de la variable que se desea pronosticar, en este caso se toma como variable relevante del cambio climático a la temperatura ya que es

muy importante para el problema [54]. Se selecciona una serie de datos grande, con muchos datos históricos para así obtener una mejor predicción posteriormente.

3. Análisis de los datos.

En este dato se revisan los datos de la serie de tiempo seleccionada, en esta se revisa visualmente si hay datos faltantes o datos atípicos, después se rellenan en este trabajo con los promedios de los 4 datos más cercanos. Estos datos también deben ser inspeccionados para reconocer patrones, como los componentes de tendencia, estacional y cíclicos, que normalmente son marcados en las series de temperatura promedio ambiental. Se hacen las pruebas de normalidad y homogeneidad, siendo la de Shapiro-Wilk y Levene respectivamente. La serie de tiempo original de esta investigación se constituye de promedios diarios de temperatura, pero después de realizar experimentación se nota que el pronóstico mejora si se utilizan datos mensuales. Es por esto que se transformó esta serie de datos diarios a datos mensuales.

4. Selección del modelo y ajuste.

Para este trabajo se hizo una prueba preliminar con los métodos de suavizamiento exponencial citados anteriormente y el que mejor desempeño tuvo fue el suavizamiento exponencial triple aditivo. También se seleccionó el modelo SARIMA porque es el que tiene mejor desempeño en el estado del arte para trabajos parecidos y dentro de los métodos ARIMA y el más adecuado para la serie de tiempo utilizada en este trabajo. Se ajustaron los modelos a los datos: el ajuste de atenuación, tendencia, y estacionalidad para el SET aditivo y p , d , q y estacionalidad para SARIMA.

5. Validación del modelo.

En esta parte se evaluó el desempeño de los pronósticos, comparando los datos pronosticados con los reales. Se hace un ajuste de las constantes del modelo con la serie de tiempo de temperatura. En este trabajo, para verificar el error entre la validación y los datos reales, para hacer un ajuste, se utilizó la métrica MAPE (Error Porcentual Absoluto Medio) y también RMSE (Raíz del Error Cuadrático Medio)

6. Implementación del modelo.

En este ya se realizó la selección del mejor modelo en la etapa de validación, por lo que se ejecuta este y se obtienen los resultados. Se hace una prueba y se mide el MAPE y RMSE, estos se pueden comparar con los del estado del arte ya que son de los más usados. Sumado a esto, a la mejor predicción de los métodos obtenidos se le hace un ajuste usando suavizamiento exponencial simple, doble y triple. Se revisa el error de cada uno y el que tenga el más bajo es el seleccionado.

7. Monitoreo del desempeño del modelo de pronóstico.

Se realiza para asegurar que el modelo se ejecuta de manera satisfactoria.

Para este trabajo se ha realizado la siguiente metodología:

1. Estudio e investigación de los métodos clásicos de pronóstico para determinar su aplicabilidad en variables del cambio climático.
 - Se realiza el estudio de modelos de suavizamiento exponencial y ARIMA. Es la investigación de usos y precisión de cada familia de métodos para tener un amplio conocimiento y así poder seleccionar y ajustar el que sea más útil posteriormente.
 - Aprendizaje para correr modelos de predicción por suavizamiento exponencial y ARIMA. Se estudia sus modelos matemáticos para poder implementarlos en el lenguaje de programación Python. Se realiza para lograr correr estos métodos, los que son seleccionados como los mejores para el área de cambio climático de acuerdo al estado del arte, en específico la temperatura en la región de Tampico, Tamaulipas y Xalapa, Veracruz.
2. Selección y empleo del mejor método de predicción para el problema de cambio climático.
 - Se realiza una limpieza de datos, revisando si hay datos faltantes con la ayuda de Excel, rellenando estos promediando los cuatro valores más cercanos. Posteriormente se realiza una gráfica de dispersión con los datos para así

revisar visualmente los datos atípicos. Se hace con el método de z-score un análisis de outliers y se corrigen. Se hace una normalización de los datos con z-score para tener un mejor pronóstico y al final se vuelve a transformar para comparar con los datos reales.

- Selección de los métodos de predicción. Se selecciona el o los modelos de pronóstico, en este caso se eligieron Suavizamiento Exponencial Simple, Holt's Linear, Holt-Winters additive method y Holt-Winters multiplicative method y el Modelo Autorregresivo Integrado de Media Móvil estacional (SARIMA)
- Validación y entrenamiento del modelo. Se entrena el modelo con el 70% de los datos de la serie de tiempo, esto se realiza para enseñarle al modelo cómo se comporta la serie. Después de esto se realiza una predicción con el 15% de los datos, y se compara con los datos reales que ya se tienen, analizando con los residuos de error de estos. Se toman muchos datos de validación ya que estos se utilizan para un ajuste final. En la figura se puede observar la división de las series de tiempo en estas etapas.

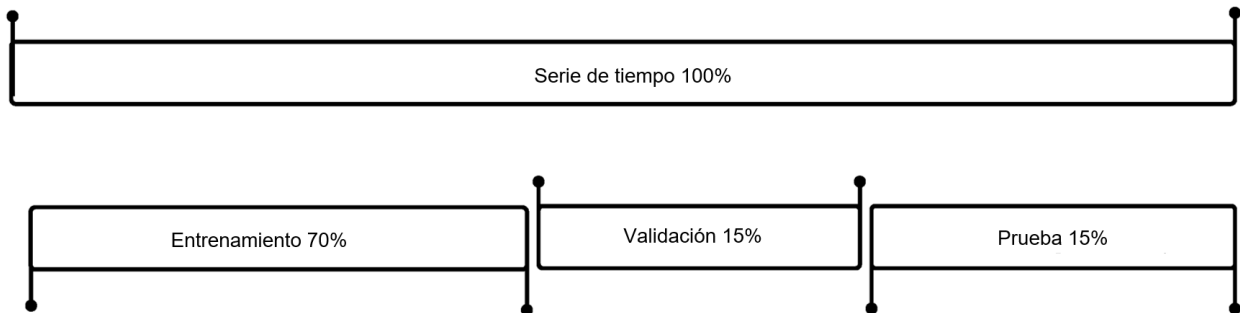


Figura 4.1. Entrenamiento, validación y prueba.

- Ejecución del modelo. Se corren los métodos con menor error con los datos históricos seleccionados, en la fase de prueba del 15% de los datos restantes. Estos también conforman gran parte de la serie porque se necesita hacer una predicción a mediano plazo para los datos que se tienen disponibles.
- Obtención de resultados.
- Monitoreo del desempeño del modelo. Sirve para asegurarse que el modelo se ejecutó de forma satisfactoria.

3. Obtención del error del pronóstico.

- Determinación del nivel de precisión del pronóstico por el cálculo de Root Mean Square Error (RMSE) y Mean Absolute Percentage Error (MAPE).
- Comparación del error de los métodos de pronóstico. Se compara a algunos pronósticos del estado del arte.
- Se grafican los resultados.

En la figura 4.2 se muestra un esquema de la metodología del pronóstico. Se utilizan los datos mensuales de temperatura de cada ciudad. En la primera fase se hace la estimación con HW, después se ajusta este pronóstico con varios métodos y se elige el mejor en una segunda etapa de validación. En la fase 2 se hace la estimación con SARIMA, después el pronóstico se ajusta igual que en la fase uno, con el mejor método de SE. Finalmente se evalúa el error de la predicción de ambos comparándolo con lo real y se elige el mejor de estos.

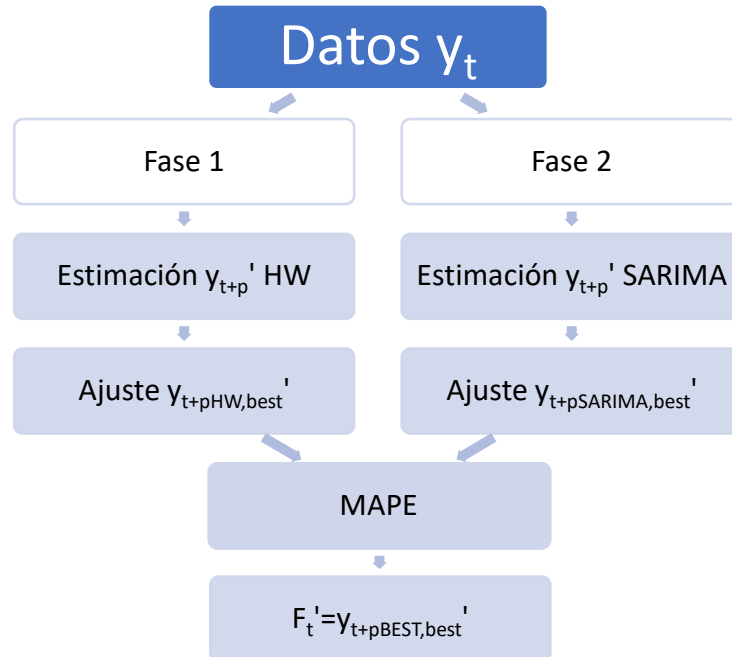


Figura 4.2. Diagrama metodología.

5 Análisis y Resultados

En esta sección se describen los datos utilizados para este trabajo, y resultados experimentales.

5.1 Preselección de modelo de suavizamiento exponencial

Para iniciar se hace esta experimentación para elegir el método más conveniente de suavizamiento exponencial para este trabajo. Para esta sección se utilizó una serie de tiempo de 100 días de temperatura promedio diaria en la ciudad de Tampico, Tamaulipas México, con fecha del 23 de septiembre de 2014 al 30 de diciembre del mismo año. Obtenidas del sitio CLICOM.

En la figura 5.1 se muestra una gráfica de los datos de la serie de tiempo utilizada.

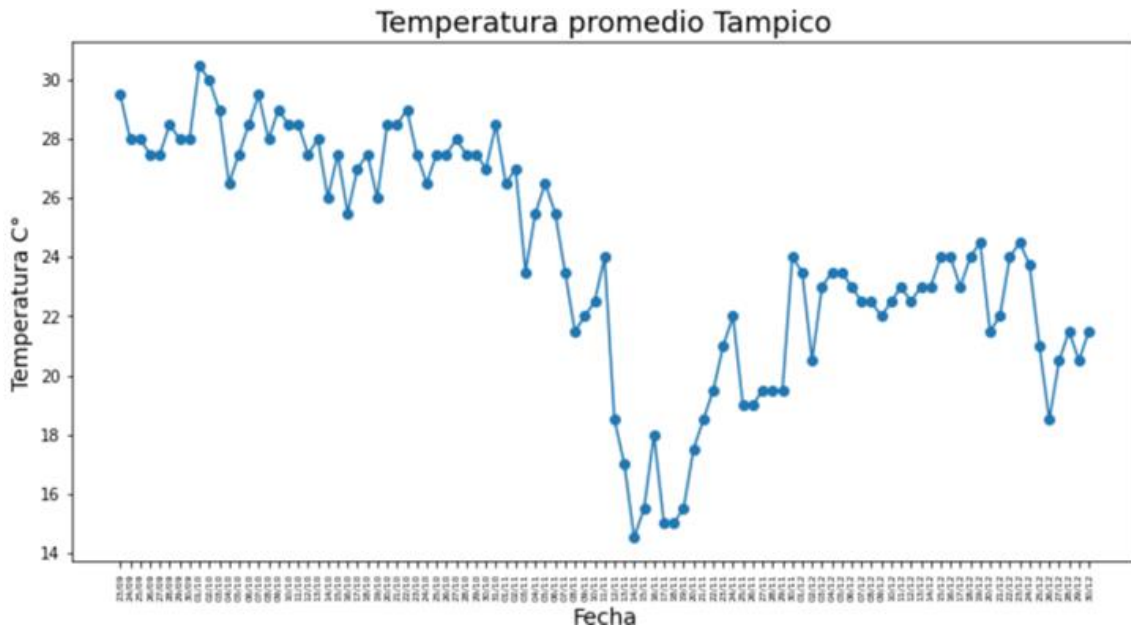


Figura 5.1. Datos de temperatura promedio Tampico C°.

Se empezó por la etapa de validación para elegir el método con menor error.

En la figura 5.2 se muestra la gráfica de comparación de validación ajustada con; el modelo de suavizamiento exponencial simple, con una constante de suavización $\alpha = 0.6$, suavizamiento exponencial doble- Holt con un $\alpha = 0.4$ y un $\beta = 0.1$, suavizamiento exponencial triple Holt-Winters aditivo con un $\alpha = 0.8$ un $\beta = 0.3$ y un $\gamma = 0.6$, también se incluye el suavizamiento exponencial triple Holt-Winters multiplicativo con un $\alpha = 0.8$ un $\beta = 0.2$ y un $\gamma = 0.3$. En esta etapa se puede notar que la predicción de SES es una línea que no sigue tendencia, todos los valores son el mismo, característica importante de este método. No tiene buenas predicciones más allá de un paso delante. En el modelo de Holt se puede notar que este sí sigue una tendencia, y finalmente en las predicciones con Holt Winters se nota que siguió una estacionalidad que logró captar con los datos de entrenamiento.

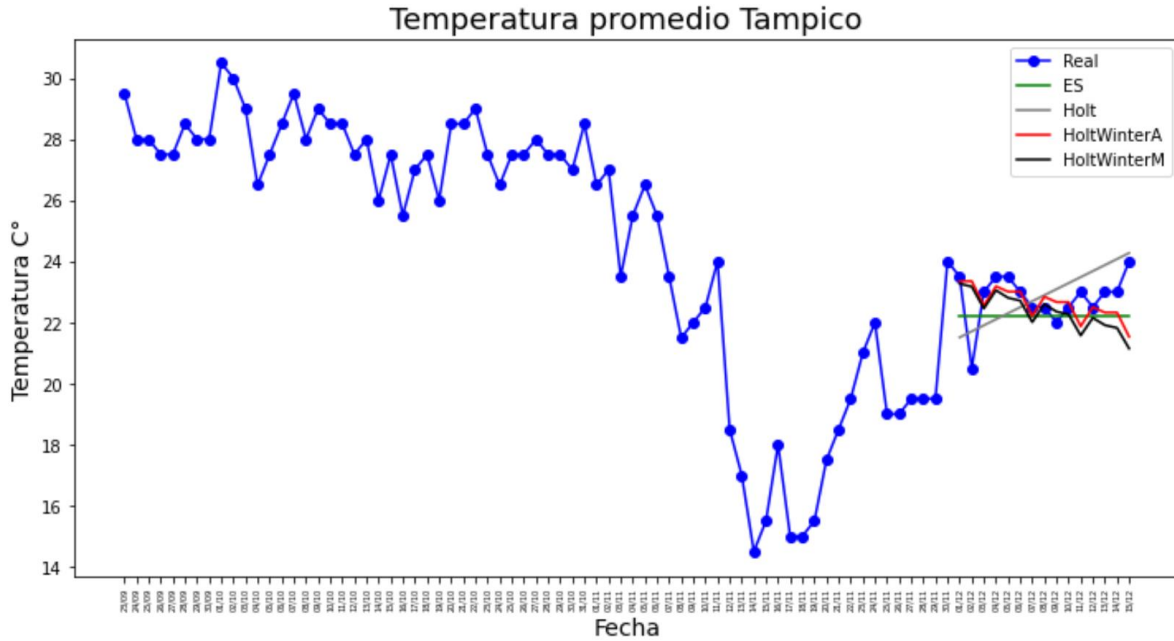


Figura 5.2. Gráfica de comparación de modelos en validación.

Para esta etapa de validación se compararon los reales contra los pronosticados obteniendo un MAPE como se observa en la tabla 5.3. En donde se ve que el método con mejor desempeño obteniendo el menor MAPE fue el Holt-Winters aditivo.

Tabla 5.3. MAPE en etapa de validación de modelos de SE.

Temperatura promedio Tampico, Tamaulipas, México.				
	Método de pronóstico			
	SES	SED Holt	SET Holt-Winters aditivo	SET Holt-Winters multiplicativo
MAPE (%)	3.7	4.0	3.1	3.8

Se seleccionó el mejor modelo de la etapa de validación, y se toma este para pronosticar la prueba siendo 15 pasos a futuro, el 15% restante de la serie de tiempo. En la figura 5.4 se muestra una gráfica de los resultados de este pronóstico.

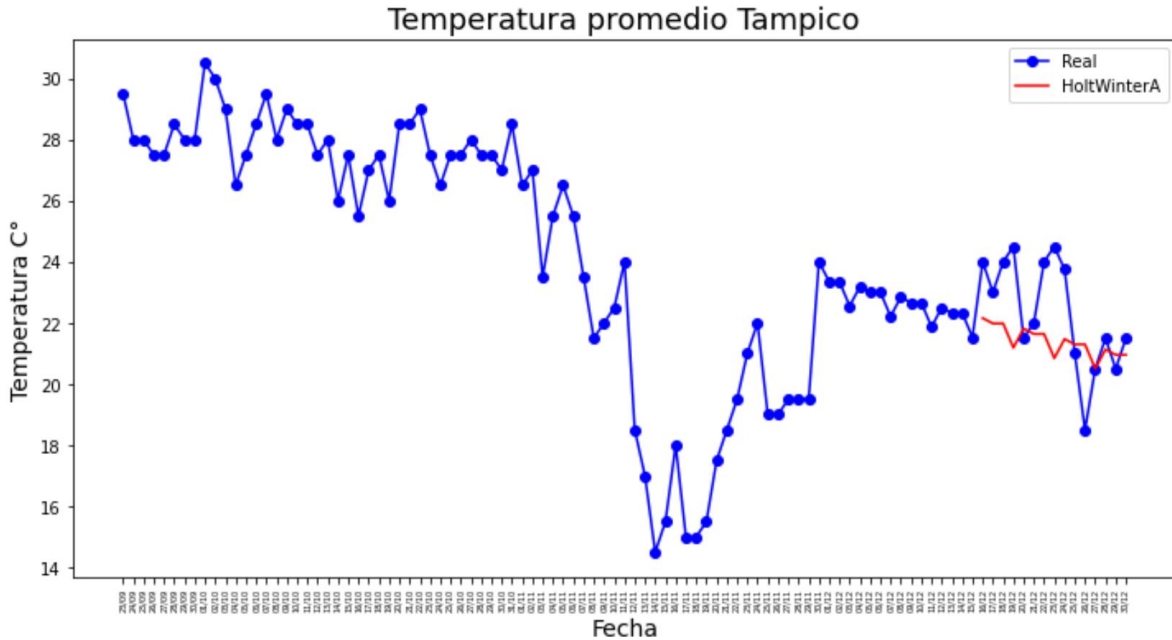


Figura 5.4. Gráfica de etapa de prueba Holt Winters A.

Se mide el error de este pronóstico, obteniendo un MAPE de 6.3%.

5.2 Experimentación con Holt-Winters y SARIMA

En esta parte se utiliza una serie de tiempo de temperatura ambiental en Tampico de 21 años tomando la temperatura promedio mensual, iniciando en enero de 1993 a mayo de 2014. Inicialmente se contaba con esta misma serie de tiempo con datos diarios, pero debido al mal desempeño del pronóstico con Holt-Winters y SARIMA con errores medidos con MAPE del 25% se decidió cambiar a datos mensuales para tener un mejor desempeño a mediano plazo. Se muestra en la figura 5.5. Una gráfica de estos datos, donde se puede observar la estacionalidad de la serie.

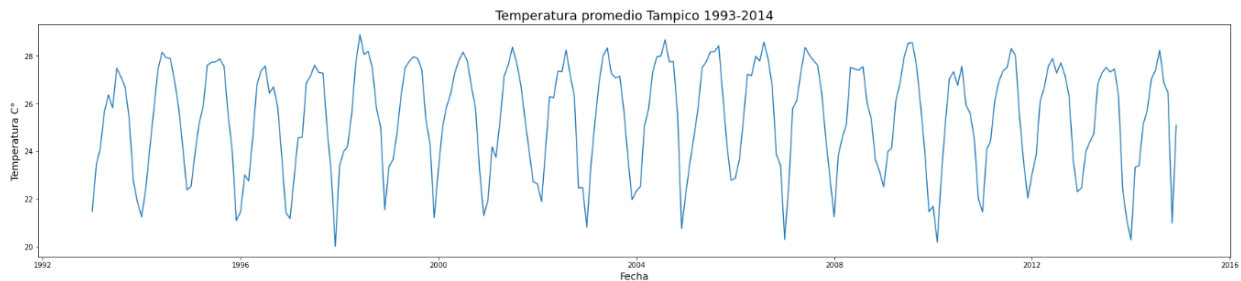


Figura 5.5. Gráfica datos de temperatura promedio mensuales.

Para la selección del método entre suavizamiento exponencial triple y SARIMA y el ajuste de los parámetros se dividió la serie de tiempo desde enero 1993 a mayo 2008 para entrenamiento (226 meses); de junio 2008 a agosto 2011 para validación (38 meses); y desde Septiembre 2011 a Noviembre 2014 para prueba (38 meses).

Los resultados de la predicción mensual de la temperatura en el cambio climático usando SE triple Holt-Winters, aditivo en Tampico, se muestran en la figura 5.6. Se muestra una gráfica de la predicción utilizando el método SE triple Holt-Winters aditivo, con los siguientes ajustes; $\alpha = 0.1$ un $\beta = 0.1$ y un $\gamma = 0.6.$, y una estacionalidad correspondiente a los meses del año. Obteniendo un error MAPE de 3.0%. También se realizó un ajuste de residuales de este pronóstico con SES, Holt linear, Holt-Winters y se logró con este último reducir el MAPE a 2.8%.

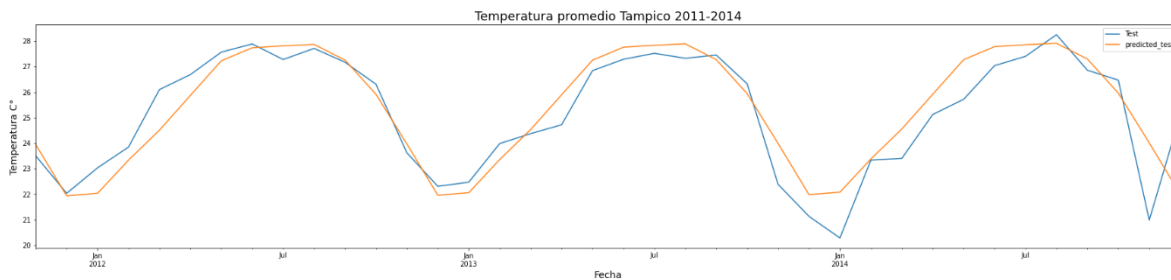


Figura 5.6. Prueba HW comparado con real.

Por otro lado, los resultados de la predicción mensual de la temperatura en el cambio climático usando SARIMA en Tampico, se muestran en la figura 5.7. Se muestra una gráfica de la predicción utilizando SARIMA (0,1,0), (1,0,1) y una estacionalidad correspondiente a los meses del año. Obteniendo un error MAPE de 3.5%. También se realizó un ajuste de residuales de este pronóstico con métodos de suavizamiento exponencial, pero no se logró disminuir el error.

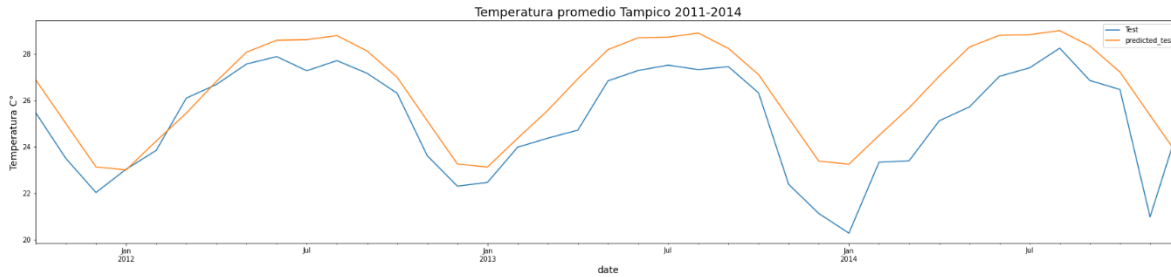


Figura 5.7. Prueba SARIMA comparado con real

En la figura 5.8 se muestra una gráfica con los resultados de la predicción de ambos métodos, en el que se logra apreciar cómo la predicción con suavizamiento exponencial triple es más acertada a los valores reales.

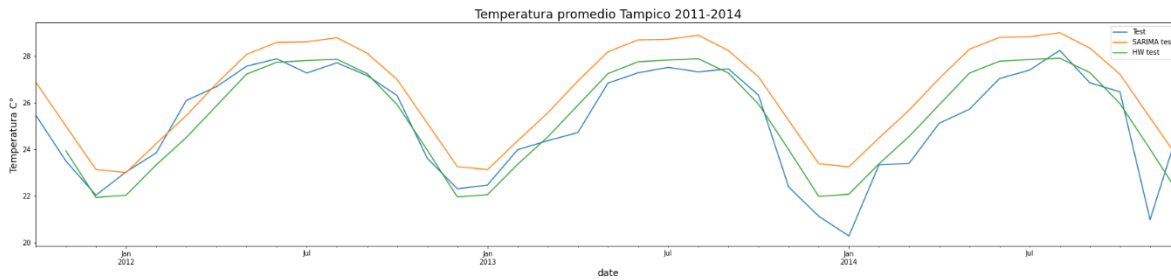


Figura 5.8. Comparación de pronósticos

En la tabla 5.9 se muestra una comparación de los resultados de MAPE y RMSE para los dos métodos utilizados SARIMA y Holt-Winters aditivo, comparando este último con los resultados del análisis de residuales

Tabla 5.9. Comparación de MAPE y RMSE Tampico 1993-2014.

Temperatura promedio Tampico, Tamaulipas, México.			
	Pronóstico		
	SET Holt-Winter aditivo	Ajuste residuales SET	SARIMA
MAPE (%)	3.0	2.8	3.5

5.3 Predicción de la temperatura en Tampico y Xalapa

Para esta parte de la investigación se usan datos recientes de Tampico, Tamaulipas ciudad de interés y de Xalapa, Veracruz, ya que cuentan con una base de datos históricos muy amplia.

Se busca obtener un buen pronóstico dentro de la serie de tiempo y obtener el error por medio de las métricas MAPE y RMSE para encontrar qué método es más conveniente para cada ciudad y utilizarlo para predecir la temperatura promedio mensual de los años siguientes del 2023.

En estas ciudades se hizo una recopilación de los datos de temperatura ambiental promedio mensual de 30 años desde septiembre de 1993 a septiembre de 2023. En la figura 5.10 se observa la gráfica de la serie de tiempo de Tampico y en la figura 5.11 de Xalapa. Se revisaron estas series de tiempo para no tener datos faltantes y también se les hizo una prueba para verificar outliers con un boxplot y posteriormente para corroborar esto se revisó con z-score. Se tomó en cuenta un umbral de 3 desviaciones estándar para los outliers.

Se debe recalcar que esta experimentación se hizo dos veces, la primera fue con los datos en crudo y la segunda fue con los datos normalizados con z-score, resultando con menor error la segunda prueba siendo la que se muestra en esta sección.

Debido a su buen desempeño en pruebas anteriores, en esta parte se utilizan los métodos Holt-Winters aditivo y SARIMA para realizar la predicción a mediano plazo de las ciudades mencionadas.

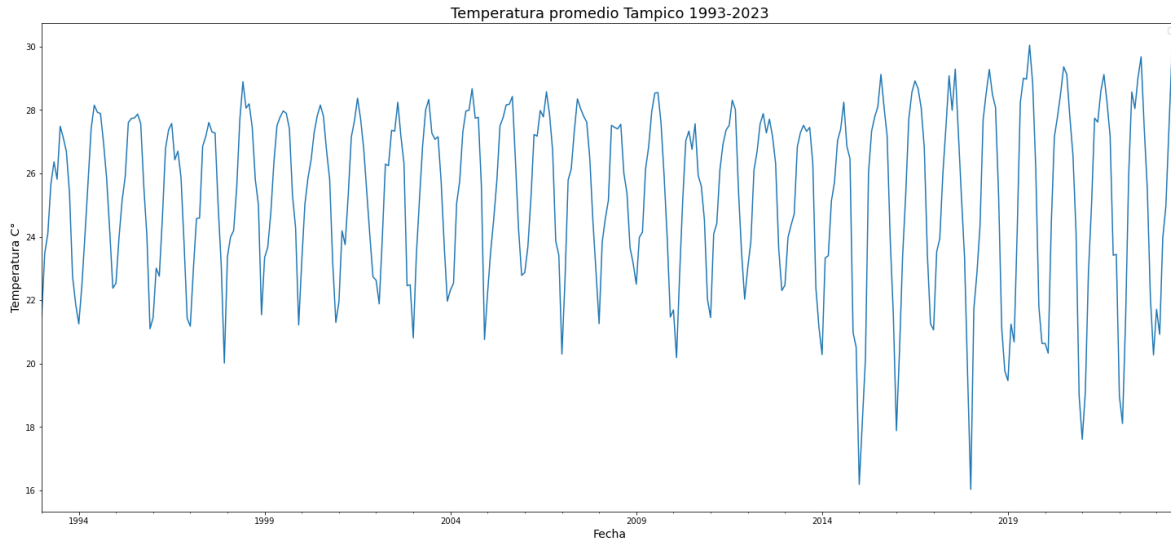


Figura 5.10. Datos históricos de Tampico 1993-2023.

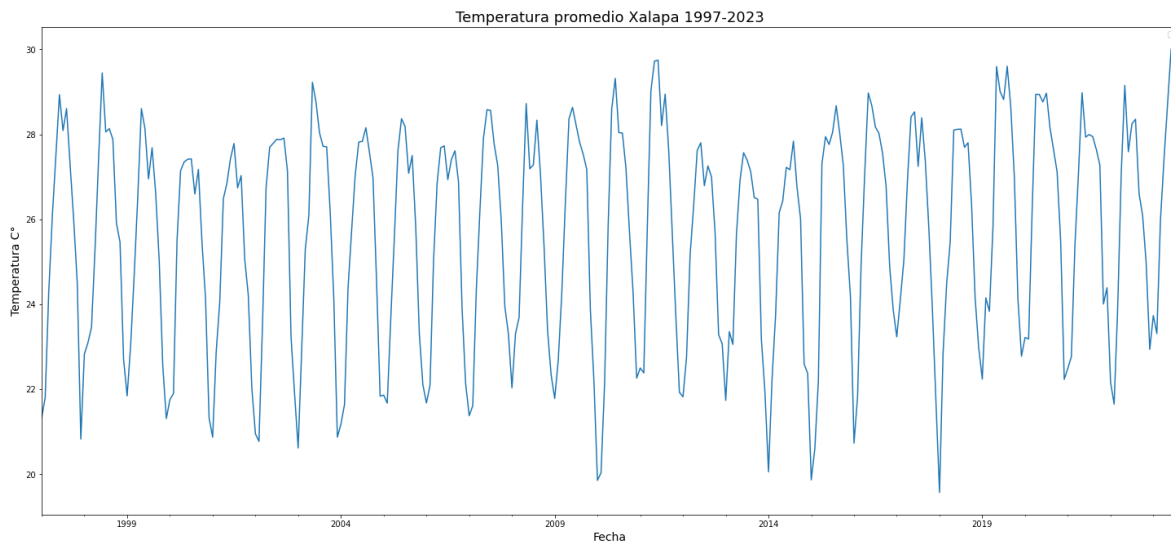


Figura 5.11. Datos históricos de Xalapa 1993-2023.

5.3.1 Prueba Tampico

En la ciudad de Tampico se hace una experimentación con los dos métodos mencionados: Holt-Winters y SARIMA. Para esta serie de tiempo se tomó para la etapa de entrenamiento 70% (226 datos) validación 15% (47 datos) y prueba 15% (47 datos de 10/2019 a 08/2023).

Holt-Winters

En la etapa para el pronóstico de prueba con el método de Holt-Winters aditivo se utilizaron los siguientes ajustes; $\alpha = 0.3$ un $\beta = 0.0$ y un $\gamma = 0.2$., y una estacionalidad correspondiente a los meses del año obtenido del entrenamiento. En la figura 5.12 se puede observar una gráfica donde se comparan los datos reales (azul) con los datos pronosticados (rojo) de la etapa de prueba. Se puede observar que el método pudo seguir los datos reales debido al ajuste de tendencia y estacionalidad aditiva obteniendo un MAPE de 3.7% y un RMSE de 1.11.

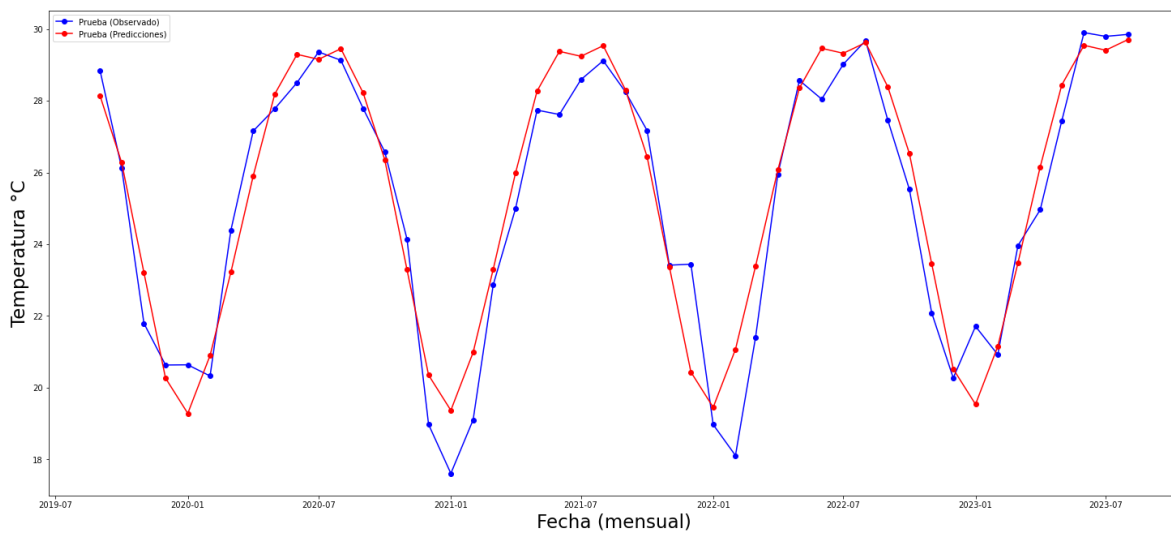


Figura 5.12. Prueba pronóstico Holt-Winters Vs valores reales en Tampico de 2019-2023

SARIMA

En esta etapa para el pronóstico de prueba se utilizó SARIMA (2,0,0), (1,0,1,12). Estos parámetros se ajustaron con auto arima mediante una búsqueda exhaustiva con los datos de entrenamiento. En la figura 5.13 se puede observar una gráfica donde se comparan los datos reales (azul) con los datos pronosticados (rojo) de la etapa de prueba. En la figura se observa que lo estimado sigue la tendencia, estacionalidad y también se nota que es cíclico debido a que son varios años. Con este método se obtiene un MAPE de 3.64% y un RMSE de 1.16.

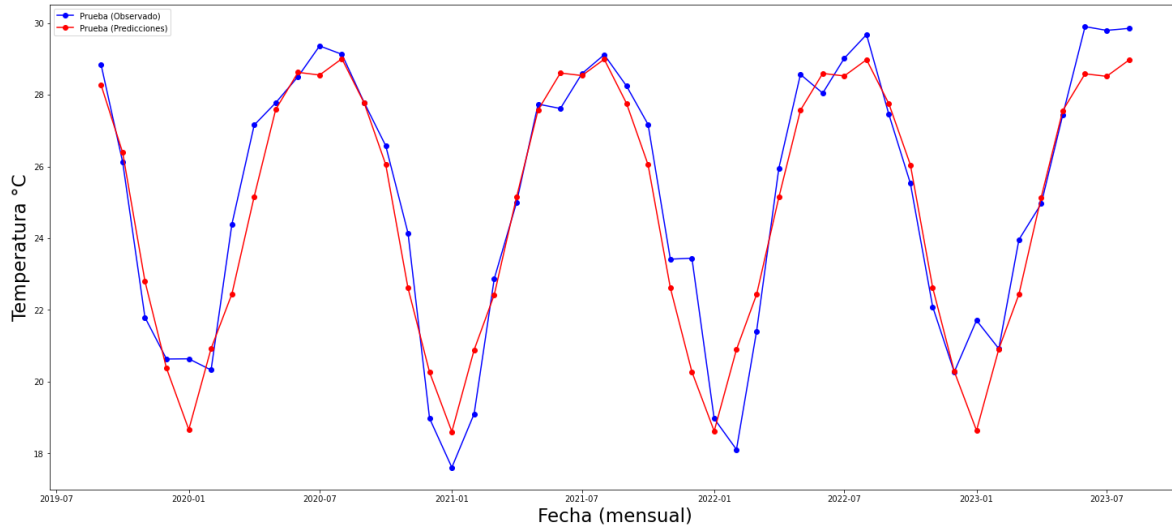


Figura 5.13. Prueba pronóstico SARIMA Vs real en Tampico 2019-2023.

5.3.2 Prueba Xalapa

Para esta ciudad se toma al igual que la anterior para la etapa de entrenamiento 70% (226 datos) validación 15% (47 datos) y prueba 15% (47 datos de 10/2019 a 08/2023).

Se hace la predicción con los modelos de Holt-Winters aditivo y SARIMA.

Holt-Winters

En la etapa para el pronóstico de prueba con el método de Holt-Winters aditivo se utilizaron los siguientes ajustes; $\alpha = 0.3$ un $\beta = 0.2$ y un $\gamma = 0.0.$, y una estacionalidad correspondiente a los meses del año obtenido del entrenamiento. En la figura 5.14 se puede observar una gráfica donde se comparan los datos reales (azul) con los datos pronosticados (rojo) de la etapa de prueba. Se puede observar que el método pudo seguir los datos reales obteniendo un MAPE de 2.7% y un RMSE de 0.87

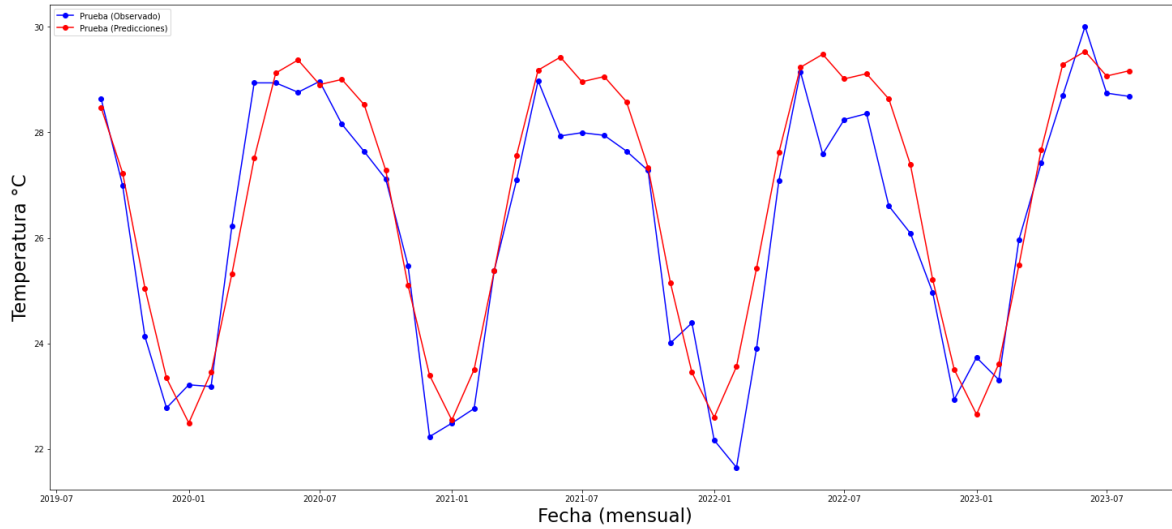


Figura 5.14. Prueba de pronóstico Holt-Winters Vs real en Xalapa 2019-2023.

SARIMA

En esta parte para el pronóstico de prueba se utilizó SARIMA (1,0,2), (2,0,2,12). Estos parámetros se ajustaron con auto arima mediante una búsqueda exhaustiva con los datos de entrenamiento. En la figura 5.15 se puede observar una gráfica donde se comparan los datos reales (azul) con los datos pronosticados (rojo) de la etapa de prueba. Con este método se obtiene un MAPE de 2.96% y un RMSE de 0.99.

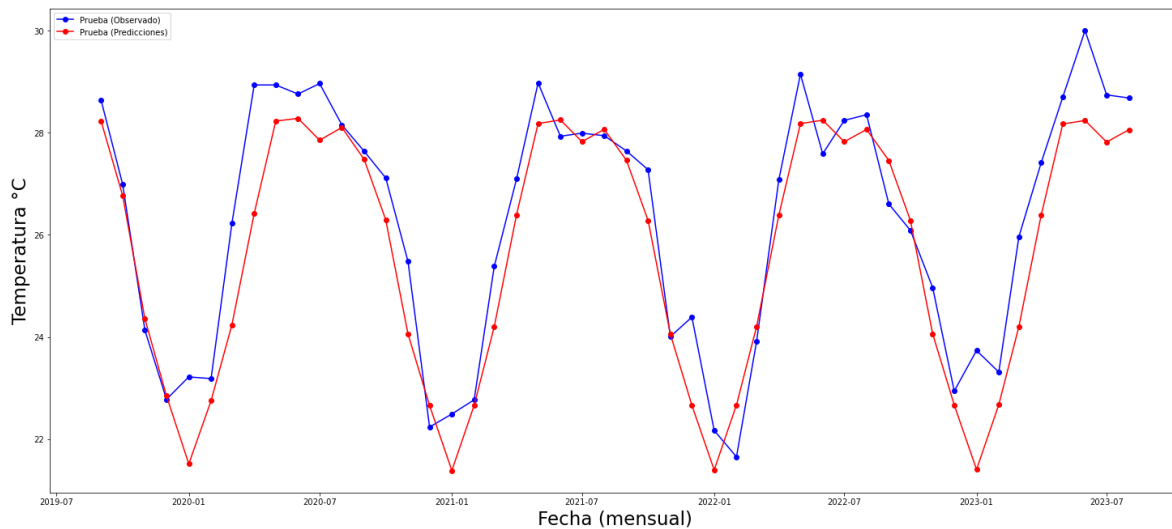


Figura 5.15. Prueba de pronóstico SARIMA Vs real en Xalapa 2019-2023.

5.3.3 Predicción de temperatura de 2023 a 2027 en Tampico y Xalapa

En la tabla 5.16 se muestra la comparación de las predicciones de temperatura en la etapa de prueba en las ciudades de Xalapa y Tampico con los dos métodos utilizados. Estas predicciones se comparan con las dos métricas utilizadas MAPE y RMSE. Como se observa, para Tampico el modelo que obtuvo un menor MAPE fue SARIMA y para Xalapa fue Holt-Winters.

Tabla 5.16. Comparación de MAPE y RMSE Tampico y Xalapa 1993-2023.

	Tampico		Xalapa	
	Holt-Winters add	SARIMA	Holt-Winters add	SARIMA
MAPE %	3.7	3.64	2.7	2.96
RMSE	1.11	1.16	0.87	0.99
MSE	1.24	1.34	0.76	0.97

Los métodos con menor error para cada ciudad se utilizan para hacer una predicción a futuro desde el último mes registrado hasta el momento que se hizo esta experimentación siendo agosto 2023. La predicción se hace desde septiembre de 2023 hasta agosto de 2027 para cada ciudad.

En la figura 5.17 se muestra una estimación de la temperatura mensual promedio en Tampico y en la 5.18 de Xalapa. Estas van de septiembre de 2023 a agosto de 2027 utilizando SARIMA y HW respectivamente. Esperando tener un error MAPE bajo como sucedió en la prueba.

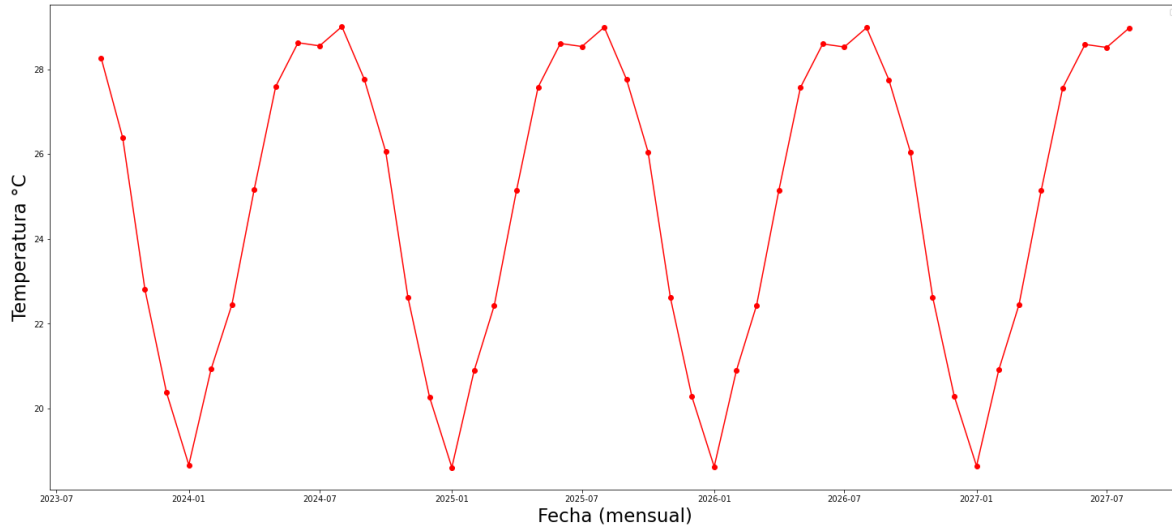


Figura 5.17. Pronóstico de temperatura mensual SARIMA Tampico 2023-2027.

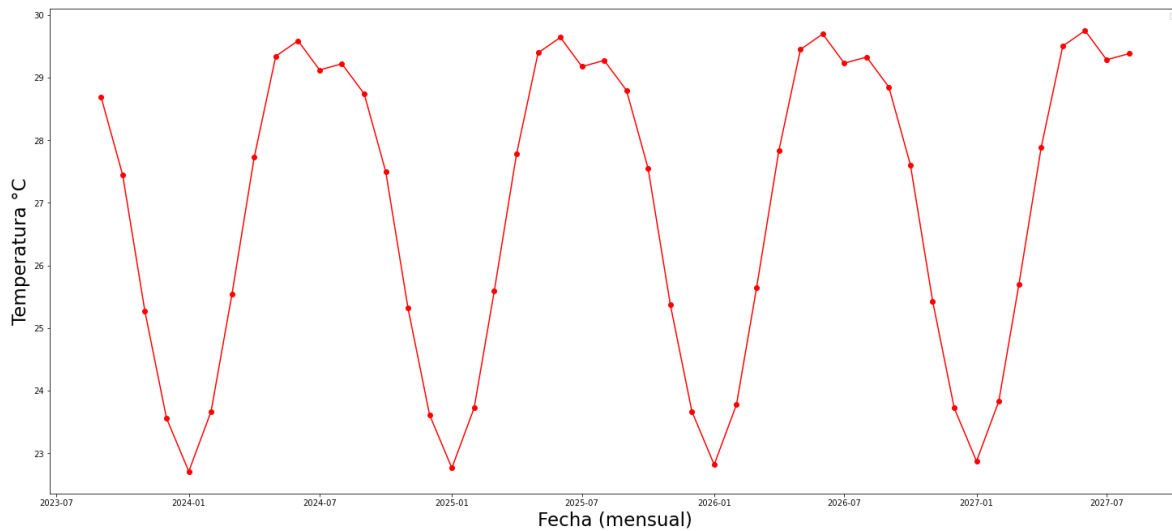


Figura 5.18. Pronóstico de temperatura mensual Holt-Winters Xalapa 2023-2027.

5.4 Prueba de hipótesis

Se hace una prueba para determinar si las series de tiempo son normales. En este caso se utiliza la prueba de Shapiro-Wilk y en ambas series de tiempo se comprueba que no son normales las series ya que los valores no tienen una distribución gaussiana, estos no se concentran alrededor de la media y no siguen una forma de campana. Después al comprobar

esto se hace la prueba de homogeneidad de varianzas de Levene y se comprueba que no hay homogeneidad en ambas series de tiempo ya que no hay suficiente evidencia de que la media, varianza y la estructura temporal permanezcan constantes a lo largo de toda la serie. Es por esto que se elige la prueba de hipótesis de Wilcoxon para determinar si el error MAPE de ambas predicciones son significativamente una mejor que la otra, como se puede ver a simple vista viendo el promedio pareciendo que si lo son.

Para esta prueba se tienen las siguientes hipótesis:

H_0 : no hay diferencia significativa entre los errores de predicción de SARIMA y Holt-Winters.

H_1 : Hay una diferencia significativa entre los errores de predicción de SARIMA y Holt-Winters.

Al realizarse esta prueba para ambas ciudades se llegó a la siguiente conclusión ya que el valor p fue más alto que el umbral alfa: No se puede rechazar la hipótesis nula. No hay suficiente evidencia para afirmar que hay una diferencia significativa entre los errores de predicción de SARIMA y Holt-Winters.

5.5 Comparación de predicción de temperatura Vs su promedio real

En la tabla 5.19 se muestran los promedios de temperatura de los años predichos con SARIMA para Tampico y para Xalapa con Suavizamiento Exponencial Triple aditivo.

Es bien conocido que el año 2023 fue el año más caluroso registrado desde 1950, lo cual se ha sabido al comparar la temperatura alcanzada en el año 2023 con respecto a la temperatura promedio de los años 1951 a 1980. Se puede realizar una comparación de las temperaturas estimadas para el año 2024 con respecto al el promedio de temperatura que se tuvo entre los años 1997 a 2010. Esa comparación se presenta en la Tabla 5.19 y, se puede observar el incremento de la temperatura promedio de los años que se han estimado 2024, 2025 y 2026 Vs 1997-2010 es para Tampico y Xalapa de 0.18°C y 1.48°C, respectivamente. por tanto, 2024 será el año más caluroso pues se sabe que 2023 fue el de la temperatura más alta respecto al promedio de 1951 a 1980.

|

Tabla 5.19. Temperatura media (1997-2010) en Tampico-Xalapa Vs estimada 2024-2026.

Año	Temperatura promedio	
	Tampico	Xalapa
1997-2010	24.61	25.41
2024	24.80	26.84
2025	24.79	26.89
2026	24.79	26.95
Promedio 2024-2026	24.79	26.89

Desde luego en la investigación presentada hubiese sido conveniente también realizar el cálculo del incremento de temperatura estimadas de 2024 a 2026 y más años hacia adelante con respecto al promedio de temperatura de 1951 a 1980, pero se requería un proceso de investigación y de procesamiento adicional lo cual se deja para trabajos posteriores. Por lo pronto se puede observar que la estimación de temperatura estimada para 2024 es superior al valor registrado para el año 2023, y dada la precisión de nuestro método de pronóstico podemos esperar que así será, y por tanto, el año 2024 será aún más caluroso que el mas caluroso registrado en las últimas décadas.

6 Conclusiones

Se obtuvo una métrica de error de pronóstico cuyos valores son del mismo orden de magnitud los obtenidos por los métodos del estado del arte, ya que para la temperatura promedio ambiental no se obtienen predicciones muy precisas habitualmente, Sin embargo, un RMSE de 0.87 demuestra un buen desempeño. Se hizo una predicción de la temperatura para los años 2024 a 2027, esperando que de acuerdo a la prueba también se tengan promedios de error bajos que se pueden comparar con los del estado del arte.

Para la serie de tiempo de temperatura promedio mensual en Xalapa, se obtuvo un pronóstico con menor promedio de error con el Suavizamiento exponencial triple, al contrario de la ciudad de Tampico, donde se obtuvo un MAPE más bajo con el método SARIMA. Esta diferencia de acuerdo a la prueba de hipótesis realizada no es muy significativa, ya que estadísticamente no hay mucha diferencia de estos errores. Este trabajo nos demuestra que el suavizamiento exponencial y SARIMA, métodos clásicos, son buenos y confiables para la predicción de temperatura ambiental a mediano plazo. Estos podrán ser utilizados para futuras predicciones y análisis del cambio climático en otras ciudades con características parecidas a Tampico y Xalapa, y se espera que obtengan resultados muy buenos. Esto para poder tomar medidas preventivas y planificar estrategias para combatir al cambio climático.

6.1 Recomendaciones

Como trabajos futuros se recomienda continuar probando con otros métodos de predicción como los de Machine Learning. Ampliar las regiones de México de estudio, agregar ciudades parecidas y recopilar datos de otras variables que influyen en el cambio climático para comparar contra métodos multivariados.

Bibliografía

- [1] B. Viguera, M. R. Martínez Rodríguez, C. I. Donatti, C. A. Harvey y F. Alpizar, El clima, el cambio climático, la vulnerabilidad y acciones contra el cambio climático: Conceptos básicos, Turrialba, Costa Rica: Proyecto CASCADA, 2017, p. 44. https://www.conservation.org/docs/default-source/publication-pdfs/cascade_modulo-1-el-clima-el-cambio-climatico-la-vulnerabilidad-y-acciones-contr-el-cambio-climatico.pdf?Status=Master&sfvrsn=576b396d_2
- [2] Masson-Delmotte V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor y T. Waterfield; Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways; IPCC, 2018.
- [3] José Clemente Rueda Abad, ¿Aún estamos a tiempo para el 1.5°C? Voces y Visiones sobre el Reporte Especial del IPCC, Programa de Investigación en Cambio Climático de UNAM Universidad Nacional Autónoma De México, 2019, <https://www.pincc.unam.mx/publicaciones/aun-estamos-a-tiempo-para-el-1-5c-voces-y-visiones-sobre-el-reporte-especial-del-ipcc/>
- [4] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, The M4 Competition: 100,000 time series and 61674 forecasting methods, *Int. J. Forecast.*, vol. 36, no. 1, 2020, pp. 54–74, 2020, DOI:675 10.1016/j.ijforecast.2019.04.014. 2019
- [5] Winters P, “Forecasting Sales by Exponentially Weighted Moving Averages,” *Management Science*, vol. 6, no. 3, pp. 324–342, 1960. <http://libsta28.lib.cam.ac.uk:2114/stable/2627346>
- [6] Universidad Nacional Autónoma de México/ Programa de Investigación en cambio climático, Reporte mexicano de cambio climático. Grupo 1. Bases científicas, modelos y modelación, vol. 1, México, D.F.: Alebrije diseño, 2015, p. 293. ISBN Volumen: 978-607-02-7522-7
- [7] T. Dietz, R. L. Shwom y C. T. Whitley, Climate Change and Society, *Annual reviews*, vol. 46, n° 1, pp. 135-158, 2020.

- [8] F. Giorgi, Climate change prediction, *Climatic change Journal*, vol. 73, n° 3, pp. 239-265, Springer, 2005. DOI: 10.1007/s10584-005-6857-4
- [9] V. Magaña y E. Caetano, Pronóstico climático estacional regionalizado para la República Mexicana como elemento para la reducción de riesgo, para la identificación de opciones de adaptación al cambio climático y para la alimentación del sistema: cambio climático por estado y por sector, Universidad Autónoma de México, México, 2007.
- [10] Siwek, K., & Osowski, S., *Data mining methods for prediction of air pollution*, International Journal of Applied Mathematics and Computer Science, 26(2), 467–478. DOI:10.1515/amcs-2016-0033 2016
- [11] CICESE, CLICOM. <http://clicom-mex.cicese.mx/>. [Último acceso: 24 mayo 2022].
- [12] Y.Radhika, M.ShashiAtmospheric, Temperature Prediction using Support Vector Machines, International Journal of Computer Theory and Engineering, Vol. 1, No. 1, 1793-8201. DOI: 10.7763/IJCTE. 2009.V1.9 2009
- [13] Huang, G., Zhu, H., Zhang, J., & Liu, B., Analysis of the Characteristics of Climate Change in the Ecologically Vulnerable Area of the Mu Us Dune Field under the Background of Global Warming, *Remote Sensing*, 13(4), 627. DOI:10.3390/rs13040627 2021
- [14] F. Behzadi, . A. Wasti, S. Haque Rahat, . J. N. Tracy y P. A. Ray, Analysis of the climate change signal in Mexico City given disagreeing data sources and scattered projections, *Journal of Hydrology: Regional Studies*, vol. 27, p. 100662, 2020.
- [15] J. Brownlee, Machine Learning Mastery, 6 Agosto 2018. <https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/>. [Último acceso: 9 Marzo 2022].
- [16] O. Folorunsho, Application of Data Mining Techniques in Weather Prediction and Climate Change Studies, I.J. Information Engineering and Electronic Business Published Online February 2012 in MECS (<http://www.mecs-press.org/>), 1, 51-59 DOI: 10.5815/ijieeb.2012.01.07. 2012
- [17] Ponce, Mirna P., Métodos híbridos para pronóstico de series financieras basadas en suavizamiento exponencial, Tesis de Doctorado, Tecnológico Nacional de México/Instituto Tecnológico de Ciudad Madero, 2021.

- [18] F. Ballester, Air Pollution, Climate Change, *Rev Esp Salud Pública*, vol. 79, n° 2, pp. 159-175, 2005
https://www.scielosp.org/article/ssm/content/raw/?resource_ssm_path=/media/assets/resp/v79n2/v79n2a05.pdf
- [19] J. R. Porter y M. A. Semenov, Crop responses to climatic variation, *Philosophical transactions of the royal society*, vol. 360, p. 2021–2035, 2005. doi:10.1098/rstb.2005.1752
- [20] J. E. Hanke and D. W. Wichern, *Pronósticos en los negocios*, Prentice Hall, 2010.
- [21] R. G. Brown, *Statistical forecasting for inventory control*, McGraw-Hill, New York, p. 1959, 1959.
- [22] C. C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, *International Journal of Forecasting*, vol. 20, no. 1, pp. 5–10, 2004.
- [23] J. G. De Gooijer, R. J. Hyndman, J. G. D. Gooijer, and R. J. Hyndman, 25 Years of Time Series Forecasting, p. 2006, 2006. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0169207006000021> doi=10.1.1.154. 9227
- [24] Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. OTexts
- [25] M. Farsi, D. Hosahalli, B. Manjunatha, I. Gad, E.-S. Atlam, A. Ahmed, G. Elmarhomy, M. Elmarhomy y O. A. Ghoneim, Parallel genetic algorithms for optimizing the SARIMA model for better forecasting of the NCDC weather data, *Alexandria Engineering Journal*, vol. 60, n° 1, pp. 1299-1316, 2021.
- [26] B. Liu, X. Yu, J. Chen y Q. Wang, Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model, *Atmospheric Pollution Research*, vol. 12, 2021. DOI: 10.1016/j.apr.2021.101144
- [27] L. Benites, Prueba de Shapiro-Wilk: definición, cómo ejecutarla en SPSS, *statologos*, 18 marzo 2022. [En línea]. Available: <https://statologos.com/prueba-de-shapiro-wilk/>. [Último acceso: 30 Noviembre 2023].
- [28] Prueba de Levene, *DATAtab*, 2023. [En línea]. Available: <https://datatab.es/tutorial/levene-test>. [Último acceso: 30 noviembre 2023].

- [29] N. Shahid, M. Ali Shah, A. Khan, C. Maple y G. Jeon, Towards greener smart cities and road traffic forecasting using air pollution data, *Sustainable Cities and Society*, vol. 72, 2021. DOI: 10.1016/j.scs.2021.103062
- [30] M. Beroho, H. Briac, R.E: Halimi, A, Ouallali, I. Boulahfa, R. Mrabet, F. Kebedi, K. Aboumaria, Analysis and prediction of climate forecasts in Northern Morocco, *Helyon* vol 6, 2020, Elsevier, Science direct, DOI: 10.1016/j.heliyon.2020.e05094
- [31] J. Frausto, L. Moya, J. González B., G. Castilla, M Ponce Maldonado, FCTA: A Forecasting Combined methodology with a Threshold Accepting Approach, *Mathematical Problems in Engineering*, Hindawi, Wiley Editors, 2022.
- [32] J. L. Ibarra Montoya y F. M. Huerta Martínez, Cambio climático y predicción de incendios al 2050 en el Bosque La Primavera, Jalisco, *Revista mexicana de ciencias forestales*, vol. 7, n° 37, pp. 39-50, 2016. DOI:10.29298/rmcf.v7i37.50
- [33] S. B. Phillips, V. P. Aneja, D. Kang y S. P. Arya, Maximum entropy modeling of species geographic distributions, *Ecological Modelling*, vol. 190, n° 3-4, pp. 231-259, 2006. DOI: 10.1016/J.ECOLMODEL.2005.03.026
- [34] A. Qin, B. Liu, Q. Guo, R. W. Bussmann, F. Ma, Z. Jian, G. Xu y S. Pei, Maxent modeling for predicting impacts of climate change on the potential distribution of *Thuja sutchuenensis* Franch., an extremely endangered conifer from southwestern China, *Global Ecology and Conservation*, vol. 10, n° 4, pp. 139-146, 2017. DOI: 10.1016/J.GECCO.2017.02.004
- [35] K. K. Rani Samal, A. Kumar Panda, K. Sathya Babu y S. Kumar Das, Multi-output TCN autoencoder for long-term pollution forecasting for multiple sites, *Urban Climate*, vol. 39, 2021. DOI: 10.1016/j.uclim.2021.100943
- [36] F. Behzadi, A. Wasti, S. Haque Rahat, J. N. Tracy y P. A. Ray, Analysis of the climate change signal in Mexico City given disagreeing data sources and scattered projections, *Journal of Hydrology: Regional Studies*, vol. 27, 2020. DOI: 10.1016/j.ejrh.2019.100662
- [37] B. Liu, X. Yu, J. Chen y Q. Wang, Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model, *Atmospheric Pollution Research*, vol. 12, 2021. DOI: 10.1016/j.apr.2021.101144