



Universidad Autónoma de Querétaro

Facultad de Informática

Doctorado en Ciencias de la Computación

Algoritmo para pronóstico de precipitaciones en tiempo real basado en el modelo predictivo CRHUDA.

TESIS

Que como parte de los requisitos para obtener el grado de
Doctor en Ciencias de la Computación

Presenta:

M.C.C. Martín Muñoz Mandujano

Dirigido por:

Dr. Martín Alfonso Gutiérrez López

SINODALES

Dr. Martín Alfonso Gutiérrez López
Presidente

Dr. Hugo Jiménez Hernández
Secretario

Dr. Ricardo Chaparro Sánchez
Vocal

Dr. Juan Pablo Molina Aguilar
Suplente

Dr. José Alfredo Acuña García
Suplente

Centro Universitario, Querétaro, Qro.
Abril 2024
México.



Dirección General de Bibliotecas y Servicios Digitales
de Información



Modelo inteligente predictivo basado en redes
neuronales artificiales para el pronóstico de
momento e intensidad de precipitaciones.

por

Martín Muñoz Mandujano

se distribuye bajo una [Licencia Creative Commons
Atribución-NoComercial-SinDerivadas 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Clave RI: IFMAC-246541

*Esta tesis esta dedicada a mis papás Martin y Vero, a mi
hermana Gaby*

y a toda mi familia y amigos.

Resumen

El pronóstico de precipitaciones ha sido una preocupación constante para la humanidad desde la antigüedad, dada su importancia en la vida diaria, tanto para la gestión de recursos hídricos como para la prevención de catástrofes naturales. La ausencia de un mecanismo confiable y preciso para predecir el inicio de las precipitaciones ha motivado la presente investigación, cuyo objetivo es desarrollar un algoritmo hidroeinformático para el pronóstico de precipitaciones, orientado a su futura implementación en sistemas de alerta temprana. El área de estudio se centra en la zona metropolitana del Estado de Querétaro, y se fundamenta en el modelo de pronóstico CRHUDA.

Utilizando una metodología de investigación basada en diseño, se generaron múltiples iteraciones en la fase de desarrollo y pruebas, explorando diversas soluciones desde la comprobación y análisis de sensibilidad del modelo CRHUDA, hasta el desarrollo de una red neuronal artificial LSTM para predecir el inicio de la precipitación. Este proceso culminó en la conclusión de la necesidad de crear una arquitectura completa que, siguiendo principios de arquitectura de software, permitiera generar una solución robusta, mantenible y capaz de ofrecer pronósticos más precisos que los obtenidos mediante un único modelo.

Palabras clave: Arquitectura de Software, Hidroeinformática, Pronóstico, Redes Neuronales Artificiales.

Abstract

The forecast of precipitation has been a constant concern for humanity since ancient times, given its importance in daily life, both for water resource management and for the prevention of natural disasters. The absence of a reliable and accurate mechanism to predict the onset of precipitation has motivated the present research, whose objective is to develop a hydroinformatic algorithm for precipitation forecasting, aimed at its future implementation in early warning systems. The study area focuses on the metropolitan area of the State of Querétaro, and is based on the CRHUDA forecast model.

Using a design-based research methodology, multiple iterations were generated in the development and testing phase, exploring various solutions from the verification and sensitivity analysis of the CRHUDA model, to the development of an LSTM artificial neural network to predict the onset of precipitation. This process culminated in the conclusion of the need to create a complete architecture that, following software architecture principles, would allow generating a robust, maintainable solution capable of offering more accurate forecasts than those obtained through a single model.

Keywords: Software Architecture, Hydroinformatics, Forecasting, Artificial Neural Networks.

Agradecimientos

Primeramente me gustaría extender mi agradecimiento a la Universidad Autónoma de Querétaro por abrirme las puertas y darme la oportunidad de continuar mi desarrollo académico y profesional, así como a los profesores que me brindaron sus conocimientos, apoyo y acompañamiento a lo largo de todo este proceso.

Agradezco a mi director de tesis el Dr. Alfonso Gutiérrez López quien a lo largo de los años que llevo de conocerlo, me ha brindado su apoyo y confianza para poder llevar acabo esta investigación, siendo para mi una inspiración en cuanto al tipo de investigador que deseo ser y sabiendo que es y será un gran colega y amigo.

Me gustaría expresar de igual forma mi agradecimiento a la Dra. Gabriela Xicoténcatl Ramírez, directora de la Facultad de Informática, quien a lo largo de mi vida profesional a confiado en mis capacidades, apoyándome y brindándome retos para poder superarme y seguir creciendo.

Desde mi corazón y de forma muy especial, me gustaría agradecer de forma prioritaria a mi familia. Primeramente a mi papá Martín Muñoz Paredes y mi mamá Verónica Mandujano Rodríguez, ya que confiaron y me apoyaron a lo largo de este gran reto, independientemente si me desvelaba, si debía quedarme los fines para trabajar e incluso llendo a visitarme con el almuerzo los domingos que no podía ir a verlos por trabajar en los pendientes, siempre han sido un ejemplo para mi y pilares fundamentales en mi vida. De igual forma otro pilar fundamental y apoyo

emocional es mi hermana Verónica Gabriela Muñoz Mandujano (Gaby para mas fácil), pues nos apoyamos mientras vivimos juntos aunque a mitad del camino tuvo que mudarse para trabajar en su propia meta, siempre a sido mi confidente y esa persona a la que le puedo contar todos mis planes, ideas y estrategias a futuro, a veces apoyándome y otras dándome el avión, se que siempre esta para mi. Le agradezco enormemente a toda mi familia, a todos mis tíos y tías, primos y primas que siempre me han dado su apoyo y visto mi crecimiento y de forma muy especial a mis abuelitos Martín, Juana, Gloria y Willebaldo, aunque no puedan estar 3 de ellos para ver este logro tan importante, los 4 han formado parte fundamental de mi crecimiento como persona y se que están sumamente orgullosos y felices por este gran logro.

En mi corazon existe una segunda familia en la cual he pertenecido desde el lejano 2014/2015, me refiero a todos los integrantes de Centro de Desarrollo quienes han sido parte fundamental en mi estancia durante el doctorado, en mi formación como profesionista e incluso como persona. Muchas gracias a todos y cada uno de ustedes.

Por supuesto que de manera particular me gustaría agradecer a 4 integrantes fundamentales de centro de desarrollo, le agradezco enormemente a Alex, Diego, Mau y Adrian quienes me han apoyado de una forma especial cada uno de ellos. Alex quien desde licenciatura me a brindado su apoyo y confianza, apoyándome en todo lo que se me ocurre y guiándome, siendo una persona que me permite poner los pies sobre la tierra y analizar mis opciones. Diego, un gran amigo y ejemplo de como ser docente, una persona que siempre se preocupa por los demás, muchas gracias por todo tu apoyo. Mau, un amigo muy importante para mi, y alguien a quien puedo decir que es mi hermano de otra madre, ya que siempre me ha apoyado, escuchando cada locura que se me ocurre y aconsejándome ya sea para hacerla realidad o para descartarla por completo. Finalmente Adrian, eres un gran amigo y que últimamente nos hemos unido mas, sin duda hoy en día eres de las pocas personas que me conocen al 100 como soy fuera del entorno laboral y se que puedo

contar contigo para lo que necesite.

Índice general

Resumen	I
Abstract	II
Agradecimientos	III
Índice General	VI
Índice de Figuras	IX
Índice de Tablas	XII
1. Introducción	1
1.1. Descripción del Problema	2
1.2. Hipótesis	3
1.3. Justificación	4
1.4. Objetivos	5
1.4.1. Objetivo General	5
1.4.2. Objetivos Específicos	5
1.5. Alcances y Limitaciones	6
1.6. Aportación y Pertinencia	6
2. Estado del Arte	8

2.1.	Mecanismos para el pronóstico de series de tiempo.	10
2.1.1.	Modelos de Aprendizaje Automático y Deep Learning	11
2.1.2.	Modelos ARMA	13
2.1.3.	Modelos ARIMA	14
2.2.	Mecanismos actuales para el pronóstico de precipitaciones.	14
2.3.	Modelo CRHUDA.	17
2.4.	Técnicas de pronóstico dentro de la inteligencia artificial	21
2.4.1.	Redes Neuronales Artificiales	21
3.	Metodología y desarrollo	24
3.1.	Definición y análisis de problema	26
3.1.1.	Análisis del estado del arte respecto al pronóstico de precipitaciones.	26
3.1.2.	Adquisición de datos climatológicos	26
3.1.3.	Análisis y preprocesamiento de datos	28
3.1.4.	Relación Clausius-Clapeyron y su influencia sobre las precipitaciones	32
4.	Análisis de sensibilidad	36
4.1.	Descripción de series del modelo CRHUDA	36
4.1.1.	Modelos autorregresivos de promedios móviles ARMA(p,q)	38
4.1.2.	Modelos estocásticos multivariados	38
4.1.3.	Autocorrelación como validación del modelo CRHUDA	39
4.2.	Análisis de sensibilidad	40
4.3.	Implementación y Validación	42
4.3.1.	Presentación del modelo CRHUDA	42
4.3.2.	Cálculo de correlogramas	45
4.3.3.	Cálculo del factor de escala	47

5. Algoritmo de pronostico mediante modelo regresivo de tipo LSTM	49
5.1. Diseño y Desarrollo	49
5.1.1. Selección de entorno de desarrollo	49
5.1.2. Herramientas de diseño de modelo	51
5.2. Algoritmo de pronóstico por medio de RNA de tipo LSTM	53
5.2.1. Resultados de Modelo	56
6. Diseño de arquitectura mediante implementación de principios arquitectonicos de software.	64
6.1. Rediseño del algoritmo de pronóstico mediante principios arquitectónicos	66
6.1.1. Resultado de Rediseño	67
7. Discusión	70
7.1. Discusión	70
7.1.1. Uso de modelo CRHUDA y análisis de sensibilidad	70
7.1.2. Algoritmo de Redes Neuronales Artificiales de tipo LSTM	71
7.1.3. Arquitectura para pronostico de precipitaciones	72
7.2. Publicaciones	73
7.3. Trabajo Futuro y lineas de investigación	74
8. Conclusiones	75
Referencias	78

Índice de figuras

2.1. <i>Acumulado de precipitación en 24 h desde el 4 de noviembre del 2018 al 28 de febrero del 2019. Obtenido de Fuentes et al. (2020)</i>	15
2.2. <i>Esquema conceptual del funcionamiento del modelo CRHUDA (CRossingHUmidity, Dewpoint, and Atmospheric pressure) (Gutierrez-Lopez et al., 2019b; Muñoz-Mandujano et al., 2022).</i>	19
2.3. <i>Resultados de la aplicación del modelo CRHUDA a los datos del Centro Histórico en Querétaro, el 16 de agosto de 2014. Elaboración propia (ver Gutierrez-Lopez et al., 2019b; Muñoz-Mandujano et al., 2022, para mas detalles).</i>	20
2.4. <i>Resultados de la aplicación del modelo CRHUDA a los datos de Atlacomulco en el Estado de México, el 23 de abril de 2018. Elaboración propia (ver Gutierrez-Lopez et al., 2019b; Muñoz-Mandujano et al., 2022, para mas detalles).</i>	20
2.5. <i>Resultados de la aplicación del modelo CRHUDA a los datos de Dzilbachen en Campeche, el 22 de febrero de 2018. Elaboración propia (ver Gutierrez-Lopez et al., 2019b; Muñoz-Mandujano et al., 2022, para mas detalles).</i>	20
3.1. <i>Actividades principales de los científicos de datos. Basado en Press, 2016</i>	28

3.2. Variaciones de los encabezados dentro de las fuentes de datos proporcionada por CONAGUA de las EMAS.	30
4.1. Readaptacion del esquema conceptual del modelo CRHUDA(CRrossingHUmidity, Dewpoint, and Atmospheric pressure). (Muñoz-Mandujano et al., 2024). 37	37
4.2. Resultados de la aplicación del modelo CRHUDA en la estación Candiles el 24-25 de junio de 2013.	43
4.3. Resultados de la aplicación del modelo CRHUDA en la estación Cimatarío el 16 de agosto de 2014 (lámina máxima 15.4 mm en 1 h; intensidad máxima 41.2 mm/h; período de retorno de 32 años).	44
4.4. Resultados de la aplicación del modelo CRHUDA a los datos del UAQ Cerro Campanas del 16 de agosto de 2014 (lámina máxima 55.2 mm en 1 h; intensidad máxima 63.8 mm/h; período de retorno de 53 años). 44	44
4.5. Correlograma para las series de tiempo de presión atmosférica. Estación Juriquilla 17 Septiembre 2017 (a) 6 horas (c) 12 horas y (d) 24 horas	46
4.6. Correlograma para las series de tiempo de humedad. Estación Juriquilla 17 Septiembre 2017 (a) 6 horas (c) 12 horas y (d) 24 horas . . .	46
4.7. Correlograma para las series de tiempo de punto de rocío. Estación Juriquilla 17 Septiembre 2017 (a) 6 horas (c) 12 horas y (d) 24 horas .	46
5.1. Representación estilo FCNN de la red neuronal artificial.	54
5.2. Gráfica de Correlación y correlación parcial de la humedad.	57
5.3. Gráfica de Correlación y correlación parcial del punto de rocío.	58
5.4. Gráfica de Correlación y correlación parcial de la presión atmosférica. 58	58
5.5. Gráfica de resultados de modelo ARMA para pronostico de punto de rocío.	61
5.6. Gráfica de resultados de modelo ARMA para pronostico de Humedad. 62	62
5.7. Gráfica de resultados de modelo ARMA para pronostico de Presión Atmosférica.	63

6.1. *Diagrama conceptual de la arquitectura para pronostico de lluvias.* . . 69

Índice de tablas

2.1. Características fundamentales de una red neuronal artificial.	21
4.1. Resultados del análisis de sensibilidad y del factor de escala para el modelo CRHUDA aplicado a datos de diversas tormentas extremas registradas en el mes de septiembre. Estación Milenio (Muñoz-Mandujano et al., 2024)	48
5.1. Resultados del análisis comparativo de las librerías de aprendizaje automático	50
5.2. Resultados de las diversas versiones del modelo con la Arquitectura 1 - LSTM Regresión.	56
5.3. <i>Resultados de modelo ARMA para pronostico de punto de rocío.</i>	61
5.4. <i>Resultados de modelo ARMA para pronostico de Humedad.</i>	62
5.5. <i>Resultados de modelo ARMA para pronostico de presión atmosférica.</i>	62

Introducción

La estructura del presente trabajo de tesis se distribuye de la siguiente manera, detallando cada uno de los componentes y capítulos que conforman la investigación:

- **Capítulo 1:** Introduce las bases del proyecto de tesis, incluyendo la descripción del problema, la hipótesis, justificación, objetivos, y también discute la aportación, los alcances y las limitaciones de la investigación.
- **Capítulo 2:** Se dedica al estado del arte en el pronóstico de series de tiempo, explorando las técnicas y algoritmos más comunes, con un enfoque particular en el pronóstico climatológico y el contexto del monitoreo y pronóstico meteorológico dentro del estado de Querétaro.
- **Capítulo 3:** Describe la metodología empleada en la investigación, la definición y análisis del problema, la adquisición de datos climatológicos, su análisis y preprocesamiento, así como la identificación de variables clave para el pronóstico de precipitaciones.
- Los capítulos sobre el desarrollo y resultados de la investigación se distribuyen en 3 capítulos diferentes, reflejando la metodología iterativa aplicada:
 - **Capítulo 4:** Detalla el modelo CRHUDA, incluyendo un análisis de sensi-

bilidad para verificar su efectividad en el pronóstico de precipitaciones y su factor de escalamiento.

- **Capítulo 5:** Propone un algoritmo para el pronóstico de precipitaciones, facilitando su implementación en un sistema de alerta temprana así como también se muestran los resultados y problemas identificados.
- **Capítulo 6:** Presenta una diagrama conceptual de un algoritmo diseñado por medio de principios arquitectónicos de software para el pronóstico de precipitaciones, siendo una posible solución a los problemas evidenciados en el capítulo 5.
- **Capítulo 7:** Aborda la discusión de la investigación, presentando las publicaciones y otros productos generados, así como las perspectivas de trabajo futuro derivadas de los resultados obtenidos.
- **Capítulo 8:** Concluye el trabajo de investigación, resumiendo las conclusiones alcanzadas a partir de los hallazgos de la investigación.

1.1. Descripción del Problema

Según Cargua et al. (2017) existe una gran variedad de dispositivos con los cuales se puede realizar un monitoreo de las condiciones climatológicas en un punto dado e incluso arrojan un pronóstico aproximado de algunas de las condiciones medidas. Al encontrar variedad de instrumentos surge la diversidad de precios en el mercado, donde la mayoría son de costos relativamente altos debido a su complejidad y funcionamiento. De igual forma, así como se pueden encontrar algunos dispositivos económicos para el monitoreo climatológico, éstos en la mayoría de los casos no son capaces de satisfacer las necesidades en cuanto a precisión, disponibilidad y durabilidad del dispositivo.

De forma adicional existen diversas plataformas que “indican” cuál es el estado

del clima actual, siendo algunas de ellas plataformas como *Google*, *Apple* y el asistente inteligente *Alexa*. Sin embargo, a pesar de “informar” sobre el clima en cada zona, en realidad emplea una interpolación de estaciones virtuales, calculando las condiciones de determinada ubicación mediante estaciones muy alejadas, lo que conlleva a tener un parámetro más cercano a una suposición que a un valor real, por lo que no es posible emplear dicha información de forma científica.

Debido a que los mecanismos actuales de pronóstico de precipitaciones en tiempo real son muy costosos, y las alternativas no poseen una efectividad aceptable o su tiempo de pronóstico es muy corto como para ser empleado en los sistemas de alerta temprana. La opción más viable es el uso del modelo CRHUDA, el cual posee hasta 8 h de aviso empleando una cantidad de variables sumamente menor, lo que en el área del Machine Learning es preferible ya que genera un coste computacional menor generando a su vez una reducción de los costes económicos del sistema, tal como lo menciona Bugaets et al. (2018) en su trabajo, así como también aumenta la eficacia de los modelos. Lamentablemente el algoritmo empleado para obtener un pronóstico requiere ser sometido a una mayor experimentación y mejora, pues el hecho de que sea un algoritmo empírico genera una enorme dificultad para su implementación en entornos reales además de que al depender del factor humano, es vulnerable a los errores humanos.

1.2. Hipótesis

Esta investigación se basa en la hipótesis de que se puede crear un algoritmo de pronóstico efectivo, utilizando el modelo CRHUDA como fundamento, para predecir las precipitaciones. Este algoritmo tendría la finalidad de ser integrado en un sistema de alerta temprana.

Esta investigación busca responder a las siguientes preguntas principales:

1. ¿Es posible usar el modelo CRHUDA para generar un algoritmo capaz de pronosticar en tiempo real el momento de inicio de una precipitación?
2. ¿Existe la posibilidad de optimizar o mejorar el modelo CRHUDA para obtener un mejor algoritmo de implementación? ¿Qué métodos de optimización podrían ser utilizados?
3. ¿Con cuánta anticipación y exactitud puede pronosticar precipitaciones el algoritmo que se pretende desarrollar?
4. ¿Qué tipo de precipitaciones pueden pronosticarse con un menor error? (convectiva, huracanada o ambas)
5. ¿Qué tan complicado será el replicar el algoritmo propuesto para pronosticar precipitaciones, incluso en lugares donde no se tengan datos históricos de precipitaciones?

1.3. Justificación

Como señala Jain y Jain (2017), el clima desempeña un rol crucial en la vida cotidiana. Sus efectos pueden influir significativamente en aspectos tan variados como la vida personal, la economía y el medio ambiente. Esto es igualmente aplicable a los habitantes de Querétaro. Fenómenos climáticos como lluvias intensas, temperaturas extremas y vientos fuertes tienen el potencial de alterar profundamente la rutina diaria no solo en Querétaro sino en cualquier ciudad. El clima, constituido por una amplia gama de variables como temperatura, humedad, presión atmosférica y punto de rocío, se ve influenciado por múltiples factores interconectados. Entre estos factores se encuentran la contaminación ambiental, la ubicación geográfica, la estación del año y eventos específicos en la región.

La ciudad de Querétaro enfrenta anualmente tormentas severas que resultan en inundaciones en varias áreas urbanas, ocasionando considerables daños y pérdidas

materiales. De acuerdo con información de la Oficina para la Reducción de Riesgos de las Naciones Unidas, se estima que, desde 1970, el Estado de Querétaro ha sufrido pérdidas económicas superiores a los 64 mil millones de pesos debido a las inundaciones (Cano, 2017).

Pese a que el modelo y el algoritmo CRHUDA es capaz de pronosticar el suceso de precipitaciones con un amplio margen de tiempo, aún es necesario el continuar con su desarrollo pues el algoritmo empírico no es lo suficientemente práctico como para ser llevado a una implementación en un sistema de alerta temprana.

Debido a lo expuesto anteriormente es necesario elaborar un algoritmo capaz de generar un pronóstico preciso de la precipitación, y a su vez, de las demás variables climatológicas relacionadas a ésta.

1.4. Objetivos

1.4.1. Objetivo General

Desarrollar un algoritmo hidroinformático para el pronóstico de precipitaciones en tiempo real para su uso en sistemas de alerta temprana, usando como base el modelo de pronóstico CRHUDA y teniendo como área de estudio la zona metropolitana del Estado de Querétaro.

1.4.2. Objetivos Específicos

Este proyecto tiene los siguientes objetivos específicos:

- Observar y simular el funcionamiento del modelo CRHUDA, así como realizar un análisis de sensibilidad para identificar sus áreas de oportunidad y cómo mejorarlo.

- Analizar y comparar los métodos existentes para la optimización de pronósticos de series de tiempo.
- Desarrollar e implementar los algoritmos y modelos hidroinformáticos necesarios para poder cumplir con el objetivo general.
- Analizar los algoritmos existentes para el pronóstico de precipitaciones en tiempo real y del algoritmo desarrollado durante la presente investigación para comparar su desempeño.
- Definir el tipo de pronóstico de intensidad óptimo para tener la mayor exactitud posible.
- Identificar el tipo de precipitación para la cual el algoritmo es más eficiente.

1.5. Alcances y Limitaciones

El proyecto de investigación presentado en esta tesis se rige por los siguientes alcances y limitaciones:

- La investigación se centra principalmente en verificar la viabilidad de pronosticar el tiempo y la intensidad de las precipitaciones. No obstante, no se aborda su implementación en sistemas de alerta temprana.
- Dada la envergadura y complejidad del proyecto, se limitará a realizar pruebas de laboratorio controladas, empleando las variables climatológicas registradas para el pronóstico.

1.6. Aportación y Pertinencia

El modelo CRHUDA representa un desarrollo innovador dentro del campo de la hidroinformática, un área que presenta amplias oportunidades de avance mediante

la integración de las ciencias computacionales. A pesar de su potencial, el modelo CRHUDA aún no ha sido aplicado en contextos reales, lo que subraya la brecha existente entre el desarrollo teórico y su implementación práctica.

La hidroinformática emerge como un campo de oportunidad significativo, principalmente porque la mayoría de las investigaciones en este ámbito todavía no explotan plenamente las ventajas que las tecnologías computacionales modernas pueden ofrecer. Esta disciplina, que cruza las fronteras entre la hidrología, la ingeniería ambiental y la informática, tiene el potencial de transformar la manera en que comprendemos y gestionamos los recursos hídricos y los fenómenos asociados, como las precipitaciones.

Sin embargo, una limitación notable en el estado actual de la tecnología es que las herramientas de pronóstico disponibles no logran predecir con precisión ni el inicio ni la intensidad de las precipitaciones. Esta imprecisión representa un desafío considerable para la gestión de recursos hídricos, la planificación urbana y la prevención de desastres naturales. Por lo tanto, el desarrollo y la mejora de modelos como CRHUDA, que buscan abordar estas limitaciones, son cruciales para avanzar en la precisión y la fiabilidad de las predicciones meteorológicas y hidrológicas.

Estado del Arte

México, un país reconocido por su rica biodiversidad y sus extensos centros urbanos, se enfrenta anualmente a una variedad de fenómenos meteorológicos extremos que afectan tanto sus áreas naturales como sus zonas urbanas. El estudio de estos eventos extremos, aunque no es un campo nuevo, ha sido el foco de investigaciones especializadas a lo largo de los años (Azli & Rao, 2010; Bonta & Shahalam, 2003). De particular interés es el monitoreo y pronóstico de precipitaciones extremas en grandes áreas urbanas y reservas ambientales, dada su alta prioridad. La importancia de estos pronósticos radica en su capacidad para evaluar el riesgo hidrológico asociado a la severidad de las tormentas, así como el impacto de las pérdidas económicas que estas pueden causar, afectando gravemente a las economías locales (Serinaldi & Kilsby, 2015).

Desde el período de 2000 a 2014, 2147 millones de dólares en pérdidas y 186 muertes anuales («Índice de peligro por inundación (IPI).» 2017). En materia de precipitaciones, el año 2013 fue muy intenso, especialmente para el mes de septiembre, la lámina histórica de las precipitaciones aumentó 60 %, con respecto a la media histórica, registrando una media mensual de 227.3 mm; superando el récord ocurrido en 1955 de 212.1 mm (Gutiérrez-López et al., 2019). Las inundaciones y las lluvias extremas son actualmente un grave problema que debe ser atendido. Uno

de los mecanismos utilizados por el gobierno mexicano para proteger y mitigar los daños causados por diferentes tipos de desastres es la implementación de Sistemas de Alerta Temprana (SAT). Las predicciones meteorológicas utilizan diversas variables que tratan de proporcionar un pronóstico adecuado y preciso. (Variables como: temperatura, humedad, viento, punto de rocío entre otras).

Los Sistemas de Alerta Temprana (SAT) son un conjunto de procedimientos e instrumentos diseñados para monitorear amenazas o eventos adversos, ya sean de origen natural o antrópicos, que son previsibles. Estos sistemas implican la recolección y procesamiento de datos e información, brindando pronósticos o predicciones temporales sobre la actividad de dichas amenazas y sus posibles efectos. Gracias a los SAT, millones de personas en todo el mundo han podido salvar sus vidas y proteger sus medios de subsistencia.

La relevancia de un SAT reside en su capacidad para anticipar, con un grado de certeza significativo, cuándo y dónde una amenaza o evento adverso, natural o generado por humanos, puede desencadenar situaciones de riesgo. Esto subraya la importancia de emitir alertas con suficiente antelación.

En el contexto de las inundaciones, los SAT especializados desempeñan un rol crucial al monitorear las condiciones hidrometeorológicas y el comportamiento de los ríos o cuencas hidrográficas. Esto permite pronosticar la probabilidad de inundaciones en áreas específicas.

Los SAT automatizados operan a través de la observación y el monitoreo, empleando redes telemétricas, estaciones de medición de lluvia y niveles de ríos, modelos hidrológicos computarizados, sensores remotos y satélites. Estos sistemas supervisan la cantidad de lluvia y los niveles de los ríos, permitiendo pronosticar con precisión las crecidas.

Un pronóstico aceptable de la intensidad de las precipitaciones es una cuestión esencial en los pronósticos de los fenómenos extremos de flujo en superficie (Damrath et al., 2000). Las investigaciones realizadas hasta la fecha han tendido a

centrarse en datos hidrometeorológicos como el viento, la temperatura, la humedad y la presión atmosférica. Esta información se ha utilizado en una previsión diaria de los flujos de corriente durante períodos de 5 a 7 días (Rasouli et al., 2012).

Sin embargo, el objetivo principal de un SAT de inundaciones o un sistema de predicción del tiempo; es permitir señales de alerta antes de los fenómenos extremos, principalmente para las lluvias intensas de muy corto plazo. Los recientes avances en las técnicas de aprendizaje automático han acentuado la necesidad de utilizar los datos hidrometeorológicos mencionados para un SAT eficaz a muy corto plazo (aviso de lluvias torrenciales dentro de las próximas 3 a 9 h) (Moon et al., 2019b).

2.1. Mecanismos para el pronóstico de series de tiempo.

Los métodos actuales para el pronóstico de series de tiempo han experimentado un desarrollo significativo, impulsados por los avances en campos como la estadística, la inteligencia artificial y el análisis de grandes volúmenes de datos (big data).

Se denomina serie de tiempo a cualquier conjunto de datos que consiste en mediciones, registros u observaciones realizadas en intervalos sucesivos de tiempo. En consecuencia, una serie de tiempo es una secuencia ordenada de observaciones de una variable específica a lo largo del tiempo.

Una serie estacionaria se caracteriza por tener momentos estadísticos constantes en el tiempo, tanto en su origen como en su media. Este tipo de serie se observa en situaciones donde los patrones de demanda que afectan la serie son relativamente estables y no presentan grandes variaciones a lo largo del tiempo (Vidrio-Sahagún et al., 2021).

2.1.1. Modelos de Aprendizaje Automático y Deep Learning

La Inteligencia Artificial (IA), una rama destacada de la informática, se dedica al desarrollo de algoritmos, sistemas y técnicas para dotar a las máquinas de capacidades similares a la inteligencia humana. Esto abarca habilidades como el aprendizaje automático, el procesamiento del lenguaje natural, el reconocimiento de patrones y la toma de decisiones. Existen dos categorías principales de IA: la IA débil y la IA fuerte. La IA débil se centra en sistemas diseñados para ejecutar tareas específicas y limitadas. En contraste, la IA fuerte se orienta hacia sistemas con capacidades generalistas, capaces de realizar cualquier tarea humana. La IA se ha integrado en diversos campos, incluyendo robótica, medicina, finanzas, industria y educación (Rouhiainen, 2018).

Las inteligencias artificiales son empleadas en una gran diversidad de tópicos para pronósticos, algunos de ellos son:

- **Pronóstico de series temporales:** Esta metodología se emplea para estimar futuros valores en una secuencia de datos a lo largo del tiempo, aplicándose en áreas como precios de acciones, mediciones de temperatura o niveles de producción (Hurtado & Ríos, 2008).
- **Pronóstico de demanda:** Esta técnica se centra en la proyección de la cantidad esperada de ventas de un producto o servicio en el corto plazo (Corres et al., 2009).
- **Pronóstico de riesgo:** Este enfoque se utiliza para detectar y cuantificar los riesgos inherentes a situaciones o eventos específicos (Rojas Vilches & Martínez Reyes, 2011).
- **Pronóstico de eventos:** Esta técnica implica la anticipación de la ocurrencia de eventos futuros, que pueden incluir desastres naturales o fluctuaciones del mercado (Zou et al., 2022).

- Pronóstico climático: Esta técnica se basa en modelos matemáticos y datos meteorológicos para hacer proyecciones sobre las condiciones climáticas futuras (Easterling & Stern, 1999).
- Pronóstico de fallos: Este método se aplica para anticipar posibles averías en equipos o sistemas antes de que estas sucedan (Clements & Hendry, 2002).
- Pronóstico financiero: Esta técnica se utiliza para proyectar los resultados económicos futuros de empresas o inversiones (Abu-Mostafa & Atiya, 1996).

Uno de los campos particulares de la inteligencia artificial es el Machine Learning, el cual se enfoca en desarrollar algoritmos y modelos que pueden aprender automáticamente a partir de datos, sin ser explícitamente programados. Esto permite a las computadoras mejorar su rendimiento en una tarea específica sin ser explícitamente programadas para hacerlo (Rouhiainen, 2018).

Dentro del Machine Learning existen diferentes técnicas y algoritmos, dos de las más comunes son la clasificación y la regresión.

La clasificación es un tipo de aprendizaje supervisado en el que el objetivo es asignar una etiqueta o clase a una entrada de datos. Por ejemplo, clasificar un correo electrónico como spam o no spam, o clasificar una imagen como perro o gato. Los algoritmos de clasificación populares incluyen k-nearest neighbors, decision trees, y redes neuronales.

La regresión es también un tipo de aprendizaje supervisado, pero en lugar de predecir una etiqueta o clase, el objetivo es predecir un valor numérico. Por ejemplo, predecir el precio de una casa en función de características como el tamaño, la ubicación, y el número de habitaciones. Los algoritmos de regresión populares incluyen regresión lineal, regresión logística y redes neuronales.

2.1.2. Modelos ARMA

El modelo ARMA, abreviatura de AutoRegressive Moving Average, representa un enfoque estadístico avanzado en el análisis de series de tiempo. Este modelo combina dos conceptos fundamentales: autorregresión y medias móviles, operando bajo la premisa de que las variables en estudio son estacionarias, es decir, sus propiedades estadísticas como la media y varianza permanecen constantes a lo largo del tiempo (G. E. Box et al., 2015).

En los modelos autorregresivos (AR), la idea central es que una variable dependiente en un tiempo específico (t) se puede expresar como una función de sus propios valores anteriores. Es decir, en un modelo $AR(p)$, donde p denota el orden, la variable en el tiempo (t) se pronostica en base a sus valores previos hasta $(t-p)$. Este enfoque es particularmente efectivo para realizar pronósticos a corto y mediano plazo en series temporales, ya que incorpora la dependencia temporal de los datos. El valor de p , o 'orden', indica cuán lejos en el tiempo se mira atrás para la predicción, y un mayor orden implica la inclusión de más datos históricos en el modelo (G. E. Box et al., 2015).

Por otro lado, el modelo de Medias Móviles (MA) se centra en la relación de una variable con errores aleatorios o términos estocásticos de períodos anteriores. En este modelo, se asume que la variable de interés depende linealmente no solo de su valor actual sino también de uno o varios términos de error recientes (G. E. Box et al., 2015).

Combinando estos dos enfoques, el modelo ARMA integra tanto los procesos autorregresivos como los de medias móviles para proporcionar un análisis más robusto y completo de series temporales. Esta fusión permite capturar tanto la dependencia temporal inherente en los datos (AR) como las fluctuaciones aleatorias que pueden influir en ellos (MA), ofreciendo así una herramienta poderosa para el análisis y pronóstico en series temporales (G. E. Box et al., 2015).

2.1.3. Modelos ARIMA

Un modelo ARIMA, que significa "Autoregressive Integrated Moving Average", es un modelo estadístico similar al anterior, es utilizado para el análisis y predicción de series temporales. El modelo ARIMA es especialmente útil para series temporales que muestran patrones no estacionarios, es decir, aquellos cuyas propiedades estadísticas como la media y la varianza cambian con el tiempo. La fortaleza del modelo ARIMA radica en su capacidad para modelar diferentes tipos de patrones temporales mediante la combinación de tres conceptos principales: procesos autorregresivos (AR), diferenciación integrada (I) y medias móviles (MA). Ese componente extra que no pose los modelos ARMA es la diferenciación. Se utiliza para hacer que la serie sea estacionaria, es decir, para asegurarse de que las propiedades estadísticas de la serie no dependan del tiempo. Esto es crucial porque los modelos AR y MA asumen que la serie es estacionaria (González Casimiro, 2009).

2.2. Mecanismos actuales para el pronóstico de precipitaciones.

Actualmente existen diversas investigaciones con la finalidad de realizar pronósticos de precipitaciones. Su enfoque va desde el pronóstico del inicio de la temporada de lluvias para un año en específico, el cual corresponde al rango long-range forecasting hasta el pronóstico de la lluvia acumulada que caerá en el día.

En el trabajo de Fuentes et al. (2020), se realiza una corrección del modelo matemático Weather Research and Forecasting (WRF), el cual es empleado para realizar simulaciones mediante radares de la formación de nubosidades y su evolución a lo largo del tiempo. Este mecanismo es capaz de pronosticar la cantidad de lluvia acumulada que sucederá por día en un tiempo dado. La corrección que se realizó fue una mejora de precisión mediante el uso de una red neuronal artificial.

Como se puede apreciar en la Figura 1, se muestra una comparativa de los valores obtenidos por las estaciones que usó el autor como valor observado del acumulado de precipitación en 24 h desde el 4 de noviembre de 2018 hasta el 28 de febrero de 2019 representadas por la serie de color negro. La serie azul representa el pronóstico del modelo WRF simulado mediante el programa SisPI. De color rojo se representa la serie correspondiente a su modelo mejorado mediante redes neuronales artificiales, denominado MLP.

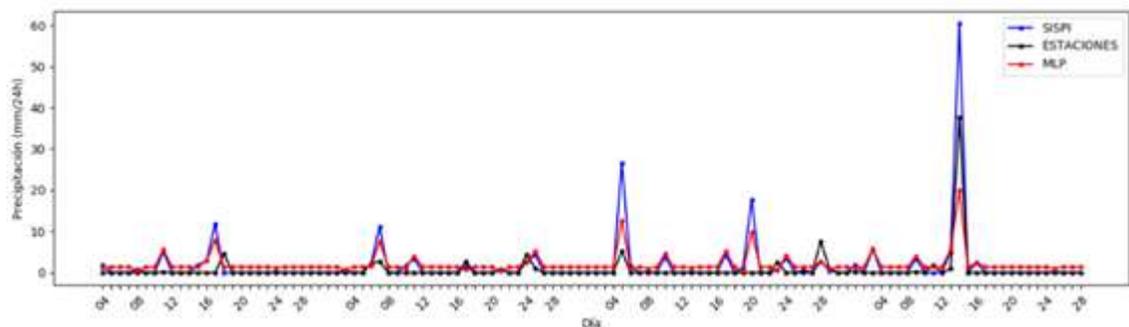


Figura 2.1: Acumulado de precipitación en 24 h desde el 4 de noviembre del 2018 al 28 de febrero del 2019. Obtenido de Fuentes et al. (2020)

Se puede observar que ambos modelos pronostican de forma adecuada la tendencia de los valores acumulados de precipitación; sin embargo, no logran pronosticar de forma adecuada el valor máximo de lluvia acumulada, aunque el modelo optimizado si logra su objetivo de mejorar respecto al modelo original WRF. Además, es importante destacar que, pese a que logra un más que adecuado pronóstico del acumulado de precipitación, únicamente realiza un pronóstico de la evolución de la tormenta, mas no indica el momento exacto del comienzo de la precipitación.

En cuanto al trabajo realizado por Urdiales y Céleri (2018), se llevó a cabo una modelación de la ocurrencia de precipitación empleando cadenas de Markov de primer orden. Como resultado de esta investigación, se concluyó que el modelo ajustado para todos los meses no logra capturar adecuadamente la estacionalidad de las precipitaciones. Aunque el modelo presentó buenos resultados durante los

meses húmedos, su efectividad disminuyó significativamente en los meses secos.

Por otro lado, Acharya et al. (2019) demostraron que el uso de redes neuronales artificiales ofrece mejores resultados en comparación con un modelo de regresión no lineal múltiple. Este hallazgo se basa en la obtención de un menor valor de error en el pronóstico de precipitaciones mediante el uso de diversas métricas de evaluación, lo que resalta la superioridad de las redes neuronales artificiales sobre los modelos de regresión convencionales en tareas de pronóstico complejas.

En el trabajo realizado por Azimi y Moghaddam (2020) se muestra un modelo del acumulado de precipitación empleando un rango de pronóstico corto, del inglés Short-range weather forecasting (SRWF). En dicho modelo se emplea una red neuronal artificial en conjunto con un clasificador gaussiano basado en el índice estandarizado de precipitación, del inglés Standardized precipitation index (SPI). Se empleó dicha clasificación, ya que se realizó una simulación de la evolución de una precipitación empleando un gráfico sobre la zona observada. Este tipo de simulaciones suelen ser realizadas mediante el empleo de satélites para monitorear su evolución.

Hernández-Leal et al. (2016), realizaron en su trabajo una comparativa del desempeño de 4 técnicas diferentes de pronóstico de precipitaciones orientando su uso a pronosticar la intensidad de una precipitación a lo largo de una serie de tiempo observada. Las técnicas que compararon fueron el método de pronóstico ingenuo, regresión lineal múltiple, redes neuronales artificiales y sistemas neurodifusos. Concluyeron que no son efectivos al pronosticar la no existencia de precipitaciones; es decir, mientras pronostiquen dentro de una tormenta obtienen resultados satisfactorios. Sin embargo, al pronosticar el valor de la intensidad de precipitación mientras no haya presente ninguna precipitación, son incapaces de obtener resultados positivos, generando en muchas ocasiones falsos positivos.

Un atributo para tener en cuenta en cuanto a los métodos de pronóstico es el coste, tanto en el rubro computacional como en el económico. En cuanto al ámbito computacional Lombardi et al. (2018) mencionan que los métodos de pronóstico

deben tener un coste computacional bajo para que sea simple su utilización, en cuando a lo económico destacaron los impactos económicos de las falsas alarmas respecto a las alarmas omitidas, conocidas también en el ámbito computacional y de la ciencia de datos como errores de tipo 1 y tipo 2 respectivamente, puesto que los costes generados por omitir una alarma (error tipo 2) son sustancialmente mayores a los costes generados por una falsa alarma (errores tipo 1) por lo que es sumamente preferible que el los modelos tiendan a tener fallas generando falsas alarmas para que el costo por errar una predicción sea el menor posible. Calvello et al. (2015) comenta que una de las formas para reducir los costes de las falsas alarmas es generar un umbral mediante el cual es posible definir un nivel de alerta en el sistema.

Hofmann y Schüttrumpf (2019) mencionan en su trabajo que un sumamente relevante es los costos de desarrollo e implementación de los sistemas de alerta temprana actuales puesto que es un reto considerable en diferentes países.

2.3. Modelo CRHUDA.

El modelo *CrossingHumidity, Dewpoint, and Atmospheric pressure* (CRHUDA) representa un enfoque innovador para el pronóstico del inicio de precipitaciones. Este modelo se caracteriza por estar compuesto por dos series independientes de variables climáticas, $S1 \subseteq S2$, donde tanto $S1$ como $S2$ se definen a través de un modelo autorregresivo de primer orden, AR(1). Esta estructura permite al CRHUDA capturar las dinámicas complejas asociadas con los cambios en la humedad, el punto de rocío y la presión atmosférica, y cómo estos factores interactúan para influir en la iniciación de eventos de precipitación (Gutierrez-Lopez et al., 2019b; Muñoz-Mandujano et al., 2022).

$$\begin{aligned}
S1 &= H_t = \phi_1 H_{t-1} + \varepsilon_t \\
S2 &= C_t = \varphi_1 C_{t-1} + \varepsilon_t
\end{aligned}
\tag{2.1}$$

Donde $\phi_1 = r_1$ y $\varphi_1 = r_1$ representan los coeficientes de correlación serial para $k = 1$ en cada una de las series. Esto implica la existencia de un coeficiente de proporcionalidad en ambas series que facilita su escalado y cruzamiento en el tiempo t_1 . Es importante destacar que el coeficiente de auto-correlación de retraso k es igual a 1, lo que juega un papel crucial en la modelización de estas series temporales.

Según la ecuación 2.1, la variable C se reemplaza por el valor de la presión atmosférica (P), y en el denominador, la variable (T) se sustituye por la temperatura del punto de rocío (Td), en línea con la descripción proporcionada por Shaw et al. (2011b). Este enfoque permite la representación gráfica de dos series temporales distintas: la primera correspondiente a la humedad ($S1$) y la segunda ($S2$), definida por la relación $\left(\frac{\partial P}{\partial Td}\right)$, que remite a la relación Clausius-Clapeyron. La simplificación de ambas series permite evidenciar su relación directa, como se ilustra en la ecuación 2.2.

$$\left. \begin{aligned}
&S1 = \text{humedad} \\
&S2 = \frac{\Delta P}{\Delta T} = \frac{\partial P}{\partial Td} = \frac{\text{presión atmosférica}}{\text{punto de rocío}}
\end{aligned} \right\} \text{CRHUDA } (S1 \subseteq S2) \rightarrow (S1 \cap S2)
\tag{2.2}$$

El modelo CRHUDA se implementa mediante un algoritmo empírico que centra su atención en la interacción entre dos series temporales clave, $S1$ y $S2$. La esencia de este algoritmo radica en la observación del cruce de estas series, donde el primer cruce ($t1 : S1 \subseteq S2$) actúa como indicador de una inminente precipitación. Posteriormente, y dentro de un intervalo de tiempo especificado T_t , un segundo cruce se produce ($t2 : S1 \cap S2$), marcando el inicio efectivo de la precipitación (E). La precisión

del modelo se ve influenciada directamente por el tamaño de Δt , la brecha temporal entre el segundo cruce y el evento de precipitación. Así, cuanto menor sea la separación entre el momento t_2 y el suceso E , mayor será la exactitud del pronóstico. Este procedimiento está ilustrado en la Figura 2.2, donde se pueden visualizar los puntos críticos de intersección entre S_1 y S_2 y su correlación con el inicio de las precipitaciones.

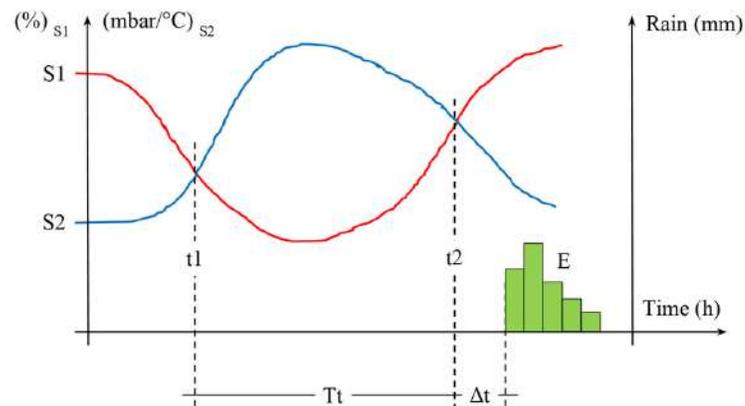


Figura 2.2: Esquema conceptual del funcionamiento del modelo CRHUDA (*CRossingHUmidity, Dewpoint, and Atmospheric pressure*) (Gutierrez-Lopez et al., 2019b; Muñoz-Mandujano et al., 2022).

Conforme a lo reportado por Gutierrez-Lopez et al. (2019b) y Muñoz-Mandujano et al. (2022), se procedió al desarrollo de una herramienta hidroeinformática diseñada específicamente para evaluar la eficacia del modelo y algoritmo CRHUDA, el cual está protegido bajo un registro de derechos de autor. Para llevar a cabo esta evaluación, se recurrió al uso de datos provenientes de varias estaciones climatológicas, enfocando el análisis en la zona metropolitana del estado de Querétaro. Este análisis contempló un total de 523 eventos de tormentas, proporcionando una muestra representativa para evaluar la aplicabilidad del modelo en dicha región.

Además, se extendió el alcance del estudio para incluir datos de 17 de los 32 estados de la República Mexicana, abarcando un total de 1237 tormentas convectivas registradas en el periodo comprendido entre 1999 y 2018. Los resultados de

estos análisis, que ilustran la capacidad del modelo CRHUDA para capturar y predecir fenómenos meteorológicos, se presentan en las Figuras 2.3, 2.4 y 2.5, donde se ofrecen ejemplos visuales de cómo el modelo logra representar el comportamiento de las tormentas analizadas.



Figura 2.3: Resultados de la aplicación del modelo CRHUDA a los datos del Centro Histórico en Querétaro, el 16 de agosto de 2014. Elaboración propia (ver Gutierrez-Lopez et al., 2019b; Muñoz-Mandujano et al., 2022, para mas detalles).

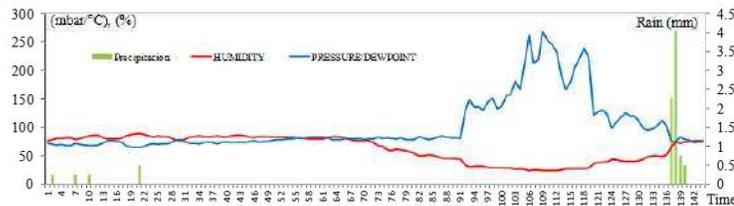


Figura 2.4: . Resultados de la aplicación del modelo CRHUDA a los datos de Atlacomulco en el Estado de México, el 23 de abril de 2018. Elaboración propia (ver Gutierrez-Lopez et al., 2019b; Muñoz-Mandujano et al., 2022, para mas detalles).

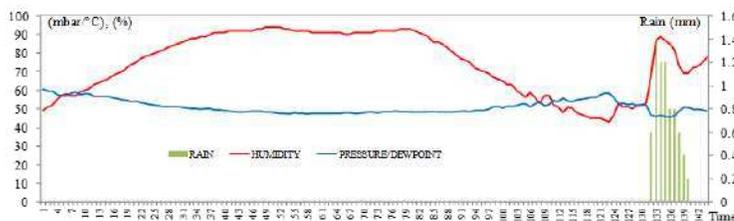


Figura 2.5: Resultados de la aplicación del modelo CRHUDA a los datos de Dzilbachen en Campeche, el 22 de febrero de 2018. Elaboración propia(ver Gutierrez-Lopez et al., 2019b; Muñoz-Mandujano et al., 2022, para mas detalles).

2.4. Técnicas de pronóstico dentro de la inteligencia artificial

2.4.1. Redes Neuronales Artificiales

Las redes neuronales artificiales (RNA) representan un modelo computacional inspirado en el procesamiento paralelo del sistema nervioso humano. Están compuestas por unidades de procesamiento básicas denominadas neuronas artificiales, las cuales imitan el comportamiento de las neuronas biológicas. Estas redes adquieren conocimiento a través de un proceso de aprendizaje, y la conexión entre las neuronas, conocida como peso sináptico, se utiliza para almacenar este conocimiento van Gerven y Bohte (2017).

Un aspecto distintivo de las redes neuronales artificiales es su arquitectura computacional distribuida, que emula la estructura del sistema nervioso, ofreciendo una robusta capacidad para el procesamiento de información compleja.

La siguiente tabla resume las tres características fundamentales de una red neuronal artificial, destacando su topología de red, la regla de aprendizaje y los tipos de entrenamiento:

Tabla 2.1: Características fundamentales de una red neuronal artificial.

Topología de Red	Permite determinar la capacidad representativa de cómo una cantidad de neuronas están distribuidas en capas y entre sí, donde la topología de red se adapta al problema a resolver.
Regla de aprendizaje	Las redes neuronales son sistemas de aprendizaje integrados, capaces de aprender a través del entrenamiento.
Tipo de entrenamiento	Las RNA pueden someterse a dos tipos de entrenamiento: uno durante la etapa de aprendizaje, donde los pesos sinápticos se ajustan, y otro durante la etapa de ejecución, cuando la red opera con datos reales.

Fuente: Elaboración propia, con base en Rivas-Asanza et al. (2018).

Debido a su constitución y fundamentos, las redes neuronales artificiales presentan una serie de ventajas y desventajas que son cruciales para comprender su aplicabilidad y limitaciones en diversos contextos. Según Rivera (2005), estas características se pueden describir de la siguiente manera:

Ventajas

- Posee la capacidad de realizar un aprendizaje adaptativo, que consiste en aprender a realizar tareas a partir de un conjunto de datos iniciales.
- Cuenta con una auto organización. Una red neuronal es capaz de crear su propia organización o representación de la información que recibe en la etapa de aprendizaje.
- Cuenta con una tolerancia parcial a fallos, pues un daño parcial a la red afecta el funcionamiento de esta, sin embargo, no la destruye por completo, esto es debido a la redundancia de la información contenida.
- Las redes neuronales pueden ser operadas en tiempo real, puesto que permiten la paralelización de la red, aunque para sacar el máximo provecho a una red neuronal artificial es necesario contar con hardware especial.

Desventajas Las redes neuronales artificiales presentan ciertas limitaciones, entre las que se incluyen:

- Complejidad creciente al enfrentar tareas de aprendizaje de gran escala.
- El tiempo de aprendizaje se incrementa con el número de patrones a reconocer y su capacidad para distinguir entre patrones similares.
- Un tiempo de convergencia elevado en aplicaciones que requieren un gran número de ejemplos para el entrenamiento.

- La falta de capacidad para interpretar los resultados de manera autónoma, dejando esta tarea a cargo del programador o la aplicación que utiliza la red.

Las redes neuronales han demostrado ser herramientas poderosas en la resolución de diversos problemas, destacándose en la predicción de series temporales y la detección de no linealidades en los datos, incluso en presencia de ruido o incompletitud. Su eficacia en el reconocimiento de patrones y la solución de problemas complejos es bien reconocida Villada et al. (2012).

Ian, Yoshua (2015) destacan la existencia de una amplia gama de tipos de redes neuronales artificiales, diferenciadas por su arquitectura específica:

- Perceptrón Multi Capa (MLP, por sus siglas en inglés).
- Red Neuronal Recursiva (RNN).
- Memoria a Corto y Largo Plazo (LSTM).
- Red Neuronal Recurrente (RNN).

La diversidad de arquitecturas se complementa con una variedad de herramientas y programas diseñados para facilitar el desarrollo de estas redes, entre los que se encuentran librerías como Keras y PyTorch, ambas destinadas al uso con el lenguaje de programación Python.

Metodología y desarrollo

Esta investigación se lleva a cabo empleando como fundamento la metodología de investigación basada en diseño. El término *diseño* se refiere al proceso de creación instructiva que se elabora, implementa y posteriormente se somete al escrutinio de la investigación. A pesar de que los procesos de investigación bajo este enfoque metodológico no están estrictamente definidos, comparten una serie de etapas fundamentales, independientemente del número de fases en las que se subdivide el proceso. Estas etapas incluyen: la definición del problema, el diseño, el desarrollo, la implementación y la evaluación de la solución propuesta (de Benito Crosetti & Ibáñez, 2016).

Este enfoque permite no solo abordar de manera sistemática el problema de investigación desde su conceptualización hasta la evaluación de las soluciones propuestas, sino que también facilita una iteración constante entre estas fases, permitiendo refinamientos continuos basados en la retroalimentación obtenida en cada etapa.

Esta investigación se estructura en cuatro etapas o fases principales, descritas a continuación:

1. Definición y análisis del problema:

- a) Análisis del estado del arte respecto al pronóstico de precipitaciones.

- b) Adquisición de la base de datos de precipitaciones diarias y horarias, en la República Mexicana y en el estado de Querétaro.
- c) Análisis y preprocesamiento de la información contenida en las bases de datos. Caracterización de las series de tiempo empleando estadísticos básicos espaciales y temporales.

2. Diseño y desarrollo:

- a) Análisis y comparativa de técnicas para la optimización de algoritmos de pronóstico.
- b) Análisis, implementación y limitaciones del modelo CRHUDA para el pronóstico de los diferentes tipos de precipitaciones.
- c) Comparativa del modelo CRHUDA respecto a otros mecanismos para pronóstico de precipitaciones.
- d) Comparativa de técnicas para la optimización de algoritmos de pronóstico y modelos probabilísticos.
- e) Desarrollo del algoritmo para la creación de los modelos de pronóstico.

3. Implementación y evaluación:

- a) Calibración, evaluación y ajuste del modelo empleando datos históricos de las variables climatológicas como datos de entrada.
- b) Validación de los resultados empleando una validación cruzada, utilizando series de tiempo históricas para reproducir la condición de precipitación con la red neuronal propuesta y evaluar los errores y dispersión del pronóstico.

4. Transferencia de conocimiento y desarrollo tecnológico:

- a) Desarrollo de tesis.
- b) Escritura y envío de artículos.

c) Participación y envío de trabajos en congresos.

3.1. Definición y análisis de problema

3.1.1. Análisis del estado del arte respecto al pronóstico de precipitaciones.

La información correspondiente a esta actividad de la metodología se encuentra descrita a detalle en el capítulo 2 del estado del arte.

A manera de resumen la información más importante sobre el estado del arte es lo siguiente:

3.1.2. Adquisición de datos climatológicos

En nuestro país, existen limitaciones en cuanto a la disponibilidad de redes de monitoreo de variables climatológicas en tiempo real. Dada la diversidad de regiones climáticas, se presenta un desafío al estudiar la distribución espacio-temporal de la precipitación Gutiérrez-López y Barragán-Regalado, 2019. En América Latina y específicamente en México, se observa que las estaciones climatológicas son más prevalentes que las estaciones hidrométricas. Esto se debe, en gran medida, a que las estaciones climatológicas son menos costosas y más sencillas de instalar que las hidrométricas.

A nivel nacional, el Servicio Meteorológico Nacional (SMN) cuenta actualmente con más de 160 Estaciones Meteorológicas Automáticas (EMAS). Estas EMAS no solo están estratégicamente situadas en terrenos adecuados, sino que también están equipadas con una variedad de instrumentos y aparatos especializados en la medición de precipitaciones.

Los datos de las EMAS fueron obtenidos a través de una solicitud realizada por el Centro de Investigaciones del Agua del estado de Querétaro (CIAQ).

La RedCIAQ, operada por el Centro de Investigaciones del Agua del estado de Querétaro (CIAQ) de la Universidad Autónoma de Querétaro, desempeña una función crucial en el monitoreo ambiental. Esta red, enfocada en la recolección de datos y el desarrollo de herramientas hidroeinformáticas, consta de más de 40 estaciones climatológicas. Estas estaciones, que registran variables hidrológicas en tiempo real cada minuto, están principalmente situadas en el área metropolitana del estado de Querétaro, aunque su alcance se extiende a todo el territorio estatal, con al menos una estación en cada municipio. Las variables monitoreadas por la RedCIAQ incluyen temperatura, presión atmosférica, punto de rocío, velocidad y dirección del viento, precipitación, radiación solar y humedad relativa.

RedCIAQ proporciona acceso a sus datos a través de su sistema web. Las mediciones se encuentran en archivos individuales con extensión CSV por cada estación y agrupadas en carpetas por mes y año. un esquema del arbol de directorios se muestra a continuación:

- /2012
 - /enero
 - estacion1.csv
 - estacion2.csv
 - etc.
 - /febrero
- /2013
- /2014
- /2015

- etc.

A diferencia de los datos de las EMAS, para obtener los datos de la RedCIAQ es necesario descargarlos desde su pagina, y para hacerlo hay que descargar cada archivo de forma individual, lo cual es un trabajo demasiado tardado, debido a ello se realizo una herramienta que por medio de la técnica de scraping descarga cada uno de los archivos de la RedCIAQ.

3.1.3. Análisis y preprocesamiento de datos

El análisis y preprocesamiento de datos constituyen tareas de vital importancia y requieren una dedicación considerable de tiempo en los campos de la inteligencia artificial y *Data Science*.

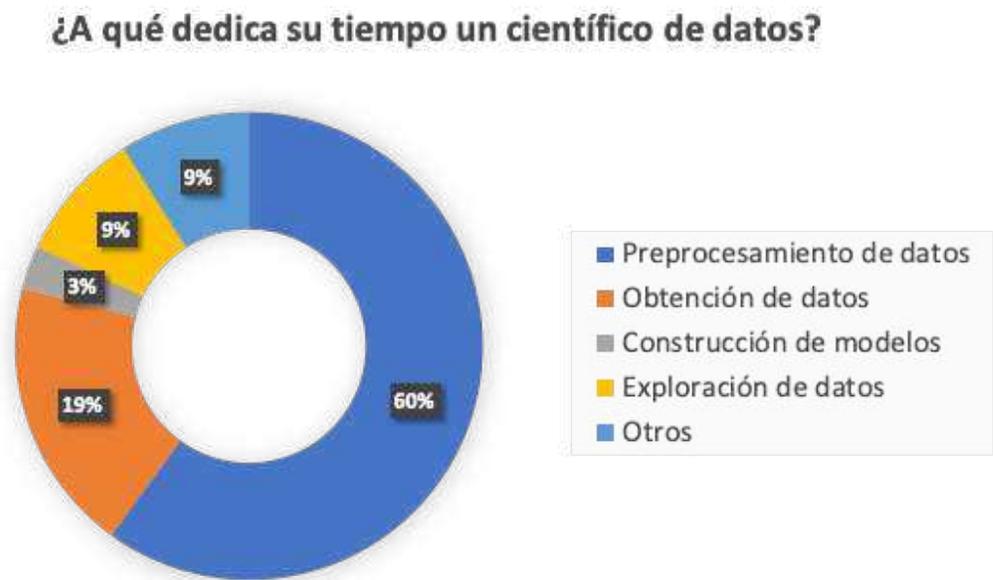


Figura 3.1: *Actividades principales de los científicos de datos. Basado en Press, 2016*

Conforme a una encuesta realizada por Press 2016, cuyos resultados se ilustran

en la figura 3.1, se identifica que la principal actividad de un científico de datos es el preprocesamiento de datos.

El objetivo de llevar a cabo tal análisis y preprocesamiento es abordar y mitigar cuatro problemas clave que pueden surgir en cualquier conjunto de datos, independientemente de su origen, formato o método de adquisición. Estos problemas son:

- Ausencia de valores.
- Inconsistencia de datos.
- Valores duplicados.
- *Outliers*.

La ausencia de valores representa un desafío común en el manejo de extensos conjuntos de datos. Las causas de esta ausencia pueden ser variadas, como errores en la medición, fallos en el almacenamiento o recuperación de información, entre otras. La principal preocupación radica en que estas ausencias dificultan el entrenamiento efectivo de sistemas de aprendizaje automático, dado que la falta de datos no es numéricamente manejable.

Entre las posibles soluciones a la problemática de datos ausentes, se destaca la técnica de interpolación, especialmente útil en conjuntos de datos temporales. Esta consiste en completar los valores faltantes utilizando la información de los puntos adyacentes. Otra alternativa es el relleno con un valor fijo, como la media, la moda o incluso el valor cero. Además, se puede emplear técnicas de regresión para predecir los valores ausentes.

La inconsistencia en los datos a menudo surge debido a errores en el formato o tipo de datos durante su procesamiento, lectura o almacenamiento. Por ejemplo, discrepancias en la estructura de fechas o caracteres no numéricos en campos numéricos. Para asegurar la coherencia, es esencial realizar diversas validaciones y correcciones.

La presencia de valores duplicados puede introducir sesgos significativos, especialmente al entrenar redes neuronales, ya que estos sesgos podrían ser aprendidos por la red.

Los *Outliers*, o valores anómalos, que se desvían significativamente de los límites normales, pueden ser el resultado de errores en la medición, almacenamiento o inserción de datos. Estos valores pueden distorsionar la distribución de los datos, afectando negativamente el proceso de aprendizaje.

Adicionalmente, la información proporcionada por CONAGUA sobre las EMAS se almacena en archivos en formato XLS (Formato de Microsoft Excel 97-2004 Workbook). Estos archivos están organizados por estación y contienen datos mensuales. Cada registro de medición se almacena con un intervalo de 10 minutos.

En adición, se dispone de datos que abarcan desde el año 2000 hasta el 2022. Como es de esperar, estos datos presentan una variedad de formatos y nomenclaturas para las variables, lo que complica su uso. Se identificaron específicamente ocho diferentes combinaciones de formatos y nombres de encabezados, las cuales se detallan a continuación:

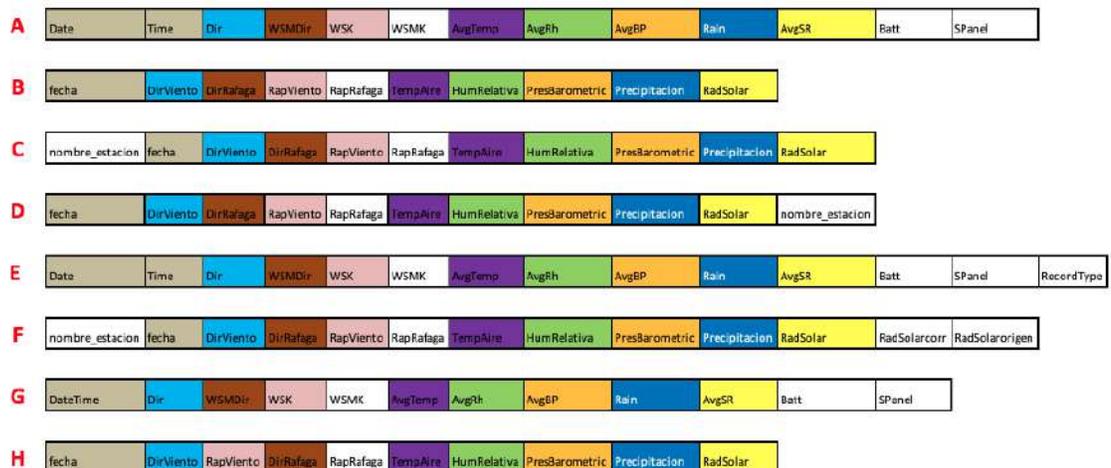


Figura 3.2: Variaciones de los encabezados dentro de las fuentes de datos proporcionada por CONAGUA de las EMAS.

Como se muestra en la figura 3.2, las diferencias en los formatos de datos, aunque puedan parecer menores a primera vista, presentan varios desafíos. Por ejemplo, algunos formatos combinan inglés y español, las fechas y horas pueden estar en un solo campo o divididas en dos, y se utilizan tanto el sistema métrico como el imperial en diferentes conjuntos de datos.

El principal problema surge al intentar tratar todos estos datos de manera uniforme. Las discrepancias significativas entre ellos impiden su uso directo en aplicaciones como el entrenamiento de redes neuronales. Por esta razón, se decidió desarrollar una herramienta para normalizar estos datos, facilitando su manipulación no solo para el entrenamiento de redes neuronales, sino también para cualquier otra tarea que requiera acceso a esta información.

La herramienta desarrollada, denominada provisionalmente DATEX (*Data Extractor*), es un programa de línea de comandos distribuido en formato ejecutable y compatible con múltiples plataformas, incluyendo Windows, Linux y Mac OS. Para su uso, debe ejecutarse a través de comandos en CMD o *PowerShell* en entornos Windows, o mediante la terminal en sistemas Linux y Mac OS.

La herramienta DATEX está diseñada para ejecutar tres tareas principales:

1. Normalización de archivos.
2. Concatenación de archivos.
3. Detección de tormentas.

La normalización de archivos es la función primordial de DATEX, creada específicamente para estandarizar los archivos procedentes de las estaciones EMAS de CONAGUA. Esta funcionalidad ajusta los valores y nombres de columnas a un formato estándar, generando un nuevo archivo con extensión XLS. Por razones de usabilidad, esta tarea se ejecuta automáticamente en las otras funciones, previendo la

posibilidad de que un usuario requiera estas funcionalidades sin haber normalizado previamente el archivo.

La función de concatenación de archivos se introdujo considerando la necesidad de trabajar con más de un mes de datos para entrenamientos, pronósticos y diversos cálculos. Esta tarea toma múltiples archivos XLS y los combina en un solo archivo ya normalizado.

La detección de tormentas es una funcionalidad diseñada para identificar eventos de lluvia que superen un umbral definido por el usuario. Genera un archivo XLS para cada tormenta detectada, registrando datos desde 24 horas antes del inicio de la tormenta hasta 50 minutos después del último registro de precipitación.

3.1.4. Relación Clausius-Clapeyron y su influencia sobre las precipitaciones

El agua en la naturaleza se presenta en tres estados - sólido, líquido y gaseoso - cuya transición entre uno y otro está influenciada principalmente por la temperatura y la presión atmosférica (Hou et al., 2013). El cambio de agua de estado sólido a líquido se conoce como punto de fusión, mientras que la transición directa de sólido a gas se denomina sublimación (J. X. Wang & Gaffen, 2001). Por otro lado, la presión absoluta en la cual el agua pasa de líquido a gaseoso se llama presión de vapor, y esta transición ocurre a una temperatura específica conocida como punto de ebullición (Suparta et al., 2017).

El concepto de humedad se refiere a la presencia de vapor de agua en la atmósfera. La presión parcial ejercida por las moléculas de agua en el aire alcanza un límite en el que se establece un equilibrio dinámico entre la evaporación y la condensación, conocido como punto de saturación. Este equilibrio es altamente dependiente de la temperatura T , pero notablemente independiente de la presión del aire seco (Camuffo, 2014). En el contexto de los cambios de presión ΔP y su relación con el

trabajo termodinámico W y el calor latente C , se observa que el trabajo realizado es proporcional al cambio en la temperatura ΔT , como se muestra en la ecuación 3.1 (Capparelli, 2013; Pumo et al., 2019).

$$\Delta W = C \frac{\Delta T}{T} \quad (3.1)$$

Además, para que ocurra un cambio de estado, es necesario considerar la variación en el volumen específico, es decir, la diferencia entre los volúmenes específicos del gas (V_g) y del líquido (V_l) (Romps, 2016).

$$W = \Delta P(V_g - V_l) \quad (3.2)$$

Al comparar esta ecuación con la formulación previamente mencionada para el trabajo y el calor latente, se obtiene:

$$\frac{\Delta P}{\Delta T} = \frac{\partial P}{\partial T} = \frac{C}{T \Delta V} \quad (3.3)$$

Esta ecuación, conocida como la relación Clausius-Clapeyron (C-C), fue deducida inicialmente por Clapeyron basándose en la teoría de Carnot y posteriormente comprobada por Clausius. Esta relación es fundamental en termodinámica para calcular la temperatura de un sistema cerrado durante un cambio de fase, conociendo los volúmenes específicos y la relación entre el aumento de la presión de saturación y la temperatura T (Agard & Emanuel, 2017b; Camuffo, 2014; Lorenz & DeWeaver, 2007). En un cambio de fase, la temperatura y la presión son constantes por definición, lo que permite que la relación C-C describa el comportamiento del sistema en estas condiciones.

La hipótesis central de la relación C-C postula que, con el aumento de la temperatura, la humedad relativa se mantiene constante mientras que la humedad específica se incrementa debido al aumento de la disponibilidad de humedad en la atmósfera (Park & Min, 2017b). En ciertas regiones tropicales, el aumento total de

las precipitaciones puede superar las estimaciones de la relación Clausius-Clapeyron, lo que sugiere la necesidad de un enfoque compensatorio distinto (Chang et al., 2016). La precisión en la predicción de precipitaciones exige un uso meticuloso de la relación C-C, tal como lo respalda (Lenderink, Barbero, Loriaux & Fowler, 2017), quien señala que "las actuales relaciones de escalamiento de precipitación-temperatura indican una respuesta al calentamiento que supera la relación Clausius-Clapeyron para extremos de precipitación horaria". Esto se visualiza claramente en gráficos de temperatura ($1/T$) contra presión ($\ln P$) (Bürger et al., 2014; Peleg, Marra, Fatichi, Molnar et al., 2018). Según (Maíllo & Fernández-Pineda, 2011), existen hasta 11 reglas distintas para deducir la Ecuación 3.3, y diversos métodos han sido desarrollados para establecer la relación C-C entre precipitaciones extremas subdiarias y la temperatura media diaria (Park & Min, 2017b).

La temperatura del punto de rocío, T_d , es aquella a la cual el aire alcanza la saturación si se enfría a presión constante (Seidel, Grant, Pszeny & Allman, 2007). T_d corresponde a la temperatura donde la presión de vapor iguala la presión de saturación del aire. De igual forma, el volumen de vapor en la atmósfera se puede representar a través de la presión que este genera (Chang et al., 2016; Millán et al., 2010a). Sin embargo, la presión atmosférica total es la suma de la presión del aire seco y la presión del vapor de agua (Harder & Pomeroy, 2013b; Holley et al., 2014). Así, la máxima presión de vapor posible depende de la temperatura superficial (Dahm et al., 2019), y conforme aumenta la temperatura, mayor cantidad de vapor de agua puede contener el aire (Mohr & Kunz, 2013; Myoung & Nielsen-Gammon, 2010). Esta dinámica puede describirse mediante la relación Clausius-Clapeyron (Gao et al., 2007; Y. Wang et al., 2018).

Por ende, cuando el aire está saturado de vapor de agua, la presión de este último depende únicamente de la temperatura (Sim et al., 2019b). Además, la temperatura de corrientes ascendentes húmedas iniciadas en la superficie y la profundidad de las nubes están claramente influenciadas por el punto de rocío superficial

(Lenderink, Barbero, Loriaux & Fowler, 2017). Por tanto, es esencial proponer un modelo sencillo que relacione la presión y el punto de rocío ($\frac{\partial P}{\partial T_d}$), y que también valide la importancia crucial de la humedad superficial en la actividad convectiva. Considerando que parámetros meteorológicos como la presión, temperatura y humedad relativa varían con la altitud (Liu et al., 2015), se justifica la implementación de un modelo sinóptico que integre datos de humedad, punto de rocío y presión atmosférica.

Análisis de sensibilidad

4.1. Descripción de series del modelo CRHUDA

La temperatura de punto de rocío, T_d , es la temperatura a la cual el aire está saturado si se enfría a presión constante (Agard & Emanuel, 2017a; Shaw et al., 2011a). Es decir, T_d es la temperatura a la cual la presión de vapor es igual a la presión de saturación del aire. De la misma manera, el volumen de vapor presente en la atmósfera puede expresarse a través de la presión que este vapor genera (Camuffo, 2014; Lorenz & DeWeaver, 2007).

Sin embargo, la presión total de la atmósfera es la suma de la presión causada por el aire seco más la presión producida por el vapor de agua (Holley et al., 2014; Romps, 2014). Así, la presión máxima de vapor que puede estar presente depende de la temperatura ambiente (Chang et al., 2016). A medida que la temperatura aumenta, más presión de vapor puede contener el aire (Bürger et al., 2014; Lenderink, Barbero, Loriaux & Fowler, 2017). Esto puede expresarse mediante la relación Clausius-Clapeyron (Peleg, Marra, Fatichi, Molnár et al., 2018; Velasco & Fernández-Pineda, 2008). Por lo tanto, cuando el aire está saturado con vapor de agua, la presión del vapor de agua depende únicamente de la temperatura (Harder & Pomeroy, 2013a; Millán et al., 2010b; Seidel, Grant, Pszenny & Allman, 2007).

De esta manera, se trazan dos series temporales, la primera es los datos de Humedad Z^1 y la segunda es la serie definida por el Punto de rocío y la Presión atmosférica Z^2 . El cruce de estas dos series indicará el comienzo de la alerta, ocho horas más tarde las series se cruzarán nuevamente y en ese momento comenzará la precipitación (figura 1).

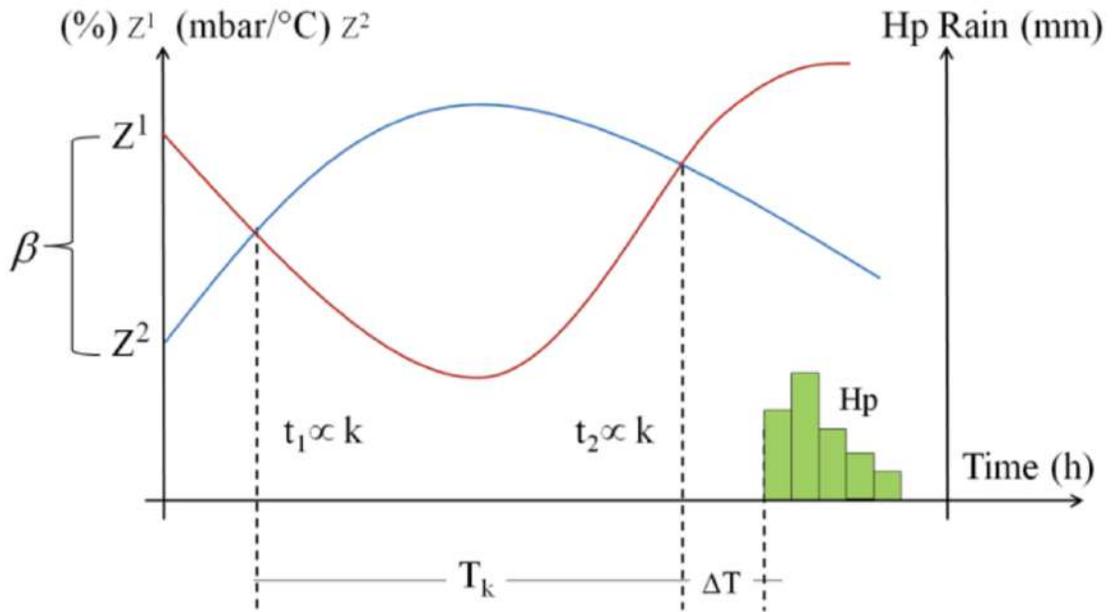


Figura 4.1: *Readaptación del esquema conceptual del modelo CRHU-DA (Crossing Humidity, Dewpoint, and Atmospheric pressure).* (Muñoz-Mandujano et al., 2024).

Esto significa que para que la precipitación comience, debe ocurrir en $t_1 \rightarrow (Z^1_{t-k} \cap Z^2_{t-k})$, y considerando $T_k + \Delta T$ en $t_2 \rightarrow (Z^1_{t-k} \cap Z^2_{t-k})$. Si $\Delta T = 0$ el pronóstico del inicio del evento de precipitación H_p es preciso. Si $\Delta T \neq 0$, hay un retraso en el tiempo en el inicio del evento de precipitación H_p (Gutierrez-Lopez et al., 2019a; Muñoz-Mandujano et al., 2022; Muñoz-Mandujano et al., 2024).

4.1.1. Modelos autorregresivos de promedios móviles ARMA(p,q)

Estos modelos se aplican a series anuales; son una combinación de los modelos autorregresivos AR(p) y de los modelos de promedios móviles MA(q) (G. E. P. Box & Jenkins, 1976); el valor de p corresponde al número de parámetros asociados a la parte autorregresiva y el q es el número de parámetros a la componente de promedios móviles. La representación general de los modelos ARMA(p,q) es:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (4.1)$$

donde ε_t es una variable aleatoria independiente del tiempo, ϕ_p son parámetros de la componente autorregresiva, θ_q son parámetros de la componente de promedios móviles. Si se cuenta con una muestra representativa del proceso, los parámetros ϕ_p se determinan de manera que se reproducen las correlaciones r_1, r_2, \dots, r_p calculadas con los datos. Este método considera que los datos y los residuales son normales e independientes. Una propiedad de los modelos ARMA(p,q) es que las autocovariancias de orden 1 a orden q dependen de los parámetros autorregresivos ϕ_p y de los promedios móviles θ_q , mientras que para órdenes mayores solo dependen de los parámetros autorregresivos. Para el caso de series discretas la autocovariancia se calcula como:

$$C_k = \frac{1}{n} \sum_{t=1}^{N-K} (x_t - \bar{x})(x_{t-k} - \bar{x}); \quad 0 < k < N \quad (4.2)$$

4.1.2. Modelos estocásticos multivariados

La práctica del análisis multivariado combinado con series de tiempo temporales en temas hidrometeorológicos ha sido desarrollada para estimar variables hidrológicas en sitios con registros escasos o nulos. Mientras que las funciones espectrales permitían estudiar las relaciones de frecuencia entre dos series tempo-

rales, la cuestión central en el ámbito temporal era cómo mejorar las predicciones de los modelos univariantes. Los hidrólogos utilizan el dominio del tiempo (frecuencia) en series multivariadas para poder considerar fenómenos climáticos de manera simultánea. Las predicciones de Z_t pueden hacerse utilizando la historia pasada de la variable única Z_t , o considerando un conjunto de información más amplio $(Z_{t-k}^1, Z_{t-k}^2, \dots, Z_{t-k}^n; k = 0, 1, 2, \dots)$ y construyendo un modelo multivariante $Z_t = f(Z_{t-k}^1, Z_{t-k}^2, Z_{t-k}^3, \dots, Z_{t-k}^n, \varepsilon_{t-k}^1, \varepsilon_{t-k}^2, \varepsilon_{t-k}^3, \dots, \varepsilon_{t-k}^n; k > 0)$. Esta modelización multivariante tiene un doble interés: proporciona un esquema explicativo más que descriptivo de la evolución de Z_t , y teóricamente debería dar lugar a mejores previsiones que las proporcionadas por el modelo univariante $\Phi(B)Z_t \approx \Theta(B)\varepsilon_t$. Con estos antecedentes, se propone el uso de modelos estocásticos para describir el comportamiento de las series del modelo CRHUDA $Z_t = f(Z_{t-k}^1, Z_{t-k}^2, Z_{t-k}^3, \dots, Z_{t-k}^n, \varepsilon_{t-k}^1, \varepsilon_{t-k}^2, \varepsilon_{t-k}^3, \dots, \varepsilon_{t-k}^n; k > 0)$, modelos del tipo multivariante distribuidos en el tiempo; en los cuales las entradas X_t son variables climáticas.

4.1.3. Autocorrelación como validación del modelo CRHUDA

La función de autocovarianza C_k mide la covarianza para un par de valores separados por un intervalo de longitud k denominado tiempo de retraso. Proporciona información sobre la variabilidad de la serie y los vínculos temporales entre los distintos componentes de la serie X_t . Si solo nos interesa estudiar las correlaciones entre los distintos pares (X_t, X_{t+k}) , definimos la función de autocorrelación (r_k) del proceso X_t es decir:

$$r_k = \frac{C_k}{C_0} = \frac{\sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{n \sum_{t=1}^N (X_t - \bar{X})^2} \quad (4.3)$$

Donde:

$$C_0 = n \sum_{t=1}^N (X_t - \bar{X})^2 \quad (4.4)$$

$$\text{Cov} = [X_t, X_{t-k}] = E[(X_t - \bar{X})(X_{t-k} - \bar{X})] \quad (4.5)$$

r_k Representa el coeficiente de correlación para un par de valores separados por un tiempo k . Los valores r_0, r_1, r_2, \dots serán las tres series de tiempo del modelo CRHUDA. Es decir:

$$CRHUDA (Z_{t-k}^1 \subseteq Z_{t-k}^2) \rightarrow (Z_{t-k}^1 \cap Z_{t-k}^2) \begin{cases} Z_{t-k}^1 & = \text{humidity} \\ Z_{t-k}^2 & = \text{atmospheric pressure} \\ & \text{dewpoint} \end{cases} \quad (4.6)$$

4.2. Análisis de sensibilidad

Se introduce un factor de escala β que será proporcional a la serie de valores de autocorrelación serial.

$$(Z_{t-k}^1 \subseteq Z_{t-k}^2) \propto \beta(Z_{t-k}^1 \cap Z_{t-k}^2) \quad (4.7)$$

Se puede determinar la función de autocorrelación $r_k = f(Z_{t-k}^1 \subseteq Z_{t-k}^2)$ para cualquier proceso, proporcional a $f = [\beta(Z_{t-k}^1 \cap Z_{t-k}^2)]$. De esta forma se propone un punto de análisis de sensibilidad del modelo CRHUDA. Los residuales de las dos series $\varepsilon (Z_{t-k}^1 \subseteq Z_{t-k}^2)$ deben tener media cero y no deben estar correlacionados, lo cual garantiza su independencia del tiempo k . Esto es: $\varepsilon = \{\varepsilon_t, t \in N\}$; $E[\varepsilon_t | \varepsilon_{t-1}] = 0$ y $\text{Var}[\varepsilon_t] = \sigma_\varepsilon^2$. De las propiedades de la función de autocorrelación se tiene que $C_0 = E[(X_t - \bar{X})^2] = \text{Var}[X_t] = \sigma_x^2$ cuando $k = 0$ entonces C_0 es la varianza estacionaria en t_0 . Por lo tanto si las series Z_{t-k}^1 y Z_{t-k}^2 son independientes, entonces $r_k = \frac{C_k}{C_0} \rightarrow 0$ tiende a cero. Sin embargo, al ser un modelo estocástico multivariado las series deben cumplir la condición de independencia y estacionariedad conjunta. Esto puede verificarse de manera sistemática con la ayuda de los límites del correlograma-multivariado; de esta forma se propone que:

Lemma 1. Un coeficiente de autocorrelación r_k calculado a partir de una serie estocástica multivariada $f(Z_{t-k}^1, Z_{t-k}^2, Z_{t-k}^3, \dots, Z_{t-k}^n)$ fuera de los límites de confianza del correlograma; representa un punto de quiebre en las series Z_{t-k}^n . Este punto de quiebre ubica a su vez los tiempos t_1 y t_2 de cruce entre las series que forman el modelo CRHUDA.

Lemma 2. El tiempo k de cruce entre las series del modelo CRHUDA; puede ser modificado en un tiempo T_k empleando un factor de escala β tal que $(Z_{t-k}^1 \subseteq Z_{t-k}^2) \propto \beta(Z_{t-k}^1 \cap Z_{t-k}^2)$.

Ajustando la función de autocorrelación se tiene:

$$C_k = E[(X_t - \bar{X})(X_{t+k} - \bar{X})] = E[(X_{t-k} - \bar{X})(X_\beta - \bar{X})] \text{ con } \beta = t + k \quad (4.8)$$

Para asociar el tiempo Lag que ocurre en r_k es necesario hacer:

$$(Lag)_k = Lag(Lag + 1)(Lag + k - 1) = \frac{\Gamma(Lag + k)}{\Gamma(Lag)} \approx \frac{(Lag + k - 1)!}{(Lag - 1)!} \quad (4.9)$$

Si $\beta > 0$; $(\beta)_0 = 1$ y $(1)_k = k!$ entonces k es el tiempo Lag que ocurre en r_k (Lemma 1) para $\frac{\beta}{\beta-1} > 0$

$$\begin{aligned} (\beta)_k &= \beta(\beta + 1)(\beta + 2) \dots (\beta + k - 1) \\ (\beta)_k &= (r + 1)(r + 2) \dots (r + k) \\ (\beta)_k &= \frac{1}{r}(r + 1)(r + 2) \dots (r + k) = \left(-\frac{1}{r}\right) (r)_{k+1} \end{aligned} \quad (4.10)$$

Entonces una relación de recurrencia puede escribirse como:

$$\begin{aligned}
(\beta)_k &= \frac{(\beta - 1)_{k+1}}{\beta - 1} \\
(\beta)_{k+1} &= \beta(\beta + 1)_k
\end{aligned}
\tag{4.11}$$

4.3. Implementación y Validación

Para este paso se emplearon los datos de la red de monitoreo de precipitaciones extremas RedCIAQ recolectados en el capítulo anterior. El procedimiento consistió en tres etapas. En la primera se extrajeron los datos de lámina de precipitación, humedad, punto de rocío y presión atmosférica a más de 500 tormentas entre los años 2012 y 2021 medidos a cada minuto para formar las series Z_{t-k}^1 y Z_{t-k}^2 . La segunda etapa consiste obtener los valores del coeficiente de autocorrelación y así graficar los correlogramas de todas las series $r_k = f(Z_{t-k}^1 \subseteq Z_{t-k}^2)$ para evaluar los puntos de quiebre. En la tercera etapa se calcula y se verifica el factor de escala β .

4.3.1. Presentación del modelo CRHUDA

Se aplicó el modelo CRHUDA (ecuación 6) a los datos de las tormentas que ocurrieron entre 2012 y 2021 en la zona metropolitana de la ciudad de Querétaro. En un sentido sinóptico es posible ver, sorprendentemente, que el cruce de estas dos series temporales $CRHUDA (Z_{t-k}^1 \subseteq Z_{t-k}^2) \rightarrow (Z_{t-k}^1 \cap Z_{t-k}^2)$ genera un punto en el tiempo que permite generar una alerta entre 9 y 10 horas antes del comienzo de la tormenta. Por ejemplo, en la 4.2 se muestran los resultados de la aplicación del modelo CRHUDA en la EMA-Candiles para los registros de la tormenta del 24 y 25 de junio de 2013. Aproximadamente a las 10:20 se observa el primer cruce, la humedad desciende rápidamente, lo que permite iniciar la alerta. Después de este punto, se presenta el cruce en sentido contrario e inicia la precipitación. Los resultados mostraron un tiempo medio de 10 horas $T_t + \Delta t = 619.58$ min entre la alerta t_1 y el inicio de la precipitación t_2 , con una mediana de 8.9 horas (534 min). El

mismo análisis puede realizarse para la tormenta del siguiente día. Con este ejemplo es evidente que la escala del eje de las ordenadas es importante e influye de forma contundente para ubicar los momentos (t_1 & t_2) como instantes precisos de cruces.

Aquí la importancia de los *Lemma 1* y *2*. Encontrar los instantes precisos de cruces t_1 & t_2 es vital para generar las alertas y comunicar a las autoridades locales de protección civil. De esta forma el factor de escala entre las dos series debe ser visualizado y calculado ($Z_{t-k}^1 \subseteq Z_{t-k}^2$) $\propto \beta(Z_{t-k}^1 \cap Z_{t-k}^2)$. Los resultados de la calibración visual del factor de escala fueron entre 0.4 y 2.6 con una media de 1.784.

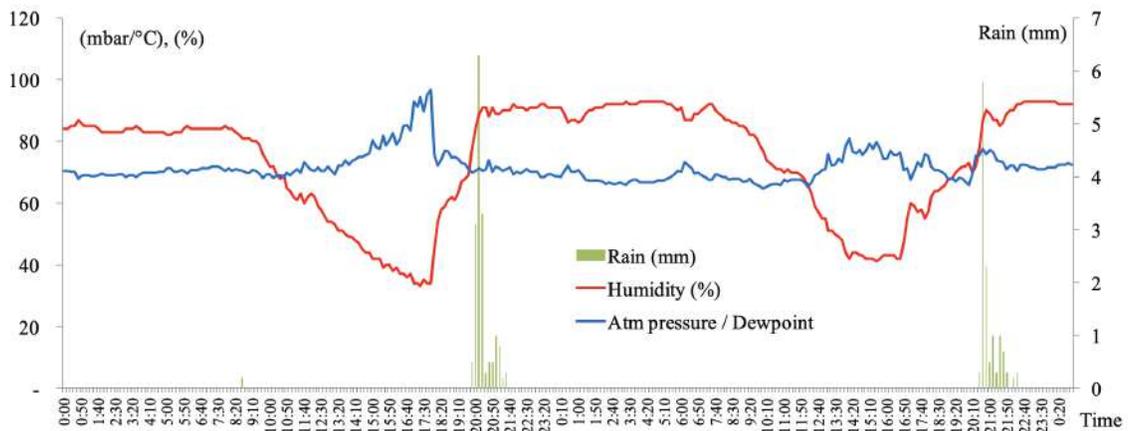


Figura 4.2: Resultados de la aplicación del modelo CRHUDA en la estación Candiles el 24-25 de junio de 2013.

En las figuras 4.3 y 4.4 se muestran los resultados para otras estaciones y con datos de tormentas diferentes. Especialmente se muestran tormentas en donde es evidente la necesidad de estimar el factor de escala. Por ejemplo en la figura 4.3, el cruce entre las dos series de tiempo es imperceptible en algunos tiempos. Un reconocimiento sinóptico es muy comprometido para asociar el resultado con un sistema de alerta. El Lemma 2 entonces debe ser probado matemáticamente a través de los componentes y parámetros de la serie de tiempo tratada como un modelo estocástico autorregresivo multivariado.

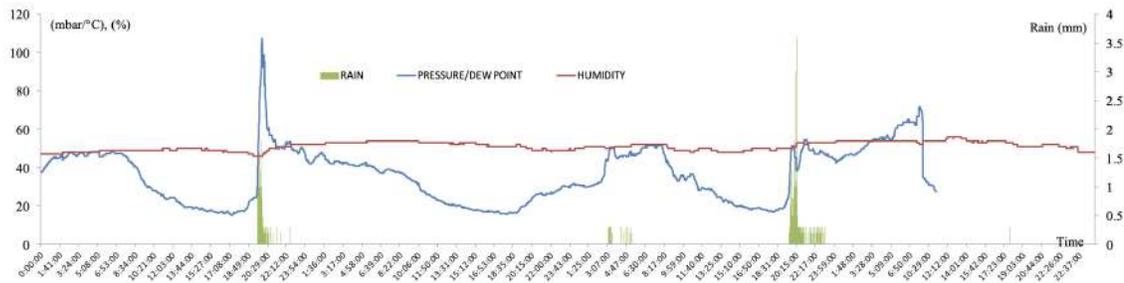


Figura 4.3: Resultados de la aplicación del modelo CRHUDA en la estación Cimatario el 16 de agosto de 2014 (lámina máxima 15.4 mm en 1 h; intensidad máxima 41.2 mm/h; período de retorno de 32 años).

La figura 4.4, muestra una gran secuencia de tormentas para la estación del Centro Universitario. La aplicación del modelo CRHUDA es reveladora debido al número de tormentas que se presentaron. En todos los casos los cruces son identificados perfectamente, lo que permite que se pueda realizar una predicción muy confiable del inicio de la precipitación. A continuación es importante verificar matemáticamente el cálculo del factor de escala β empleando los términos de la función de autocorrelación.

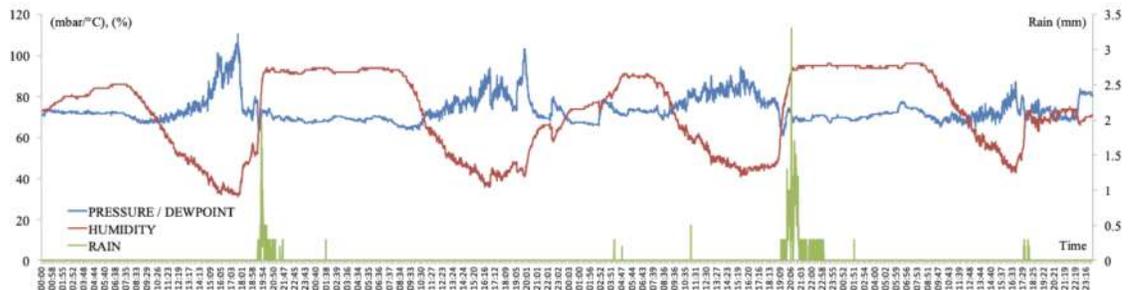


Figura 4.4: Resultados de la aplicación del modelo CRHUDA a los datos del UAQ Cerro Campanas del 16 de agosto de 2014 (lámina máxima 55.2 mm en 1 h; intensidad máxima 63.8 mm/h; período de retorno de 53 años).

4.3.2. Cálculo de correlogramas

Un coeficiente de autocorrelación r_k calculado a partir de una serie estocástica multivariada $f(Z_{t-k}^1, Z_{t-k}^2, Z_{t-k}^3, \dots, Z_{t-k}^n)$ fuera de los límites de confianza del correlograma; representa un punto de quiebre en las series Z_{t-k}^n . Este punto de quiebre ubica a su vez los tiempos t_1 y t_2 de cruce entre las series que forman el modelo CRHUDA. Con esta hipótesis se calcularon los valores de la función de autocovarianza y autocorrelación de lo cada una de las variables, series y modelo completo CRHUDA. La pregunta clave sigue siendo ¿cuál es el tiempo de pronóstico de los eventos de lluvia?, ya que la respuesta tiene implicaciones importantes, ya que un buen modelo de previsión de lluvias pueda utilizarse en un sistema de alerta temprana de inundaciones. De esta forma los correlogramas que se obtienen a partir de las ecuaciones 2, 3 y 4 permiten calcular de forma precisa el factor de escala. La *figura 5* muestra los correlogramas para las series de tiempo de la presión atmosférica para la estación Juriquilla durante la tormenta del 17 Septiembre 2017 (a) una hora, (b) 6 horas (c) 12 horas y (d) 24 horas. Para la misma tormenta, las *figuras 6 y 7* muestran los correlogramas para las series de humedad y punto de rocío respectivamente. En especial esta tormenta provocó graves daños en la zona metropolitana de Querétaro. Un socavón de grandes dimensiones y pérdidas de infraestructura vial, por mencionar algunos. Cabe mencionar que el manejo de las precipitaciones extremas en esta ciudad ha sido ya estudiado en su componente espacial asociada al sistema de alerta de la ciudad (Gutierrez-Lopez, 2021).

Los resultados de los correlogramas de las series analizadas en forma individual muestran casi en todos los casos una ruptura en alguno de los tiempos k . Con estos resultados se verifican dos cosas:

1. Los correlogramas permiten ubicar en el tiempo los cambios en el comportamiento de las series; identificados desde el correlograma como puntos fuera de los límites de confianza.

2. Los valores estimados de los coeficientes de autocorrelación permiten calcular un factor de escala.

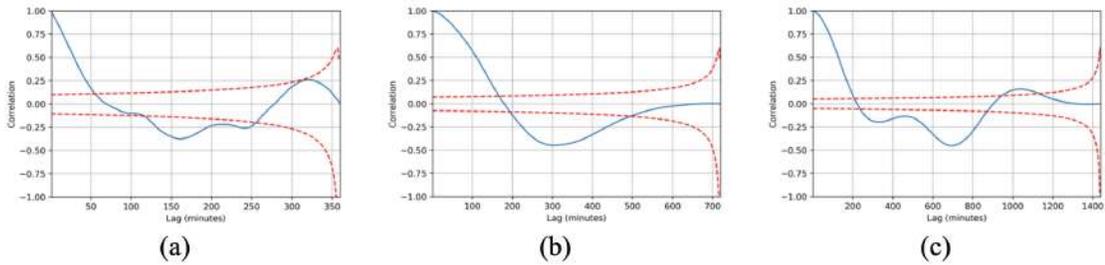


Figura 4.5: Correlograma para las series de tiempo de presión atmosférica. Estación Juriquilla 17 Septiembre 2017 (a) 6 horas (c) 12 horas y (d) 24 horas

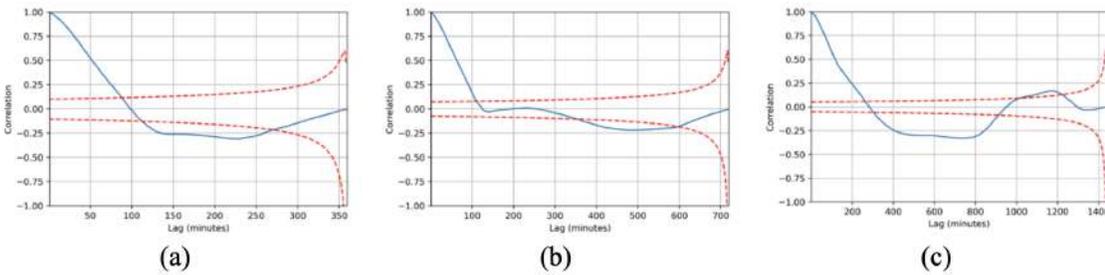


Figura 4.6: Correlograma para las series de tiempo de humedad. Estación Juriquilla 17 Septiembre 2017 (a) 6 horas (c) 12 horas y (d) 24 horas

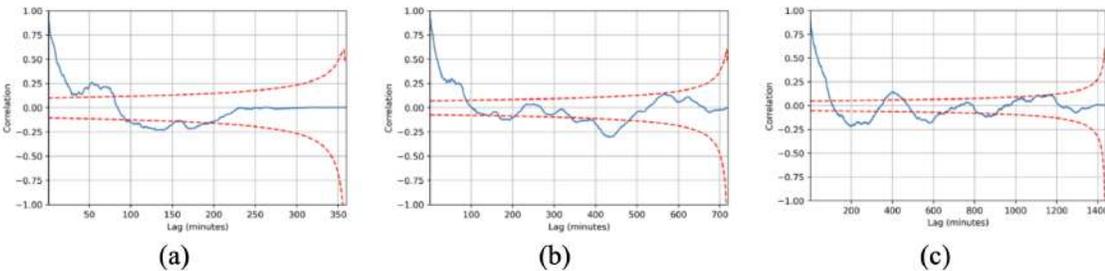


Figura 4.7: Correlograma para las series de tiempo de punto de rocío. Estación Juriquilla 17 Septiembre 2017 (a) 6 horas (c) 12 horas y (d) 24 horas

4.3.3. Cálculo del factor de escala

Estos resultados demuestran que como se pensó cuando se dedujo la ecuación (9), es posible representar el factor de ajuste empleando el modelo completo. Entonces los resultados de los correlogramas del modelo completo deben proporcionar los valores necesarios de los coeficientes de autocorrelación para encontrar el factor de escala. El cálculo del factor de escala se realiza a partir de la premisa propuesta en este trabajo. En donde el tiempo k de cruce entre las series del modelo CRHUDA; es modificado en un tiempo T_k empleando un factor de escala β tal que $(Z_{t-k}^1 \subseteq Z_{t-k}^2)$ o $\beta(Z_{t-k}^1 \cap Z_{t-k}^2)$. A partir de los valores de la función de autocorrelación, se calculan los correlogramas del modelo completo. Por ejemplo, para una de las tormentas más destructivas en los últimos años en la zona metropolitana de la ciudad de Querétaro, ocurrida el 17 de septiembre de 2017. La figura 8, muestra los resultados de los correlogramas del modelo completo CRHUDA para dicha tormenta. Como se observa solo en algunos momentos (tiempos k) el correlograma sale de sus límites. Estos puntos son los valores de $(\beta)_{k+1} = \beta(\beta + 1)_k$ para el cálculo del factor de escala. Como complemento de lo anterior, en la tabla 1 se muestran los resultados de calcular los coeficientes de autocorrelación y todos los parámetros del modelo CRHUDA para tormentas ocurridas en septiembre, el mes más lluvioso del año en la ciudad de Querétaro. El promedio del factor de escala para el mes de septiembre es de 1.107 el valor promedio de $\Delta T = 0 : 24 : 01$ y el promedio de $T_k = 1 : 46 : 45$ De lo anterior se deduce que para el mes de septiembre puede darse una alerta con un tiempo de aproximadamente dos horas de anticipación ($1 : 46 : 45 + 0 : 24 : 01 = 2 : 10 : 46$), siempre y cuando se utilice el factor de escala con los valores calculados. Es importante mencionar que solo se muestran resultados de algunas tormentas y algunas estaciones. Los resultados completos para todas las tormentas y todos los datos de origen están disponibles a solicitud.

Tabla 4.1: Resultados del análisis de sensibilidad y del factor de escala para el modelo CRHUDA aplicado a datos de diversas tormentas extremas registradas en el mes de septiembre. Estación Milenio (Muñoz-Mandujano et al., 2024)

Fecha de la tormenta	β	t_2	ΔT	t_1	T_k
08 Sept 2013	1.23	16:40:00	0:13:00	16:37:00	0:03:00
16 Sept 2013	1.31	15:15:00	0:06:00	14:51:00	0:24:00
21 Sept 2013	1.28	13:54:46	0:44:14	13:54:16	0:00:30
25 Sept 2016	1.24	15:37:00	0:05:00	15:36:00	0:01:00
17 Sept 2017	0.72	0:00:00	0:00:00	0:00:00	0:00:00
26 Sept 2017	0.93	19:26:00	1:40:00	19:12:00	0:14:00
09 Sept 2018	1.24	21:18:00	0:39:00	21:16:00	0:02:00
19 Sept 2019	1.14	20:52:00	0:02:00	11:41:00	9:11:00
29 Sept 2019	1.22	18:10:00	0:18:00	18:08:00	0:02:00
07 Sept 2021	0.77	18:17:00	0:13:00	10:27:00	7:50:00

Algoritmo de pronostico mediante modelo regresivo de tipo LSTM

5.1. Diseño y Desarrollo

5.1.1. Selección de entorno de desarrollo

En el campo dinámico y en constante evolución del Machine Learning, especialmente en el sector de las redes neuronales, los desarrolladores y científicos de datos tienen a su disposición una vasta gama de herramientas y lenguajes de programación diseñados para facilitar la construcción, entrenamiento y despliegue de modelos complejos. Esta riqueza de recursos permite a los profesionales elegir la solución que mejor se ajuste a sus necesidades específicas, ya sea por facilidad de uso, flexibilidad, eficiencia en el procesamiento o compatibilidad con tecnologías existentes.

Entre las bibliotecas más reconocidas y ampliamente utilizadas para la implementación de modelos de Machine Learning en Python, destacan varias por su robustez, extensiva documentación y comunidad activa de usuarios. *Keras* se presenta como una interfaz de alto nivel, centrada en la simplicidad y la facilidad de

uso, que permite una experimentación rápida y eficiente (Chollet et al., 2015). Por otro lado, *PyTorch* es elogiado por su flexibilidad y dinamismo, ofreciendo un enfoque más intuitivo para la definición de modelos por su capacidad para trabajar con cambios en tiempo real (Paszke et al., 2019). *TensorFlow*, desarrollado por Google, se distingue por su vasta escala de operación, desde investigaciones académicas hasta despliegues a gran escala en producción, abarcando una amplia gama de tareas con su complejo ecosistema de herramientas (Abadi et al., 2016). Finalmente, *Theano* es una de las bibliotecas pioneras en este ámbito, aunque su desarrollo se haya detenido, sirvió como piedra angular para muchas de las innovaciones actuales en computación simbólica y optimización de cálculos matemáticos (Al-Rfou et al., 2016). Estas herramientas, cada una con sus propias ventajas y particularidades, forman la columna vertebral de la actual infraestructura de Machine Learning, posibilitando avances significativos en la inteligencia artificial y el análisis de datos.

De acuerdo con el análisis realizado por Gevorkyan et al. 2019, el rendimiento de estas librerías tiende a ser muy similar. Sin embargo, se señala que PyTorch puede ser ligeramente más lento en comparación con las demás. Los resultados de esta comparativa se muestran en la tabla 5.1.

Tabla 5.1: Resultados del análisis comparativo de las bibliotecas de aprendizaje automático

Librería	Exactitud (%)	Tiempo (segundos)
Keras	98.56	113.80
TensorFlow	89.07	63.48
Theano	97.38	257.29
PyTorch	98.07	1492.29

Fuente: elaborado con base en Gevorkyan et al. 2019.

Tras realizar una revisión meticulosa de la documentación de las diversas bibliotecas de Machine Learning disponibles, se determinó que Keras representa la opción más adecuada para el desarrollo de nuestro proyecto. Esta elección se basa primordialmente en la versatilidad de Keras para integrar y operar con modelos sustentados tanto en TensorFlow como en Theano, lo que le confiere una destacada

flexibilidad y una amplia compatibilidad. Este atributo es crucial para proyectos que buscan adaptabilidad y la capacidad de experimentar con diferentes backends sin comprometer la estabilidad o la eficiencia del desarrollo.

La decisión de utilizar Keras se fundamenta en varias características distintivas de la biblioteca que se alinean estrechamente con los requerimientos del proyecto. En primer lugar, Keras ofrece una variedad de capas prediseñadas, facilitando la experimentación y la rápida iteración al construir la arquitectura del modelo, ya sea mediante la adición o eliminación de capas y neuronas según sea necesario. Además, Keras posee una documentación exhaustiva, especialmente en lo que respecta a la clasificación y regresión de valores numéricos, proporcionando un recurso invaluable para resolver dudas y optimizar el desarrollo. Otro factor determinante es la capacidad de Keras para aprovechar el procesamiento por GPU cuando opera sobre TensorFlow, lo cual es esencial para acelerar significativamente la fase de entrenamiento del modelo.

Aunque PyTorch es otra biblioteca líder para la creación de redes neuronales artificiales, se decidió no utilizarla debido a su enfoque predominante hacia el procesamiento de imágenes. Esto se refleja en su documentación y en la comunidad de usuarios, que está más orientada hacia aplicaciones de visión por computadora, en contraste con la versatilidad de Keras para abordar problemas de regresión y clasificación en dominios no visuales. Por tanto, la elección de Keras, apoyada en TensorFlow, se alinea con nuestros objetivos de construir un modelo eficiente y versátil, capaz de manejar diversas tareas de Machine Learning más allá del procesamiento de imágenes, asegurando así un desarrollo ágil y una implementación efectiva.

5.1.2. Herramientas de diseño de modelo

En esta sección, se detalla el proceso seguido para la elaboración de diversos modelos de pronóstico, cuyo objetivo es generar predicciones de precipitaciones

utilizando como base el modelo estadístico CRHUDA. Para la implementación de estos modelos, se optó por el lenguaje de programación Python, ampliamente reconocido por su eficacia y versatilidad en el ámbito del análisis de datos y el Machine Learning. A lo largo del proceso de desarrollo, se hizo uso de un conjunto de librerías especializadas, seleccionadas por su funcionalidad en diferentes etapas del modelado, desde el manejo de datos hasta la visualización de los resultados.

Las librerías empleadas incluyen:

- *Keras*: Ofrece una interfaz de alto nivel para redes neuronales, permitiendo una rápida experimentación y compatible con TensorFlow, lo que facilita el desarrollo de modelos de aprendizaje profundo (Chollet et al., 2015).
- *TensorFlow*: Proporciona un marco de trabajo exhaustivo y flexible para la construcción y entrenamiento de modelos complejos de Machine Learning, con amplio soporte para cálculos intensivos y capacidad de ejecución en GPUs (Abadi et al., 2016).
- *Scikit-learn (Sklearn)*: Esencial para el procesamiento de datos y análisis predictivo, brinda una extensa gama de algoritmos de Machine Learning listos para usar, desde clasificación y regresión hasta clustering (Pedregosa et al., 2011).
- *Pandas*: Indispensable para la manipulación y el análisis de datos, permite trabajar eficientemente con estructuras de datos tabulares para la preparación de datos (McKinney et al., 2010).
- *NumPy*: Apoya las operaciones matemáticas avanzadas y la manipulación de arrays, fundamental para el procesamiento numérico y análisis de datos (Harris et al., 2020).
- *Matplotlib*: Una librería de gráficos que facilita la visualización de datos, crucial para el análisis exploratorio y la presentación de resultados (Hunter, 2007).

Al emplear estas librerías se genera una plataforma robusta para el desarrollo de modelos predictivos, facilitando desde la manipulación inicial de los datos hasta la evaluación y visualización del rendimiento de los modelos. Este enfoque integrado garantiza un proceso eficiente y efectivo para la predicción de fenómenos meteorológicos, aprovechando la potencia y flexibilidad que el ecosistema de Python ofrece al Machine Learning.

5.2. Algoritmo de pronóstico por medio de RNA de tipo LSTM

Para la construcción de este algoritmo, se eligió trabajar con la librería Keras debido a su enfoque práctico y eficiente para el manejo de redes neuronales. Además, Keras facilita la utilización de GPU para la aceleración hardware en la etapa de entrenamiento, lo cual es vital para procesar grandes volúmenes de datos y complejas arquitecturas de red en tiempos razonables.

El modelo propuesto se basa en la arquitectura de red neuronal “Secuencial”, que simplifica la construcción de modelos al permitir la incorporación secuencial de capas. Esta arquitectura es especialmente adecuada para la experimentación rápida y la iteración en el diseño de redes, ofreciendo una gran flexibilidad para ajustar y optimizar la estructura del modelo.

Para las capas ocultas, se ha optado por utilizar redes neuronales *Long Short-Term Memory* (LSTM). Las LSTM son especialmente eficaces para modelar secuencias temporales gracias a su capacidad para retener información a lo largo del tiempo. A diferencia de las redes neuronales recurrentes estándar, las LSTM pueden manejar dependencias de largo plazo, lo que las hace idóneas para tareas de pronóstico donde el contexto histórico es crucial para la precisión de las predicciones.

El diseño final del modelo incluye una capa de entrada, tres capas ocultas LSTM,

y una capa de salida, estructuradas para optimizar la generación de pronósticos precisos. Este enfoque permite al modelo capturar y procesar efectivamente las dinámicas temporales complejas inherentes a los datos de precipitación. La figura 5.1 muestra una representación gráfica del modelo utilizando el diagrama Fully Connected Neuronal Network (FCNN).

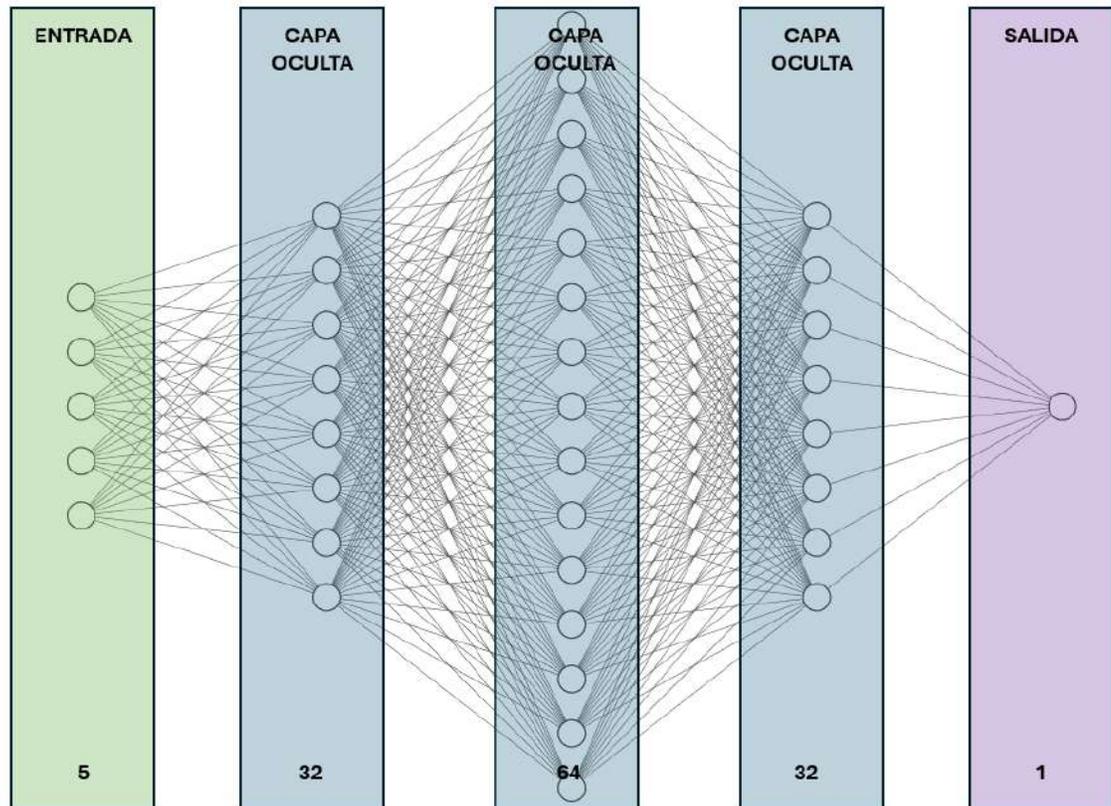


Figura 5.1: *Representación estilo FCNN de la red neuronal artificial.*

La figura anterior ilustra la composición detallada de cada capa del modelo de red neuronal. La capa de entrada del modelo consta de 5 nodos, cada uno correspondiente a una variable de entrada específica. Las capas ocultas, diseñadas para procesar y analizar estas entradas, presentan una distribución variada en términos de nodos: la primera capa oculta tiene 23 nodos, la segunda 64 nodos y la tercera 32 nodos. Por su parte, la capa de salida del modelo se compone de un único nodo, encargado de generar el valor de pronóstico final. El modelo descrito presenta

una complejidad de red neuronal (NN) consistente en 134 nodos. Este modelo es responsable de procesar un total de 4483 parámetros, los cuales están distribuidos entre 4418 coeficientes y 65 sesgos (bias). Esta configuración subraya la capacidad del modelo para manejar y aprender de una cantidad significativa de información a través de su estructura optimizada, lo que le permite generar pronósticos con una alta precisión. La distribución y la cantidad de parámetros reflejan la profundidad y la sofisticación del modelo, destacando su potencial para capturar las complejidades inherentes a los datos de precipitación y traducirlas en predicciones fiables y precisas.

Las variables de entrada elegidas para nutrir el modelo incluyen temperatura, precipitación, humedad, presión atmosférica y punto de rocío. La selección de estas variables está fundamentada en el modelo matemático CRHUDA, destacando su importancia en la comprensión y pronóstico de fenómenos meteorológicos. Esta cuidadosa elección busca abarcar un espectro amplio y representativo de factores climáticos, habilitando al modelo para analizar de forma efectiva los patrones y tendencias inherentes a los datos climatológicos.

Basándose en esta estructura, se desarrollaron múltiples modelos variando las longitudes de datos tanto en la entrada como en la salida. El objetivo de este enfoque es ampliar progresivamente el alcance del pronóstico generado por el modelo, explorando su capacidad para prever condiciones meteorológicas a diferentes intervalos de tiempo futuro. Esta experimentación con la longitud de los datos permite optimizar la precisión del modelo en función del horizonte temporal de predicción deseado, ajustando la arquitectura del modelo para maximizar su rendimiento y fiabilidad.

Tabla 5.2: Resultados de las diversas versiones del modelo con la Arquitectura 1 - LSTM Regresión.

Registros de entrada (minutos)	Registros de salida(minutos pronosticados)	Error (MSE)
30	1	0.00031
60	30	0.00037
240	120	0.00033
480	240	0.00034

5.2.1. Resultados de Modelo

Problemas detectados

Pese a que los resultados son muy favorables existe un inconveniente, el porcentaje de lluvias totales en la muestra usada es en extremo pequeño, ya que únicamente el 0.56 % de los datos poseen registros con lluvias mayores a 0, lo cual significa que el modelo esta teniendo ese valor en el MSE al pronosticar cuando no llueve.

Lo anterior es una problemática bastante importante ya que la proporción tan abismal entre valores con lluvia y sin ella, genera que el generar un pronostico preciso sobre la intensidad de lluvia con amplio margen a futuro sea algo extremadamente complicado o hasta imposible, es por ello que una solución seria realizar una solución diferente, la cual se encuentra descrita en el siguiente capítulo.

Uso previo de las series del modelo CRHUDA para pronostico de precipitaciones

Para la implementación del modelo CRHUDA es necesario el poder pronosticar los valores de las series S1 y S2 es por ello que siguiendo con lo descrito anteriormente, se desarrollo una primer versión de un modelo estadístico de tipo ARMA(p,q) para poder llevar acabo el pronóstico de las series.

Para la implementación del modelo se emplearon datos provenientes de estacio-

nes meteorológicas automáticas de la República Mexicana.

El primer paso en la implementación es analizar la serie de datos, las variables necesarias son el tiempo, contemplando fecha y hora de la medición, presión atmosférica, humedad y punto de rocío.

Para poder generar el modelo ARMA(p,q), es necesario determinar los valores p y q para el modelo, existiendo 2 formas para generarlo, una de ellas es empleando gráficas de correlación según la ecuación 4.3 y correlación parcial. Las gráficas con la correlación y correlación parcial se muestran a continuación.

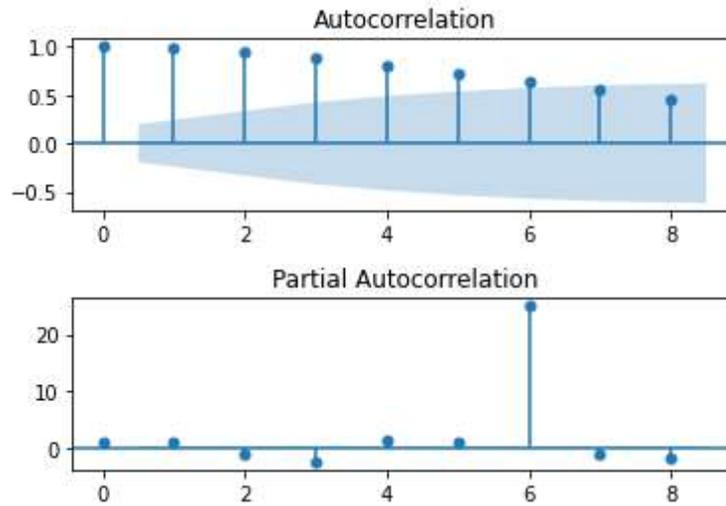


Figura 5.2: Gráfica de Correlación y correlación parcial de la humedad.

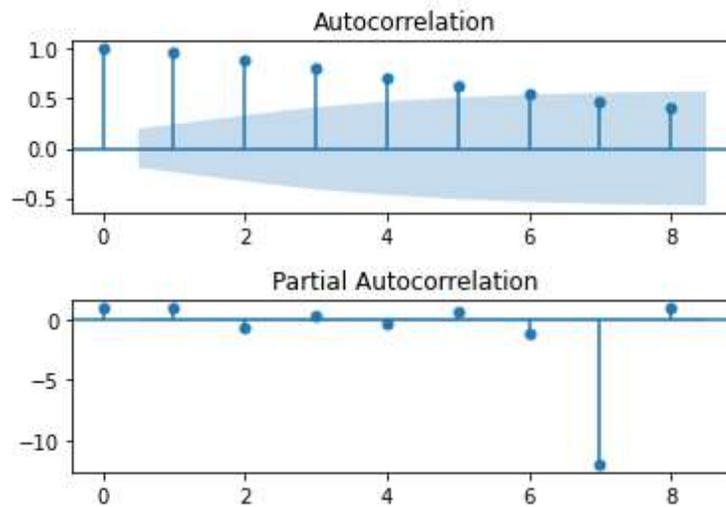


Figura 5.3: Gráfica de Correlación y correlación parcial del punto de rocío.

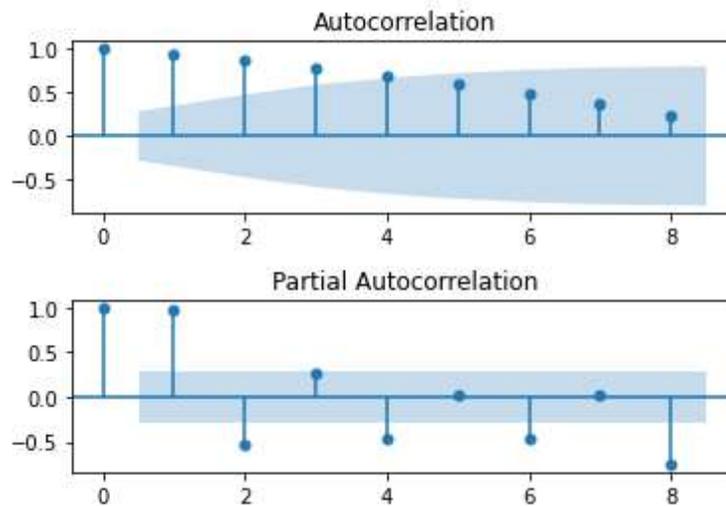


Figura 5.4: Gráfica de Correlación y correlación parcial de la presión atmosférica.

El otro método consiste en emplear la prueba y error para determinar cuáles son los valores más óptimos para el desempeño del modelo ARMA(p,q), midiendo su efectividad con diferentes valores de p y de q, y de esa forma, poder seleccionar los valores con menor tasa de error.

Un factor importante a considerar es la frecuencia de los datos, ya que este factor

puede influir enormemente en cuanto a la efectividad del pronóstico y la ventana de tiempo en la cual el modelo es capaz de predecir. Se consideraron frecuencias de 1 minuto, 10 minutos, 15 minutos y 30 minutos.

En cuanto a la serie de punto de rocío, empleando el método de prueba y error y frecuencias de 1 minuto se determinó que los valores de p y q adecuados correspondieron a un valor de 0 y 1 respectivamente. Para frecuencias de 10 minutos se determinó que los valores de p y q adecuados correspondieron a un valor de 1 y 0 respectivamente. Para frecuencias de 15 minutos se determinó que los valores de p y q adecuados correspondieron a un valor de 8 y 10 respectivamente. Finalmente para frecuencias de 30 minutos se determinó que los valores de p y q adecuados correspondieron a un valor de 1 y 0 respectivamente.

En cuanto a la serie de Humedad, empleando el método de prueba y error y frecuencias de 1 minuto se determinó que los valores de p y q adecuados correspondieron a un valor de 2 y 1 respectivamente. Para frecuencias de 10 minutos se determinó que los valores de p y q adecuados correspondieron a un valor de 1 y 1 respectivamente. Para frecuencias de 15 minutos se determinó que los valores de p y q adecuados correspondieron a un valor de 1 y 1 respectivamente. Finalmente para frecuencias de 30 minutos se determinó que los valores de p y q adecuados correspondieron a un valor de 2 y 1 respectivamente.

En cuanto a la serie de presión atmosférica, empleando el método de prueba y error y frecuencias de 1 minuto se determinó que los valores de p y q adecuados correspondieron a un valor de 1 y 2 respectivamente. Para frecuencias de 10 minutos se determinó que los valores de p y q adecuados correspondieron a un valor de 0 y 1 respectivamente. Para frecuencias de 15 minutos se determinó que los valores de p y q adecuados correspondieron a un valor de 1 y 1 respectivamente. Finalmente para frecuencias de 30 minutos se determinó que los valores de p y q adecuados correspondieron a un valor de 2 y 1 respectivamente.

El siguiente paso es la construcción del modelo ARMA; para simplificar el proceso se empleó una librería en python cuyo enfoque es la creación de modelos ARMA, ARIMA, SARIMA, entre otros modelos estadísticos, el nombre de la librería es statsmodels. Se uso un modelo ARIMA, el cual es el modelo predilecto para el desarrollo de modelos $AR(p)$, $MA(q)$, $ARMA(p,q)$ y $ARIMA(p,d,q)$, basándose únicamente en los valores que recibe como orden.

Al realizar las diferentes implementaciones del modelo ARMA con las diferentes series se obtuvieron diferentes resultados con una tasa de error cuadrático medio. A continuación se muestran diferentes tablas con los resultados de cada serie.

Punto de Rocío			
Frecuencia (minutos)	p	q	RMSE
1	0	1	0.45
10	1	0	1.095
15	8	10	1.178
30	1	0	0.0716

Tabla 5.3: Resultados de modelo ARMA para pronostico de punto de rocío.

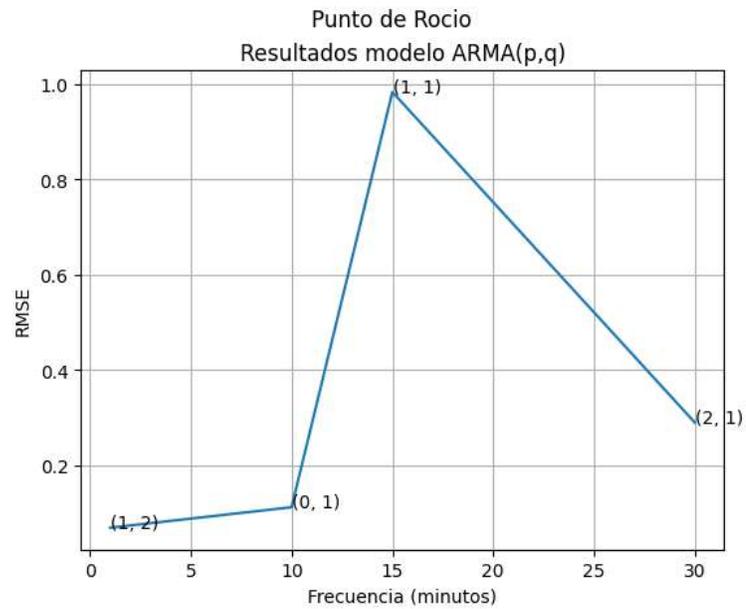


Figura 5.5: Gráfica de resultados de modelo ARMA para pronostico de punto de rocío.

Humedad			
Frecuencia (minutos)	p	q	RMSE
1	2	1	0.892
10	1	1	1.742
15	1	1	3.375
30	2	1	2.737

Tabla 5.4: Resultados de modelo ARMA para pronostico de Humedad.

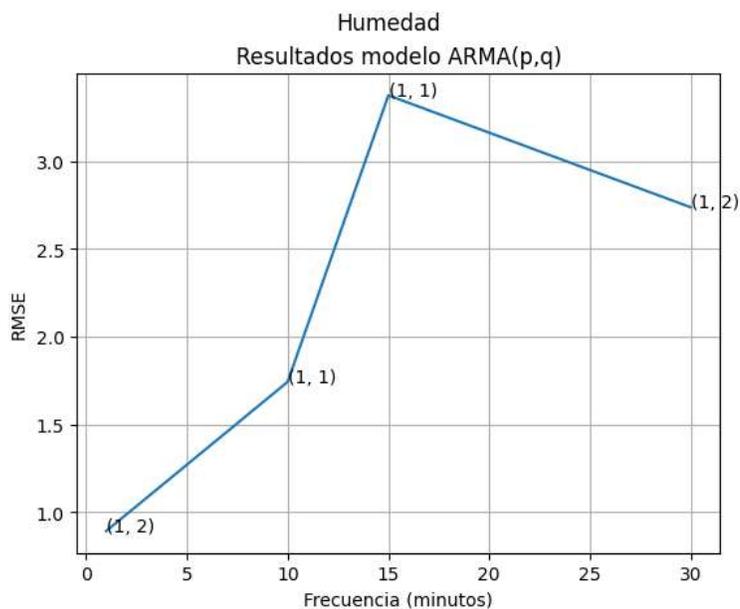


Figura 5.6: Gráfica de resultados de modelo ARMA para pronostico de Humedad.

Presión Atmosférica			
Frecuencia (minutos)	p	q	RMSE
1	1	2	0.069
10	0	1	0.112
15	1	1	0.984
30	2	1	0.290

Tabla 5.5: Resultados de modelo ARMA para pronostico de presión atmosférica.

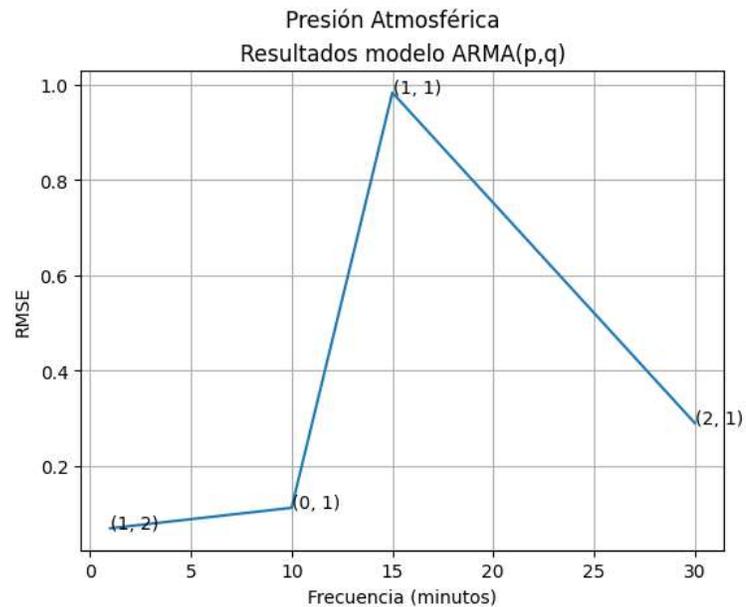


Figura 5.7: Gráfica de resultados de modelo ARMA para pronóstico de Presión Atmosférica.

Los resultados mostrados anteriormente son muy satisfactorios, pero no son útiles para poder pronosticar el inicio de una precipitación, ya que son pronósticos a 1 paso, es decir, pronostican el valor siguiente en la serie de tiempo. Para la implementación del modelo CRHUDA es necesario contar con al menos 8 horas de las series pronosticadas, para lo cual, es necesario generar un modelo ARMA para cada serie con pronósticos a 480 pasos para frecuencias de 1 minuto, 48 pasos para frecuencias de 10, 32 pasos para frecuencias de 15 y 16 pasos para frecuencias de 30 minutos.

A partir de las pruebas hechas con los diferentes modelos ARMA con las series de tiempo, se determinó que la frecuencia más adecuada para el pronóstico es entre 10 y 30 minutos por medición y el factor determinante en cuanto al tiempo adecuado sería que tan afectado se ve el fenómeno del cruce al reducir el tamaño de la muestra, ya que al ampliar el tamaño entre mediciones, la cantidad de datos ingresados al modelo disminuiría.

Diseño de arquitectura mediante implementación de principios arquitectónicos de software.

El desarrollo de modelos de Machine Learning constituye un ciclo iterativo compuesto por múltiples pasos esenciales, diseñados para orientar a científicos de datos y desarrolladores desde una comprensión inicial del problema hasta el despliegue efectivo de un modelo en entornos de producción. Este proceso, intrínsecamente iterativo, permite la mejora continua del modelo, ya sea simplificándolo o aumentando su complejidad, ajustando el número de parámetros, o modificando las variables involucradas y la manera en que estas se introducen en el modelo.

A pesar de las iteraciones y mejoras, es fundamental reconocer que el fenómeno que se busca pronosticar puede ser de tal complejidad que escape a la capacidad de un único modelo para capturar todas sus dinámicas. Por tanto, se hace necesario considerar la exploración de alternativas adicionales y el análisis de distintos enfoques, manteniendo siempre un equilibrio entre la complejidad del modelo y su capacidad para generalizar y hacer predicciones fiables sobre datos no vistos. Este enfoque debe realizarse mediante algún mecanismo que permita obtener resultados

óptimos fragmentando el modelo.

Debido a lo anterior surgió la idea de combinar 2 tópicos que pese a no ser nuevos, no se encontró alguna investigación o trabajo previo donde se usaran al mismo tiempo, siendo estos tópicos la generación de modelos de Machine Learning y el desarrollo de software por medio de principios arquitectónicos como lo son los principios SOLID.

Los principios SOLID constituyen un conjunto de cinco guías de diseño orientado a objetos, ideados para fomentar prácticas de desarrollo de software más eficientes y sostenibles. Acuñados y popularizados por Robert C. Martin (Martin, 2003), estos principios buscan facilitar la creación de sistemas de software que sean fáciles de mantener, ampliar y comprender. El acrónimo **SOLID** se desglosa en:

1. ***Single Responsibility Principle (Principio de Responsabilidad Única)***: Postula que una clase debe tener una sola razón para cambiar, promoviendo la modularidad y la cohesión.
2. ***Open/Closed Principle (Principio de Abierto/Cerrado)***: Indica que las entidades de software deben estar abiertas para la extensión pero cerradas para la modificación, favoreciendo la escalabilidad y la reusabilidad.
3. ***Liskov Substitution Principle (Principio de Sustitución de Liskov)***: Establece que objetos de una superclase deben poder ser sustituidos por objetos de una subclase sin afectar la corrección del programa, asegurando la interoperabilidad y la sustituibilidad.
4. ***Interface Segregation Principle (Principio de Segregación de Interfaz)***: Aconseja que ninguna clase debería depender de métodos que no utiliza, evitando acoplamientos innecesarios y mejorando la claridad del diseño.
5. ***Dependency Inversion Principle (Principio de Inversión de Dependencias)***: Sugiere que los módulos de alto nivel no deben depender de módulos de bajo

nivel, sino que ambos deben depender de abstracciones, contribuyendo a la flexibilidad y la desacoplabilidad del sistema.

Al seguir estos principios, permite a los desarrolladores construir sistemas más robustos, flexibles y mantenibles, facilitando así el crecimiento y la adaptación del software a lo largo del tiempo.

6.1. Rediseño del algoritmo de pronóstico mediante principios arquitectónicos

Para el rediseño del algoritmo, se tomó la decisión de aplicar específicamente el **Principio de Responsabilidad Única (SRP)** junto con el **Principio de Separación de Preocupaciones (SoC)** descritos en (Martin, 2003) y (Blancarte, 2020). La adopción de estos principios facilita la descomposición del problema en múltiples componentes, cada uno enfocado en una tarea específica, previniendo la sobrecarga de un único modelo con múltiples responsabilidades.

La propuesta contempla el desarrollo de una arquitectura modular, donde cada módulo o capa se encarga de un aspecto distinto del proceso de predicción. Esta aproximación no solo simplifica la gestión y mantenimiento del sistema, sino que también mejora su escalabilidad y flexibilidad. Al dividir el problema en diferentes capas y modelos, es posible especializar cada uno en aspectos particulares del conjunto de datos o en tipos específicos de predicción, como modelos dedicados exclusivamente a la predicción de eventos de lluvia y otros enfocados en variables como la temperatura o la presión atmosférica.

Este enfoque modular, respaldado por el SRP y el SoC, permite optimizar cada modelo para su tarea específica, lo que puede llevar a mejoras significativas en la precisión y el rendimiento general del sistema. Además, facilita la implementación de mejoras o cambios en segmentos específicos del sistema sin impactar el resto de la arquitectura, promoviendo un desarrollo ágil y sostenible a largo plazo.

6.1.1. Resultado de Rediseño

Tras analizar diversas soluciones, se optó por desarrollar una arquitectura compuesta por varios modelos en lugar de uno solo, permitiendo así dividir las responsabilidades eficazmente. La solución final incluye un modelo capaz de pronosticar la ocurrencia de lluvia y un segundo modelo que, una vez determinada la probabilidad de precipitación, puede pronosticar su intensidad.

Esta solución surge del análisis de la problemática desde una perspectiva de ingeniería de software, aplicando principios recomendados para un desarrollo de software de alta calidad. El **Principio de Separación de Preocupaciones (SoC)**, un principio fundamental del diseño de software, propone que un sistema debe dividirse en componentes independientes, cada uno centrado en una tarea específica. Esto permite a los desarrolladores trabajar en distintas partes del sistema de manera concurrente, sin interferir entre sí.

SoC implica la separación lógica del sistema en distintas capas o componentes, cada uno con sus responsabilidades específicas. Estos se comunican a través de interfaces bien definidas, facilitando la independencia y la reutilización. La implementación de SoC mejora la calidad del software, la eficiencia en el desarrollo y simplifica el mantenimiento y evolución del sistema, como se evidencia en la separación entre la lógica de negocio, la interfaz de usuario, la capa de acceso a datos, y entre el núcleo del sistema y las herramientas de monitoreo y supervisión.

Al analizar el problema se determinó que el modelo anterior intentaba realizar más de 2 tareas a la vez, es decir tenía "2 preocupaciones" ya que el modelo pretendía pronosticar si llovería y la intensidad de la misma, dejando como una solución más adecuada el dividir la problemática en 2 modelos diferentes, simplificando la complejidad de la solución y permitiendo solucionar el principal problema del modelo anterior.

La arquitectura propuesta para el sistema de pronóstico se estructura alrededor de tres modelos diferenciados, diseñados para trabajar de manera secuencial y

complementaria.

El primer modelo actúa como un regresor, encargado de generar pronósticos para series de datos específicas: presión atmosférica, punto de rocío y humedad. Estas series son esenciales para alimentar el modelo CRHUDA, tal como se detalla en el capítulo 4. Este modelo regresor establece las bases para las predicciones subsiguientes, proporcionando las entradas necesarias para el análisis posterior.

El segundo modelo, un clasificador binario, recibe como entrada las series generadas por el primer modelo. Su función es determinar, basado en la adaptación del modelo CRHUDA, si los valores actuales y proyectados de las series indican la probabilidad de lluvia en un futuro cercano. Este paso es crucial para filtrar los escenarios que requieren una atención detallada respecto a la precipitación.

Finalmente, el tercer modelo se activa solo si el clasificador binario anticipa la ocurrencia de lluvia. Este modelo está especializado en la regresión de la intensidad de la lluvia, con el objetivo de proporcionar predicciones precisas sobre la cantidad de precipitación esperada. La implementación de este modelo es fundamental para integrar un sistema de alerta temprana efectivo, capaz no solo de predecir la lluvia sino también de indicar su intensidad, lo que es vital para la toma de decisiones y la preparación ante eventos meteorológicos adversos.

La arquitectura completa está visualmente representada en la figura 6.1, ofreciendo una perspectiva clara de cómo cada componente interactúa dentro del sistema de predicción propuesto.

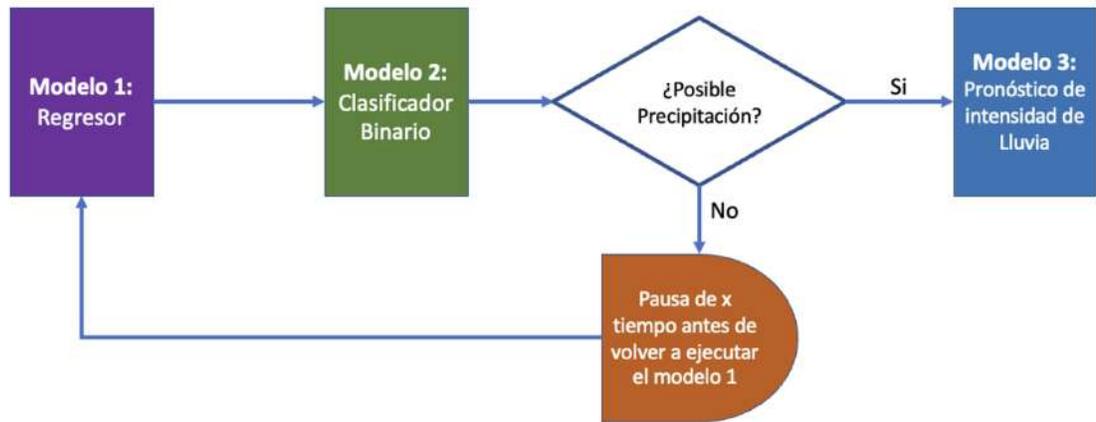


Figura 6.1: *Diagrama conceptual de la arquitectura para pronóstico de lluvias.*

Discusión

7.1. Discusión

7.1.1. Uso de modelo CRHUDA y análisis de sensibilidad

En este punto se ha probado que el modelo CRHUDA que se diseñó para encontrar el efecto de la relación Clausius-Clapeyron (C-C) y la combinación de la presión atmosférica, el punto de rocío y la humedad; es capaz de predecir el inicio de la precipitación con gran precisión. Estas variables ya han sido utilizadas, pero de forma individual y en modelos meteorológicos Rogers et al., 2007; Sim et al., 2019a; Velasco y Fernández-Pineda, 2008. Sin embargo, los resultados son coherentes con los de otros estudios que sugieren que la dependencia de la temperatura del punto de rocío de la superficie sigue dos veces la relación C-C, apoyados por el simple argumento físico de que este comportamiento C-C surge de la física de las nubes de convección (Danladi et al., 2018; Lenderink, Barbero, Loriaux & Fowler, 2017). De esta forma el CRHUDA, como un simple modelo sinóptico, puede predecir el inicio de la precipitación de manera fiable. En promedio nueve a diez horas antes del inicio de la precipitación; el pronóstico parece aceptable. Esto también concuerda con nuestras observaciones anteriores que mostraron que es posible utilizar el

conjunto de datos hidrometeorológicos en tiempo real minuto a minuto (Li & Lai, 2004; Moon et al., 2019a). Por lo tanto, se puede sugerir que una base de datos de precipitaciones diarias permite la correcta desagregación espacio-temporal de las precipitaciones horarias distribuidas espacialmente (Dyson et al., 2015; Segond et al., 2006). Este hallazgo tiene importantes implicaciones para el desarrollo de un modelo de previsión de precipitaciones para proporcionar previsiones con antelación, ya que sólo es necesario encontrar la combinación adecuada de variables climáticas (Berg & Haerter, 2013; Park & Min, 2017a).

7.1.2. Algoritmo de Redes Neuronales Artificiales de tipo LSTM

Tras confirmar la eficacia del modelo CRHUDA en la predicción del comienzo de eventos de precipitación, se podría suponer que la investigación había alcanzado un punto culminante. No obstante, el desarrollo de un modelo predictivo utilizando Redes Neuronales Artificiales de tipo Long Short-Term Memory (LSTM) reveló que el camino hacia una predicción precisa aún enfrentaba desafíos significativos. Aunque este modelo demostró ser altamente preciso, la discrepancia entre los pronósticos generados y los resultados esperados era mínima. Este margen estrecho de variación representa una barrera para la optimización del modelo, impidiendo mejoras sustanciales en la calidad de las predicciones. Además, se observó que al predecir la intensidad de la lluvia en situaciones de ausencia de precipitaciones anticipando cero lluvia se inflaba artificialmente la tasa de éxito del modelo, debido a que un alto porcentaje de las predicciones coincidían con la realidad simplemente por pronosticar la no ocurrencia de lluvia.

Aunque estos resultados preliminares fueron prometedores, no proporcionaron una solución viable para predecir el inicio de la precipitación con la precisión deseada. La naturaleza de los pronósticos generados por el modelo LSTM, que se limitan a predecir el valor inmediatamente siguiente en la secuencia de tiempo, resulta insuficiente para las necesidades del modelo CRHUDA. Este último requiere, para

su correcta implementación, pronósticos que abarquen al menos 8 horas de datos futuros.

7.1.3. Arquitectura para pronóstico de precipitaciones

A partir de lo visto en el capítulo 6, es evidente la necesidad de desarrollar una arquitectura robusta para la construcción de modelos de machine learning, subrayando la importancia de aplicar principios bien establecidos en la arquitectura de software. Dado que la construcción de modelos de machine learning está intrínsecamente vinculada al desarrollo de software, la aplicación de principios como el *Single Responsibility Principle* (SRP) y la *Separation of Concerns* (SoC) resulta fundamental para lograr un diseño óptimo. El concepto de adaptar etapas, técnicas o procesos vistos en la Ingeniería de Software al desarrollo de Machine Learning no es nuevo, pues podemos ver un ejemplo muy notorio en el Machine Learning Operations, siendo este una adaptación de la metodología convencional de DevOps.

Al adherirse al principio de SoC, se facilita la organización general y la segregación de los componentes dentro del sistema de pronóstico, garantizando que cada módulo se concentre en una única preocupación. Esto no solo permite abordar cada etapa del proceso de pronóstico de manera más efectiva sino que también promueve soluciones óptimas y especializadas para cada problema específico. En paralelo, la aplicación del SRP dentro de estos módulos asegura que cada clase o componente maneje una única responsabilidad, evitando así la complejidad y el acoplamiento excesivo.

La sinergia de estos principios arquitectónicos da lugar a un sistema de pronóstico altamente flexible y eficiente. Esta aproximación facilita la integración y gestión de múltiples modelos de manera efectiva, ofreciendo una plataforma que es tanto mantenible como escalable. Por lo tanto, la adopción de prácticas sólidas en la arquitectura de software no solo mejora la estructura del sistema de pronóstico sino que también potencia su capacidad para evolucionar y adaptarse a nuevas necesidades

o desafíos futuros.

7.2. Publicaciones

A lo largo del desarrollo de la presente investigación se obtuvieron los siguientes productos académicos:

- Publicación en la revista STATS en su volumen 7 del artículo *Sensitivity Analysis of Start Point of Extreme Daily Rainfall Using CRHUDA and Stochastic Models*, participando como autor principal.
- Publicación en la revista NTHE en su número especial de Vulnerabilidad Urbana, con el artículo *Modelo y algoritmo de pronóstico CRHUDA aplicado a las estaciones de la red de monitoreo RedCIAQ*, participando como primer autor.
- Publicación en la revista NTHE en su número especial de Vulnerabilidad Urbana, con el artículo *Construcción de herramienta para el manejo y consulta de datos climatológicos en la investigación de vulnerabilidades.*, participando como coautor.
- Publicación en la revista NTHE en su número especial de Vulnerabilidad Urbana, con el artículo *Propuesta normativa para considerar la vulnerabilidad urbana ante inundaciones*, participando como coautor.
- Participación en el 2° Congreso Internacional en Computación y Tecnología Educativa como ponente.
- Participación en el 1° Congreso de Gestión Integral de Recursos Hídricos como ponente.
- Participación en el 2° Congreso Nacional de Riesgos Hidrogeológicos como ponente.

Es importante destacar que como parte de mi formación integral como investigador se generaron diversos productos como ponencias, artículos y participación en eventos fuera del tópico principal de este trabajo. Siendo uno de los mas importantes la participación en las XIII Jornadas Internacionales de CAMPUS VIRTUALES donde se obtuvo un premio al mejor paper/ponencia en cuanto a inteligencia artificial en la educación, demostrando la habilidad de no solo trabajar en el tema presentado.

7.3. Trabajo Futuro y líneas de investigación

Gracias a este trabajo de investigación surgen diversas líneas de investigación que pueden continuar desarrollándose, ya que existe una enorme cantidad de trabajo pendiente de realizar. En primer instancia se encuentra el trabajar con el pronóstico de precipitaciones y demás variables climatológicas de una zona, con la intención de poder realizar un sistema de alerta temprana funcional capaz de operar en el estado de Querétaro.

Otra de las líneas de investigación muy interesantes que surgen gracias a los resultados y discusión de la misma, es el trabajar en el diseño y desarrollo de arquitecturas de inteligencia artificial implementando y adaptando principios de arquitectura de software, tal como lo es los principios SOLID o algún otro.

Conclusiones

Después de un análisis exhaustivo de los resultados obtenidos en sucesivas iteraciones durante el desarrollo del algoritmo, hemos llegado a la confirmación de nuestra hipótesis inicial. La realización de un análisis de sensibilidad en profundidad aplicado al modelo CRHUDA ha revelado que, de hecho, es factible su uso para la predicción del inicio de eventos de precipitación, siempre que se ajuste adecuadamente el factor de escala β . Esta investigación resalta la precisión y viabilidad del modelo CRHUDA como una herramienta para pronostico potente, capaz de anticipar fenómenos meteorológicos complejos con una anticipación considerable, lo cual abre diferentes oportunidades para investigar en el campo de la hidroinformática.

Las preguntas que se tenía como fin el contestarlas por medio de esta investigación se muestran a continuación junto a su respuesta.

1. ¿Es posible usar el modelo CRHUDA para generar un algoritmo capaz de pronosticar en tiempo real el momento de inicio de una precipitación? **En efecto, es posible pronosticar el momento exacto mediante el uso del modelo CRHUDA siempre y cuando se cuente con el pronostico de las 3 series involucradas en dicho modelo.**
2. ¿Existe la posibilidad de optimizar o mejorar el modelo CRHUDA para obtener un mejor algoritmo de implementación? ¿Qué métodos de optimización

podrían ser utilizados? **El modelo puede ser mejorado e implementado con el diseño de una arquitectura en lugar de usar un solo modelo, para ello se considera la adaptación de diversos patrones arquitectónicos de Software tal como lo son los principios SOLID y SoC.**

3. ¿Con cuanta anticipación y exactitud puede pronosticar precipitaciones el algoritmo que se pretende desarrollar? **Como bien fue mencionado en el capítulo 4, al implementar el modelo CRHUDA se pudo encontrar que el tiempo promedio entre la alerta y el inicio de una precipitación se mantiene aceptable en el promedio de 10 horas y con la exactitud del pronostico dependiendo directamente el factor de escalamiento β .**
4. ¿Qué tipo de precipitaciones pueden pronosticarse con un menor error? (convectiva, huracanada o ambas) **Debido a la magnitud y elevada dificultad del fenómeno investigado, únicamente se pudieron realizar pruebas sobre precipitaciones convectivas, quedando como un posible trabajo futuro su aplicación sobre precipitaciones huracanadas.**
5. ¿Qué tan complicado será el replicar el algoritmo propuesto para pronosticar precipitaciones, incluso en lugares donde no se tengan datos históricos de precipitaciones? **Siempre y cuando se cuente con la capacidad de pronosticar las variables involucradas en el modelo CRHUDA, es posible implementar el modelo, sin embargo es necesario desarrollar mas investigación en cuanto a su adaptación a diferentes regiones fuera del estado de Querétaro, para poder encontrar de forma adecuada el factor de escalamiento β aplicable para esa ubicación.**

Aunque esta investigación haya llegado a su fin, en realidad existe una gran cantidad de trabajo e investigación relacionada, ya que la incorporación del modelo CRHUDA en un sistema de alerta temprana eficiente enfrenta aún importantes desafíos. Principalmente debido a la complejidad asociada a la predicción de las

series temporales requeridas. La habilidad para predecir con exactitud el inicio de la precipitación depende de la comprensión y modelado preciso de un conjunto de variables meteorológicas de humedad, presión atmosférica y punto de rocío en intervalos bastante elevados, lo cual representa una tarea computacionalmente exigente y metodológicamente sofisticada. Este obstáculo destaca la necesidad de avanzar en el desarrollo de metodologías de modelado más robustas y en la implementación de tecnologías de procesamiento de datos más avanzadas, que permitan superar las limitaciones actuales y faciliten la integración del modelo CRHUDA en sistemas de alerta temprana que contribuyan significativamente a la prevención de desastres naturales y la protección de vidas y bienes.

Bibliografía

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning. *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 265-283.
- Abu-Mostafa, Y. S., & Atiya, A. F. (1996). Introduction to financial forecasting. *Applied intelligence*, 6, 205-213.
- Acharya, R., Pal, J., Das, D., & Chaudhuri, S. (2019). Long-range forecast of Indian summer monsoon rainfall using an artificial neural network model. *Meteorological Applications*, 26(3), 347-361.
- Agard, V., & Emanuel, K. (2017a). Clausius–Clapeyron Scaling of Peak CAPE in Continental Convective Storm Environments. *Journal of the Atmospheric Sciences*, 74(9), 3043-3054. <https://doi.org/10.1175/jas-d-16-0352.1>
- Agard, V., & Emanuel, K. (2017b). Clausius–Clapeyron scaling of peak CAPE in continental convective storm environments. *Journal of the Atmospheric Sciences*, 74(9), 3043-3054.
- Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., Ballas, N., Bastien, F., Bayer, J., Belikov, A., Belopolsky, A., et al. (2016). Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, arXiv-1605.

- Azimi, S., & Moghaddam, M. A. (2020). Modeling Short Term Rainfall Forecast Using Neural Networks, and Gaussian Process Classification Based on the SPI Drought Index. *Water Resources Management*, 1-37.
- Azli, M., & Rao, A. R. (2010). Development of Huff curves for peninsular Malaysia. *Journal of hydrology*, 388(1-2), 77-84.
- Berg, P., & Haerter, J. O. (2013). Unexpected increase in precipitation intensity with temperature -A result of mixing of precipitation types? *Atmospheric Research*, 119, 56-61. <https://doi.org/10.1016/j.atmosres.2011.05.012>
- Blancarte, O. (2020). *Introducción a la arquitectura de Software*. Oscar Javier Blancarte Iturral.
- Bonta, J., & Shahalam, A. (2003). Cumulative storm rainfall distributions: comparison of Huff curves. *Journal of Hydrology (New Zealand)*, 65-74.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. Holden-Day.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Bugaets, A., Gartsman, B., Gelfan, A., Motovilov, Y., Sokolov, O., Gonchukov, L., Kalugin, A., Moreido, V., Suchilina, Z., & Fingert, E. (2018). The integrated system of hydrological forecasting in the Ussuri River basin based on the ECOMAG model. *Geosciences*, 8(1), 5.
- Bürger, G., Heistermann, M., & Bronstert, A. (2014). Towards subdaily rainfall disaggregation via Clausius–Clapeyron. *Journal of Hydrometeorology*, 15(3), 1303-1311.
- Calvello, M., d'Orsi, R. N., Piciullo, L., Paes, N., Magalhaes, M., & Lacerda, W. A. (2015). The Rio de Janeiro early warning system for rainfall-induced landslides: analysis of performance for the years 2010–2013. *International journal of disaster risk reduction*, 12, 3-15.
- Camuffo, D. (2014). Theoretical Grounds for Humidity. En *Microclimate for Cultural Heritage*. <https://doi.org/10.1016/b978-0-444-63296-8.00002-0>

- Cano, J. (2017). Por inundaciones, \$64 mil millones perdidos – AMQueretaro.com. Consultado el 26 de junio de 2018, desde <http://amqueretaro.com/queretaro/2017/10/03/por-inundaciones-64-mil-millones-perdidos>
- Capparelli, A. (2013). Físicoquímica básica. *Series: Libros de Cátedra*.
- Cargua, S., Cueva Rosillo, M. E., Escobar, J., Arciniegas, S., & Nieto, E. (2017). Recurso hídrico, clima y sistemas de información geográfica.
- Chang, W., Stein, M. L., Wang, J., Kotamarthi, V. R., & Moyer, E. J. (2016). Changes in spatiotemporal precipitation patterns in changing climate conditions. *Journal of Climate*, 29(23), 8355-8376.
- Chollet, F., et al. (2015). Keras.
- Clements, M. P., & Hendry, D. F. (2002). Modelling methodology and forecast failure. *The Econometrics Journal*, 5(2), 319-344.
- Corres, G., Esteban, A., García, J. C., & Zárata, C. (2009). Análisis de series temporales. *Revista Ingeniería Industrial*, 8(1).
- Dahm, R., Bhardwaj, A., Sperna Weiland, F., Corzo, G., & Bouwer, L. M. (2019). A Temperature-Scaling Approach for Projecting Changes in Short Duration Rainfall Extremes from GCM Data. *Water*, 11(2), 313.
- Damrath, U., Doms, G., Fruhwald, D., Heise, E., Richter, B., & Steppeler, J. (2000). Operational quantitative precipitation forecasting at the German Weather Service. *Journal of Hydrology*, 239(1-4), 260-285.
- Danladi, A., Stephen, M., Aliyu, B. M., Gaya, G. K., Silikwa, N. W., & Machael, Y. (2018). Assessing the influence of weather parameters on rainfall to forecast river discharge based on short-term. *Alexandria Engineering Journal*, 57(2), 1157-1162. <https://doi.org/10.1016/j.aej.2017.03.004>
- de Benito Crosetti, B., & Ibáñez, J. M. S. (2016). La investigación basada en diseño en Tecnología Educativa. *RiiTE Revista Interuniversitaria de Investigación en Tecnología Educativa*.

- Dyson, L. L., Van Heerden, J., & Sumner, P. D. (2015). A baseline climatology of sounding-derived parameters associated with heavy rainfall over Gauteng, South Africa. *International Journal of Climatology*, 35(1), 114-127.
- Easterling, W. E., & Stern, P. C. (Eds.). (1999). *Making climate forecasts matter*. National Academies Press.
- Fuentes, A., Sierra, M., & Morfa, Y. (2020). Corrección del pronóstico cuantitativo de la precipitación mediante el uso de redes neuronales. *Revista Cubana de Meteorología*, 26(3).
- Gao, X., Li, J., & Sorooshian, S. (2007). Modeling intraseasonal features of 2004 North American monsoon precipitation. *Journal of climate*, 20(9), 1882-1896.
- Gevorkyan, M. N., Demidova, A. V., Demidova, T. S., & Sobolev, A. A. (2019). Review and comparative analysis of machine learning libraries for machine learning. *Discrete and Continuous Models and Applied Computational Science*, 27(4), 305-315.
- González Casimiro, M. P. (2009). Análisis de series temporales: Modelos ARIMA.
- Gutierrez-Lopez, A. (2021). A Robust Gaussian variogram estimator for cartography of hydrological extreme events. *Natural Hazards*, 107. <https://doi.org/10.1007/s11069-021-04641-9>
- Gutierrez-Lopez, A., Cruz-Paz, I., & Muñoz Mandujano, M. (2019a). Algorithm to Predict the Rainfall Starting Point as a Function of Atmospheric Pressure, Humidity, and Dewpoint. *Climate*, 7(11), 131. <https://doi.org/10.3390/cli7110131>
- Gutierrez-Lopez, A., Cruz-Paz, I., & Muñoz Mandujano, M. (2019b). Algorithm to Predict the Rainfall Starting Point as a Function of Atmospheric Pressure, Humidity, and Dewpoint. *Climate*, 7(11), 131.
- Gutiérrez-López, A., & Barragán-Regalado, R. (2019). Ajuste de curvas IDF a partir de tormentas de corta duración. *Tecnología y ciencias del agua*, 10(6), 1-24.
- Gutiérrez-López, A., Fortanell Trejo, M., Albuquerque Gonzalez, N. I., & Bravo Prado, F. (2019). Análisis de la variabilidad espacial en la precipitación en la zona

- metropolitana de Querétaro empleando ecuaciones de anisotropía. *Investigaciones geográficas*, (99).
- Harder, P., & Pomeroy, J. (2013a). Estimating precipitation phase using a psychrometric energy balance method. *Hydrological Processes*, 27(13), 1901-1914. <https://doi.org/10.1002/hyp.9799>
- Harder, P., & Pomeroy, J. (2013b). Estimating precipitation phase using a psychrometric energy balance method. *Hydrological Processes*, 27(13), 1901-1914.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585, 357-362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hernández-Leal, E. J., Duque-Méndez, N. D., & Moreno-Cadavid, J. (2016). Generación de pronósticos para la precipitación diaria en una serie de tiempo de datos meteorológicos. *Ingenio Magno*, 7(1), 144-155.
- Hofmann, J., & Schüttrumpf, H. (2019). Risk-based early warning system for pluvial flash floods: Approaches and foundations. *Geosciences*, 9(3), 127.
- Holley, D., Dorling, S., Steele, C., & Earl, N. (2014). A climatology of convective available potential energy in Great Britain. *International journal of climatology*, 34(14), 3811-3824.
- Hou, T., Kong, F., Chen, X., & Lei, H. (2013). Impact of 3DVAR data assimilation on the prediction of heavy rainfall over Southern China. *Advances in Meteorology*, 2013.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(3), 90-95.
- Hurtado, C., & Ríos, G. (2008). *Series de tiempo*. Universidad de Chile.
- Ian, Yoshua, A. C. (2015). Deep Learning. *Deep Learning*. <https://doi.org/10.1016/B978-0-12-391420-0.09987-X>

- Índice de peligro por inundación (IPI)*. (2017). <http://www.atlasmnacionalderiesgos.gob.mx/descargas/Metodologias/Inundacion.pdf>
- Jain, H., & Jain, R. (2017). Big data in weather forecasting: Applications and challenges. *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, 138-142. <https://doi.org/10.1109/ICBDACI.2017.8070824>
- Lenderink, G., Barbero, R., Loriaux, J. M., & Fowler, H. J. (2017). Super-Clausius–Clapeyron Scaling of Extreme Hourly Convective Precipitation and Its Relation to Large-Scale Atmospheric Conditions. *Journal of Climate*, *30*(15), 6037-6052. <https://doi.org/10.1175/jcli-d-16-0808.1>
- Lenderink, G., Barbero, R., Loriaux, J., & Fowler, H. (2017). Super-Clausius–Clapeyron scaling of extreme hourly convective precipitation and its relation to large-scale atmospheric conditions. *Journal of Climate*, *30*(15), 6037-6052.
- Li, P., & Lai, E. S. (2004). Short-range quantitative precipitation forecasting in Hong Kong. *Journal of Hydrology*, *288*(1-2), 189-209.
- Liu, Z., Chen, B., Chan, S. T., Cao, Y., Gao, Y., Zhang, K., & Nichol, J. (2015). Analysis and modelling of water vapour and temperature changes in Hong Kong using a 40-year radiosonde record: 1973–2012. *International Journal of Climatology*, *35*(3), 462-474.
- Lombardi, G., Ceppi, A., Ravazzani, G., Davolio, S., & Mancini, M. (2018). From Deterministic to Probabilistic Forecasts: The ‘Shift-Target’ Approach in the Milan Urban Area (Northern Italy). *Geosciences*, *8*(5), 181.
- Lorenz, D. J., & DeWeaver, E. T. (2007). The response of the extratropical hydrological cycle to global warming. *Journal of climate*, *20*(14), 3470-3484.
- Maíllo, S. V., & Fernández-Pineda, C. (2011). Sobre la obtención de la ecuación de Clapeyron-Clausius. *Revista Española de Física*, *22*(4).
- Martin, R. C. (2003). *Agile software development: principles, patterns, and practices*. Prentice Hall PTR.

- McKinney, W., et al. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 445, 51-56.
- Millán, H., Ghanbarian-Alavijeh, B., & García-Fornaris, I. (2010a). Nonlinear dynamics of mean daily temperature and dewpoint time series at Babolsar, Iran, 1961–2005. *Atmospheric Research*, 98(1), 89-101.
- Millán, H., Ghanbarian-Alavijeh, B., & García-Fornaris, I. (2010b). Nonlinear dynamics of mean daily temperature and dewpoint time series at Babolsar, Iran, 1961–2005. *Atmospheric Research*, 98(1), 89-101. <https://doi.org/10.1016/j.atmosres.2010.06.001>
- Mohr, S., & Kunz, M. (2013). Recent trends and variabilities of convective parameters relevant for hail events in Germany and Europe. *Atmospheric Research*, 123, 211-228.
- Moon, S. H., Kim, Y. H., Lee, Y. H., & Moon, B. R. (2019a). Application of machine learning to an early warning system for very short-term heavy rainfall. *Journal of Hydrology*. <https://doi.org/10.1016/j.jhydrol.2018.11.060>
- Moon, S. H., Kim, Y. H., Lee, Y. H., & Moon, B. R. (2019b). Application of machine learning to an early warning system for very short-term heavy rainfall. *Journal of Hydrology*. <https://doi.org/10.1016/j.jhydrol.2018.11.060>
- Muñoz-Mandujano, M., Gutiérrez-López, M. A., & Vargas-Díaz, J. A. (2022). Modelo y algoritmo de pronóstico CRHUDA aplicado a las estaciones de la red de monitoreo RedCIAQ. *NTHE*, 79-92. https://www.researchgate.net/profile/Martin-Munoz-Mandujano/publication/359770233_Modelo_y_algoritmo_de_pronostico_CRHUDA_aplicado_a_las_estaciones_de_la_red_de_monitoreo_RedCIAQ/links/624db6a5b0cee02d6954fdda/Modelo-y-algoritmo-de-pronostico-CRHUDA-aplicado-a-las-estaciones-de-la-red-de-monitoreo-RedCIAQ.pdf
- Muñoz-Mandujano, M., Gutierrez-Lopez, A., Acuña-Garcia, J. A., Ibarra-Corona, M. A., Aguilar, I. C., & Vargas-Diaz, J. A. (2024). Sensitivity Analysis of Start Point of Extreme Daily Rainfall Using CRHUDA and Stochastic Models. *Stats*, 7(1), 160-171. <https://doi.org/10.3390/stats7010010>

- Myoung, B., & Nielsen-Gammon, J. W. (2010). Sensitivity of monthly convective precipitation to environmental conditions. *Journal of Climate*, *23*(1), 166-188.
- Park, I.-H., & Min, S.-K. (2017a). Role of Convective Precipitation in the Relationship between Subdaily Extreme Precipitation and Temperature. *Journal of Climate*, *30*(23), 9527-9537. <https://doi.org/10.1175/jcli-d-17-0075.1>
- Park, I.-H., & Min, S.-K. (2017b). Role of convective precipitation in the relationship between subdaily extreme precipitation and temperature. *Journal of Climate*, *30*(23), 9527-9537.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. En *Advances in Neural Information Processing Systems 32* (pp. 8024-8035). Curran Associates, Inc. <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*(Oct), 2825-2830.
- Peleg, N., Marra, F., Fatichi, S., Molnár, P., Morin, E., Sharma, A., & Burlando, P. (2018). Intensification of Convective Rain Cells at Warmer Temperatures Observed from High-Resolution Weather Radar Data. *Journal of Hydrometeorology*, *19*(4), 715-726. <https://doi.org/10.1175/jhm-d-17-0158.1>
- Peleg, N., Marra, F., Fatichi, S., Molnar, P., Morin, E., Sharma, A., & Burlando, P. (2018). Intensification of convective rain cells at warmer temperatures observed from high-resolution weather radar data. *Journal of Hydrometeorology*, *19*(4), 715-726.

- Press, G. (2016). Cleaning big data: Most time-consuming, least enjoyable data science task, survey says. *Forbes, March, 23*, 15.
- Pumo, D., Carlino, G., Blenkinsop, S., Arnone, E., Fowler, H., & Noto, L. V. (2019). Sensitivity of extreme rainfall to temperature in semi-arid Mediterranean regions. *Atmospheric Research, 225*, 30-44.
- Rasouli, K., Hsieh, W. W., & Cannon, A. J. (2012). Daily streamflow forecasting by machine learning methods with weather and climate inputs. *Journal of Hydrology, 414*, 284-293.
- Rivas-Asanza, W., Mazon-Olivo, B., & Mejia, F. (2018). Capítulo 1: Generalidades de las redes neuronales artificiales.
- Rivera, E. (2005). *Introducción a las Redes Neuronales Artificiales*. Editorial Universidad Don Bosco. <http://rd.udb.edu.sv:8080/jspui/handle/11715/238>
- Rogers, J. C., Wang, S.-H., & Coleman, J. S. (2007). Evaluation of a long-term (1882–2005) equivalent temperature time series. *Journal of climate, 20*(17), 4476-4485.
- Rojas Vilches, O., & Martínez Reyes, C. (2011). Riesgos naturales: evolución y modelos conceptuales. *Revista Universitaria de Geografía*.
- Romps, D. M. (2014). An Analytical Model for Tropical Relative Humidity. *Journal of Climate, 27*(19), 7432-7449. <https://doi.org/10.1175/jcli-d-14-00255.1>
- Romps, D. M. (2016). Clausius–Clapeyron scaling of CAPE from analytical solutions to RCE. *Journal of the Atmospheric Sciences, 73*(9), 3719-3737.
- Rouhiainen, L. (2018). *Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro*. Alienta Editorial.
- Segond, M.-L., Onof, C., & Wheeler, H. S. (2006). Spatial–temporal disaggregation of daily rainfall from a generalized linear model. *Journal of Hydrology, 331*(3-4), 674-689. <https://doi.org/10.1016/j.jhydrol.2006.06.019>
- Seidel, T. M., Grant, A. N., Pszenny, A. A. P., & Allman, D. J. (2007). Dewpoint and Humidity Measurements and Trends at the Summit of Mount Washington,

- New Hampshire, 1935–2004. *Journal of Climate*, 20(22), 5629-5641. <https://doi.org/10.1175/2007jcli1604.1>
- Seidel, T. M., Grant, A. N., Pszenny, A. A., & Allman, D. J. (2007). Dewpoint and humidity measurements and trends at the summit of Mount Washington, New Hampshire, 1935–2004. *Journal of Climate*, 20(22), 5629-5641.
- Serinaldi, F., & Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, 77, 17-36.
- Shaw, S. B., Royem, A. A., & Riha, S. J. (2011a). The Relationship between Extreme Hourly Precipitation and Surface Temperature in Different Hydroclimatic Regions of the United States. *Journal of Hydrometeorology*, 12(2), 319-325. <https://doi.org/10.1175/2011jhm1364.1>
- Shaw, S. B., Royem, A. A., & Riha, S. J. (2011b). The relationship between extreme hourly precipitation and surface temperature in different hydroclimatic regions of the United States. *Journal of Hydrometeorology*, 12(2), 319-325.
- Sim, I., Lee, O., & Kim, S. (2019a). Sensitivity Analysis of Extreme Daily Rainfall Depth in Summer Season on Surface Air Temperature and Dew-Point Temperature. *Water*, 11(4), 771. <https://doi.org/10.3390/w11040771>
- Sim, I., Lee, O., & Kim, S. (2019b). Sensitivity analysis of extreme daily rainfall depth in summer season on surface air temperature and dew-point temperature. *Water*, 11(4), 771.
- Suparta, W., Alhasa, K., & Singh, M. (2017). Estimation water vapor content using the mixing ratio method and validated with the ANFIS PWV model. *J. Phys. Conf. Ser*, 852, 012041.
- Urdiales, D., & Célleri, R. (2018). Pronóstico de la ocurrencia de precipitación diaria en un gradiente altitudinal en el Sur del Ecuador mediante un generador climático. *Enfoque UTE*, 9(3), 29-41.
- van Gerven, M., & Bohte, S. (2017). Editorial: Artificial Neural Networks as Models of Neural Information Processing. *Frontiers in Computational Neuroscience*, 11. <https://doi.org/10.3389/fncom.2017.00114>

- Velasco, S., & Fernández-Pineda, C. (2008). Sobre la obtención de la ecuación de Clapeyron-Clausius. *Revista Española de Física*, 22(4), 7-14.
- Vidrio-Sahagún, C. T., He, J., Kasiviswanathan, K. S., & Sen, S. (2021). Stationary hydrological frequency analysis coupled with uncertainty assessment under nonstationary scenarios. *Journal of Hydrology*, 598, 125725.
- Villada, F., Muñoz, N., & García, E. (2012). Aplicación de las Redes Neuronales al Pronóstico de Precios en el Mercado de Valores. *Información tecnológica*, 23(4), 11-20. <https://doi.org/10.4067/S0718-07642012000400003>
- Wang, J. X., & Gaffen, D. J. (2001). Late-twentieth-century climatology and trends of surface humidity and temperature in China. *Journal of Climate*, 14(13), 2833-2845.
- Wang, Y., Tang, L., Zhang, J., Gao, T., Wang, Q., Song, Y., & Hua, D. (2018). Investigation of precipitable water vapor obtained by Raman lidar and comprehensive analyses with meteorological parameters in Xi'an. *Remote Sensing*, 10(6), 967.
- Zou, A., Xiao, T., Jia, R., Kwon, J., Mazeika, M., Li, R., & Hendrycks, D. (2022). Forecasting future world events with neural networks. *Advances in Neural Information Processing Systems*, 35, 27293-27305.