



Universidad Autónoma de Querétaro
Facultad de Informática
Maestría en Ciencias de la Computación

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

TESIS

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

Presenta:

I.S. Martín Muñoz Mandujano

Dirigido por:

Dr. Martín Alfonso Gutiérrez López

SINODALES

Dr. Martín Alfonso Gutiérrez López
Presidente

Dra. Ma. Teresa García Ramírez.
Secretario

Dra. Sandra Luz Canchola Magdaleno.
Vocal

Dr. Fausto Abraham Jacques García.
Suplente

M.S.I. José Alejandro Vargas Díaz.
Suplente

Centro Universitario, Querétaro, Qro.
Noviembre 2020
México.

Dirección General de Bibliotecas UAQ

Esta tesis esta dedicada a mi papá y mamá Martín y Vero, Gaby mi hermana y Anita mi novia.

Resumen

El objetivo de la presente investigación fue el desarrollar un modelo inteligente predictivo basado en redes neuronales artificiales, capaz de pronosticar el momento e intensidad de una precipitación, mediante el uso de variables climatológicas estocásticas, partiendo de la pregunta ¿Es posible el realizar un pronóstico preciso sobre el momento e intensidad de una precipitación? A lo largo de la investigación se emplearon datos pertenecientes a la red de monitoreo CIAQ y de la red de monitoreo de EMAS propiedad de CONAGUA. La investigación fue realizada empleando la metodología basada en diseño ya que permitió tener una mayor flexibilidad en cuanto a las etapas. Se diseñó un modelo matemático para el pronóstico de precipitaciones y mediante las variables involucradas en dicho modelo se desarrolló un modelo inteligente predictivo capaz de pronosticar la intensidad de la precipitación a lo largo del tiempo. A partir de los resultados obtenidos se concluyó que es posible obtener un pronóstico de hasta 8 horas empleando el modelo matemático y que mediante el modelo inteligente predictivo se obtiene una precisión del 99.9903%. Además, se concluyó que el modelo inteligente predictivo podría ser mejorado en gran medida pronosticando la intensidad de precipitación mediante una escala de diferentes intensidades (es decir una clasificación) y no de forma numérica.

Abstract

The objective of the present research was to develop an intelligent predictive model based on artificial neuronal networks, capable of predicting the moment and intensity of a precipitation, by means of the use of stochastic climatological variables, starting from the question: Is it possible to make an accurate forecast on the moment and intensity of a precipitation? Throughout the investigation, data belonging to the CIAQ monitoring network and the EMAS monitoring network owned by CONAGUA were used. The research was carried out using the methodology based on design since it allowed having a greater flexibility as far as the stages are concerned. A mathematical model was designed for the forecast of precipitation and through the variables involved in this model, an intelligent predictive model was developed, capable of predicting the intensity of precipitation over time. From the obtained results it was concluded that it is possible to obtain a forecast of up to 8 hours using the mathematical model and that by means of the predictive intelligent model a precision of 99.9903 % is obtained. In addition, it was concluded that the predictive intelligent model could be improved in great measure predicting the intensity of precipitation by means of a scale of different intensities (that is to say a classification) and not in a numerical way.

Agradecimientos

Primeramente agradezco 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 y apoyo.

Agradezco también el apoyo financiero para la realización de este trabajo al Consejo Nacional de Ciencia y Tecnología (CONACYT) mediante la beca número 736471 y a la Universidad Autónoma de Querétaro (UAQ) a través del proyecto 11029.

Agradezco enormemente a mi director de tesis el Dr. Alfonso Gutiérrez López quien creyó en mi para poder llevar a cabo esta investigación y por todo el apoyo, conocimiento y confianza que me brindó espero poder seguir colaborando con usted en la siguiente etapa que es el doctorado y conseguir grandes resultados.

De forma especial me gustaría agradecer a mi familia, empezando por mi papá Martín Muñoz Paredes y mi mamá Verónica Mandujano Rodríguez, quienes me han apoyado enormemente a lo largo de esta travesía llamada posgrado gracias por creer en mi y apoyarme en todo lo que necesitara, incluso si solo era contarles sobre como iba la investigación o cuando me emocionaba por que mi modelo funcionaba, yo se que siempre estarán ahí para mi cuando lo necesite, ustedes son los principales autores de que yo haya podido llegar hasta este punto. A mi hermana Gaby, muchas gracias por estar conmigo y aguantarme y estar para mi cuando me

frustraba o tenía que salir hasta muy tarde y tenías que esperarme, la verdad es que fue un gran apoyo el tenerte como *roomie* a lo largo de este tiempo te quiero muchísimo. Asimismo le agradezco enormemente a toda mi familia, a todos mis tíos y tías, primos y primas que espero que sepan que si no salgo tanto es debido al tiempo que le dedique a este trabajo y de forma muy particular a mis abuelitos quienes siempre han creído en mi y se que estarán muy orgullosos de mi por este logro.

Agradezco enormemente a todos los integrantes de centro de desarrollo quienes han sido parte fundamental en mi formación como profesional y persona, muchas gracias a cada uno de ustedes, tienen un lugar especial en mi corazón como mi segunda familia. Por supuesto que de manera particular me gustaría agradecer a 6 integrantes fundamentales, le agradezco enormemente a alex, diego y lalo quienes me han apoyado, preguntado por como me iba e incluso interesarse enormemente por mi trabajo, en especial alex que incluso me llevo a llamar la llamar la atención si es que no cumplía con los estándares o dedicación que esperaba de mi. A Liz y Mau por apoyarme con dudas de tuviera o incluso con solo escuchar mis desahogos, realmente son personas muy importantes para mi. por ultimo me gustaría agradecerle por todo este trabajo a JP, quien me ayudo a tomar la decisión de escoger este posgrado y aventurarme en el mundo de la investigación, estoy seguro que me esta viendo desde el cielo y esta orgulloso de mi trabajo.

Finalmente le agradezco a mi amor Ana Laura por haberme apoyado a lo largo de mi posgrado y por comprenderme en los días en que no podía ir a verla debido a todos mis pendientes generados por el posgrado, siempre estuvo presente apoyándome y escuchando mis alegrías y frustraciones, eres una persona muy especial para mi.

Índice general

Resumen	I
Abstract	II
Agradecimientos	III
Índice General	V
Índice de Figuras	VIII
Índice de Tablas	IX
1. Introducción	1
1.1. Descripción del Problema	1
1.2. Hipótesis	2
1.3. Justificación	3
1.4. Objetivos	4
1.4.1. Objetivo General	4
1.4.2. Objetivos Particulares	4
1.5. Alcances y Limitaciones	4
1.6. Estructura de la Tesis	5
2. Estado del Arte	7

2.1. El pronóstico climatológico.	7
2.2. Panorama sobre pronósticos y prevención en el estado de Querétaro.	9
2.3. Técnicas de pronostico	10
2.3.1. Método de Pronóstico Ingenuo.	11
2.3.2. Redes Neuronales Artificiales.	11
2.3.3. Sistema Neuro Difuso.	14
3. Metodología	15
3.1. Definición y análisis de problema	17
3.1.1. Adquisición de datos climatológicos	17
3.1.2. Análisis y preprocesamiento de datos	17
3.1.3. Identificación de variables	22
3.1.4. Modelo CRHUDA	25
3.2. Diseño y desarrollo	27
3.2.1. Análisis para pronóstico de series meteorológicas	27
3.2.2. Análisis Comparativo de redes neuronales artificiales	33
3.2.3. Modelo inteligente predictivo	34
4. Evaluación y Resultados	37
4.1. Calibración, evaluación y ajuste del modelo inteligente	37
4.2. Validación de resultados	38
5. Discusión	39
5.1. Discusión	39
5.2. Publicaciones	40
5.3. Trabajo Futuro	41
6. Conclusiones	42
Referencias	44
Referencias	44

A. Registro de Derecho de Autor	51
B. Constancia Congreso Internacional CIVITEC	53
C. Constancia Congreso Nacional CONACYTED	55
D. Articulo en revista CLIMATE	57
E. Articulo en revista NTHE	77
F. Articulo en revista NTHE	82

Dirección General de Bibliotecas UAG

Índice de figuras

3.1. <i>Principal labor de los científicos de datos. Creación propia basado en (Press, 2016)</i>	18
3.2. <i>Variedad de encabezados dentro de la información proporcionada por CONAGUA de las EMAS. Elaboración propia.</i>	20
3.3. <i>Esquema conceptual del modelo CRHUDA (CRossingHUmidity, Dew-point, and Atmospheric pressure). Elaboración propia.</i>	26
3.4. <i>Correlogramas de Temperatura. Elaboración propia.</i>	28
3.5. <i>Correlogramas de la Humedad. Elaboración propia.</i>	29
3.6. <i>Correlogramas de la Presión Atmosférica. Elaboración propia.</i>	30
3.7. <i>Correlogramas del Punto de Roció. Elaboración propia.</i>	31
3.8. <i>Correlogramas de CHUDA. Elaboración propia.</i>	32
3.9. <i>Diagrama del modelo inteligente predictivo. Elaboración propia.</i>	35
4.1. <i>Gráfica de validación cruzada de datos pronosticados y datos esperados. Elaboración propia.</i>	38

Índice de tablas

2.1. Rangos de pronósticos meteorológicos.	8
2.2. Características fundamentales de una red neuronal artificial.	12
3.1. Resultados del análisis comparativo de las librerías de aprendizaje automático	33

Dirección General de Bibliotecas UAQ

Introducción

1.1. Descripción del Problema

Uno de los factores que influye de manera importante en la vida diaria de las personas es el clima, puesto que es capaz de ocasionar cambios en la vida de las personas, en la economía o incluso en el medio ambiente (Jain y Jain, 2017). El clima está compuesto por una gran y diversa cantidad de variables como lo son: temperatura, humedad, presión atmosférica, punto de rocío, entre otras; todas las variables se ven afectadas por el comportamiento e interacción de cada una de las demás variables y a su vez, por una enorme gama de factores tales como: contaminación, zona geográfica, estación o los eventos que ocurran dentro de la zona.

Las precipitaciones son uno de los eventos más importantes de la investigación hidrológica, debido a los daños que pueden ocasionar (Lin y Jhong, 2015). La precipitación se clasifica, según sea líquida o sólida, en: pedrisco, aguanieve, granizo, ventisca o llovizna (para la Ciencia y la Tecnología, 2004). México es un país con una gran biodiversidad y con grandes centros urbanos. Año con año esas áreas naturales y esas zonas urbanas se encuentran expuestas a un gran número de fenómenos meteorológicos extremos. La noción de estudiar el régimen de eventos extremos no es nueva y ha sido objeto de estudios especializados a lo largo del

tiempo (Huff, 1967). En especial el monitoreo y el pronóstico de precipitaciones extremas en grandes zonas urbanas y en reservas ambientales resulta prioritario. El pronóstico es importante porque permite evaluar el riesgo hidrológico asociado a la severidad de las tormentas y de las pérdidas económicas que ocasionan graves daños a las economías locales (Serinaldi y Kilsby, 2015).

Según lo que explica Nikam y Meshram (2013), existen tres problemáticas principales del por qué no se realizan pronósticos precisos sobre la precipitación, las cuales son:

1. La lluvia es un evento demasiado aleatorio y las causas por las que la lluvia ocurre son sumamente complejas.
2. El número de variables de entrada podría ser muy grande como para realizar un pronóstico preciso de 24 horas.
3. Para poder predecir una precipitación, los expertos necesitan pronosticar el clima del área, aumentando las variables involucradas y la dificultad del proceso.

Por consiguiente, la presente investigación se enfocará en realizar un modelo inteligente predictivo empleando el pronóstico de series de tiempo mediante redes neuronales artificiales, con el objetivo de pronosticar el momento e intensidad de una lluvia, tomando como principal área de pruebas la zona metropolitana de Querétaro.

1.2. Hipótesis

Esta investigación parte de la hipótesis de que el momento e intensidad de la precipitación están relacionadas con las demás variables climatológicas y que, al ser analizadas, pueden servir para generar pronósticos.

Las principales preguntas que pretende resolver esta investigación son:

1. ¿Es posible el realizar un pronóstico preciso sobre el momento e intensidad de una precipitación?
2. ¿Qué variables afectan directamente en el momento de presentarse una precipitación?
3. ¿Qué variables afectan directamente en la intensidad de una precipitación?
4. ¿Qué técnica de inteligencia artificial es la más adecuada para realizar un pronóstico preciso en cuanto al momento e intensidad de una precipitación?

1.3. Justificación

Actualmente las inundaciones ocasionadas por las lluvias son un tema de interés en el estado de Querétaro. Según datos de la Oficina para la Reducción de Riesgos de las Naciones Unidas, desde 1970 se han perdido más de 64 mil millones de pesos en daños ocasionados por inundaciones en el estado de Querétaro (Cano, 2017).

El 27 de septiembre de 2017 ocurrió una de las lluvias más destructivas en la zona de Jurica, pues dejó como saldo 2 socavones, una persona fallecida y al menos 20 colonias afectadas (Gómez Alma, 2017). Esto se ocasionó debido al desbordamiento de 3 drenes y a la enorme cantidad de agua que cayó durante el evento (Excelsior, 2017). Es importante señalar que, según los colonos de la zona, no hubo ningún aviso ni pronóstico oportuno que la lluvia tendría tal dimensión, ellos se dieron cuenta una vez que el agua ya se encontraba en el interior de sus domicilios.

Al ser la lluvia un factor tan importante en la vida cotidiana de las personas, y a su vez es capaz de generar una enorme cantidad de daños, es sumamente necesario el mantener informada a la comunidad sobre sus pronósticos y estado actual.

Debido a lo anterior es necesario y de suma importancia el crear una manera de pronosticar las precipitaciones de manera oportuna, así como la dimensión de las mismas para poder aminorar los daños ocasionados por estas.

1.4. Objetivos

1.4.1. Objetivo General

El objetivo general de este proyecto es el desarrollar un modelo inteligente predictivo basado en redes neuronales artificiales, capaz de pronosticar el momento e intensidad de una precipitación, mediante el uso de variables climatológicas estocásticas, para la prevención de precipitaciones intensas en el estado de Querétaro.

1.4.2. Objetivos Particulares

Los objetivos particulares de este proyecto son los siguientes:

- Analizar las variables climatológicas que intervienen en la ocurrencia de la precipitación.
- Experimentar con el modelo inteligente predictivo, la ocurrencia o no de la precipitación en el corto plazo.
- Evaluar el índice de éxito del modelo inteligente predictivo, así como identificar las variables que afecten dicho índice.

1.5. Alcances y Limitaciones

El proyecto de investigación descrito en la tesis esta sujeto a los siguientes alcances y limitaciones:

- El principal objetivo de esta investigación es el comprobar si es posible pronosticar el momento e intensidad de una precipitación. Sin embargo no contempla su puesta en marcha en ningún sistema de alerta temprana.
- Debido a la magnitud del proyecto únicamente serán realizadas pruebas de laboratorio controladas con las variables climatológicas que se registren y se empleen como pronostico.

1.6. Estructura de la Tesis

La tesis se encuentra organizada de la siguiente manera:

- El capítulo 2 habla sobre el estado del arte en cuanto al pronostico climatológico y mas en especifico al pronóstico dentro del estado de Querétaro.
- El capítulo 3 habla sobre la metodología empleada durante la presente investigación así como también se subdivide en 2 secciones:
 - La sección 3.1 muestra la definición y análisis del problema donde se muestra la adquisición de los datos climatológicos, su análisis y preprocesamiento así como también describe las variables que se identificaron como clave para el pronóstico de la precipitación y finalmente explica el modelo de pronóstico CRHUDA.
 - La sección 3.2 muestra el análisis para el pronóstico de las series meteorológicas y el análisis comparativo entre los diferentes frameworks de redes neuronales que fueron considerados para desarrollar el modelo, además se explica el modelo inteligente predictivo final que se genero mediante esta investigación.
- El capítulo 4 habla sobre la evaluación y resultados, en este se muestran las observaciones que se hicieron durante la calibración, evaluación y ajuste del

modelo inteligente predictivo para llevarlo a su versión final además que se muestra la validación de los resultados obtenidos durante la investigación.

- El capítulo 5 habla sobre la discusión de la investigación, se muestran las publicaciones y demás productos generados a partir de la presente investigación así como también se habla sobre el trabajo futuro que se planea generar gracias a los resultados obtenidos por la presente investigación.
- El capítulo 6 cierra la tesis con las conclusiones que se pueden definir gracias a los resultados de la investigación.

Dirección General de Bibliotecas UAO

Estado del Arte

2.1. El pronóstico climatológico.

El pronóstico del clima es una labor que ha sido realizada desde la antigüedad, puesto que el clima ha sido históricamente un factor muy relevante para el actuar diario de los seres humanos. En el año 650 A.C. los babilonios predecían el tiempo de acuerdo con patrones de las nubes y en el año 340 A.C. Aristóteles describía los patrones del tiempo (Berovides Hernández, 2013). Alrededor del año 1915 la firma londinense “Negretti & Zambra” patentó un pronosticador climático de bolsillo, el cual, encontrándose en el hemisferio norte, posee una precisión de cerca del 94 % para las próximas horas.

Según lo define la Organización Meteorológica Mundial (2018), el pronóstico meteorológico se divide en diversos rangos de tiempo, tal como se puede apreciar en la tabla 2.1.

A partir de lo antes mencionado, podemos definir que el pronosticador de Negretti y Zambra es de rango muy corto (very short-range).

Es importante mencionar que el pronóstico de precipitaciones a corto tiempo es actualmente muy investigado y constituye toda una línea de investigación en el

Tabla 2.1: Rangos de pronósticos meteorológicos.

Rango	Descripción Temporal
<i>Nowcasting</i>	Descripción actual de los parámetros climatológicos hasta no más de 2 horas de margen.
<i>Very short-range weather forecasting</i>	Hasta 12 horas.
<i>Short-range weather forecasting</i>	Desde 12 hasta 72 horas.
<i>Medium-range weather forecasting</i>	Desde 72 hasta 240 horas.
<i>Extended-range weather forecasting</i>	Desde 10 días hasta 30 días generalmente promediando los valores por periodos.
<i>Long-range forecasting</i>	Desde 30 días hasta 2 años.

Fuente: elaboración propia, con base en Organización Meteorológica Mundial (2018).

campo de la Hidro-informática (Moon, Kim, Lee, y Moon, 2019). Numerosas técnicas se han empleado a lo largo del tiempo para explicar el comportamiento de las precipitaciones extremas, desde técnicas estadísticas básicas, hasta análisis multivariados y recientemente se ha empleado inteligencia artificial (Meredith, Rust, y Ulbrich, 2018).

La importancia del pronóstico a corto plazo es evidente si se piensa que eventos catastróficos ocurren en cuestión de minutos; por ejemplo, un tsunami, una inundación súbita, el rompimiento de una presa y el rompimiento de drenes urbanos que ocasionan desbordamientos y graves daños a la infraestructura de zonas urbanas (Thiboult, Anctil, y Ramos, 2017). Un ejemplo muy importante en el pronóstico de la precipitación a corto plazo son los estudios de erosión del suelo. Este tipo de estudios son escasos, ya que se requieren mediciones de intensidad de lluvia en intervalos de 30 minutos. Es decir, con intensidades de lluvia en 30 minutos se puede conocer la energía que proviene de la precipitación y que actúa como un factor erosivo del suelo (Bagarello, Ferro, y Giordano, 2010; Kinnell, 2014; Lin, Chen, y Kao, 2005).

2.2. Panorama sobre pronósticos y prevención en el estado de Querétaro.

Actualmente en el estado de Querétaro son dos las dependencias que poseen alguna herramienta de pronóstico de precipitaciones, siendo estas la Comisión Estatal de Aguas de Querétaro (CEA) y la Universidad Autónoma de Querétaro (UAQ). La CEA muestra en su página Web información climatológica actual de ciertos municipios, así como un pronóstico a 14 días del estado climatológico. En lo referente exclusivamente a lluvias solo muestra la probabilidad de lluvia (Comisión Estatal de Aguas, 2016).

En la ciudad de Querétaro, existe una red de estaciones meteorológicas distribuidas por todo el territorio del Estado y aproximadamente 34 estaciones se encuentran concentradas en la capital y forman parte de la Red de Monitoreo de Precipitaciones Extremas de Querétaro (RedCIAQ). La recolección de datos se realiza minuto a minuto y en tiempo real, de manera que se dispone de una base de datos de más de 20 millones de conjuntos de datos (Aguilar, Pastor, Vázquez, y Ibarra, 2018; Suparta, Alhasa, y Singh, 2017). Esta red de vigilancia del clima es uno de los sistemas más avanzados de América Latina y el Caribe, y también se han desarrollado más de 18 herramientas hidro-informáticas para permitir la realización de varios análisis en tiempo real. Esta red de monitoreo depende del Centro de Investigaciones del Agua de la Facultad de Ingeniería y del Centro de Desarrollo de la Facultad de Informática, de la UAQ.

La plataforma de monitoreo de red CIAQ, posee diversas herramientas para el monitoreo climatológico en la región metropolitana de la ciudad de Querétaro y sus alrededores. A diferencia de la plataforma de la CEA, muestra el estado climatológico actual de varios puntos en la ciudad, haciendo más preciso el monitoreo del estado climatológico actual en la ciudad. En lo referente a pronósticos de lluvias,

muestra el pronóstico que registran las estaciones (Pérez Valencia, 2015). Asimismo, se tiene la aplicación de “Hidrojurica”, que mantiene al tanto a los residentes de la zona de Jurica en lo referente a las precipitaciones, además de mostrar un semáforo de alerta de lluvias en la zona metropolitana del Municipio de Querétaro (Ibarra Corona, Cano Iturbe, Lozano Domínguez, Vargas Díaz, y Alejandro, 2018).

Las variables monitoreadas por los sensores instalados en las estaciones meteorológicas automáticas (EMA) son la temperatura de la lluvia, la velocidad y dirección del viento, la radiación solar, el punto de rocío, la humedad y la presión atmosférica, todas ellas transmitidas en tiempo real en el portal web (redciaq.uaq.mx) y son ampliamente consultadas por los ciudadanos y las sociedades académicas y científicas del Estado, además de las autoridades municipales y estatales para la implementación de alertas y programas de apoyo a los sectores vulnerables. La figura 2 muestra la pantalla de monitoreo de una de las estaciones de RedCIAQ. Cuando se selecciona una estación, sus variables meteorológicas se muestran en tiempo real, con un gráfico de las últimas 24 horas.

2.3. Técnicas de pronóstico

Existen varias técnicas para pronosticar el clima, pero tal como lo explican Salman, Kanigoro y Heryadi (2015), últimamente, gracias al avance tecnológico, ha sido posible combinar técnicas de modelado y predicción estadística con técnicas de Machine Learning, consiguiendo así el optimizar esta labor y permitiendo que se desarrolle de manera más precisa y veloz (Moon y cols., 2019) Algunas de estas técnicas son:

- Método de Pronóstico Ingenuo.
- Redes Neuronales Artificiales.
- Sistema Neuro Difuso.

2.3.1. Método de Pronóstico Ingenuo.

Según Makridakis, Wheelwright, y Cortés Fregoso (2004), en estadística, es conocido también como modelo de caminata aleatoria. Este modelo consiste en tomar como pronóstico la información más reciente con relación al último valor real. Esto puede ser expresado con la fórmula 3.1:

$$P_{t+1} = Y_t \quad (2.1)$$

Donde:

- P = el pronóstico del período t+i.
- t = el período actual.
- i = es el número de período adelantados pronosticados.
- Y = el ultimo valor real.

2.3.2. Redes Neuronales Artificiales.

Las redes neuronales artificiales (RNA) son un modelo computacional de procesamiento paralelo, el cual está compuesto por un conjunto de elementos neuronales simples. Este tipo de redes se relacionan con el cerebro, por la manera en la que se adquiere conocimiento mediante un proceso de aprendizaje y la cohesión presente entre las neuronas (peso sináptico), la cual se emplea para almacenar conocimiento (van Gerven y Bohte, 2017).

Algo que caracteriza a las redes neuronales artificiales es su esquema computacional distribuido, el cual básicamente se asimila a una estructura del sistema nervioso.

La Tabla 2, muestra las tres características fundamentales de una red neuronal artificial:

Tabla 2.2: Características fundamentales de una red neuronal artificial.

Topología de Red	Permite determinar la capacidad representativa de cómo una cantidad de neuronas están distribuidos en capas y distribuidas entre sí. En donde, la topología de red se basa en el problema a resolver.
Regla de aprendizaje	Una red neuronal es una integración de diversos sistemas de aprendizaje, es por lo que tienen la capacidad de aprender a través de un entrenamiento previo.
Tipo de entrenamiento	Una RNA presenta dos tipos de entrenamiento que, durante la etapa de aprendizaje, la red se entrena para que los pesos sinápticos se adecuen a la red. La segunda es una etapa de ejecución en donde la red ya pasa a ser operativa y toma un valor de funcionamiento real.

Fuente: elaboración propia, con base en Rivas-Asanza, Mazon-Olivo, y Mejia (2018).

Debido a la constitución y fundamentos que poseen, las redes neuronales artificiales cuentan con una variedad de ventajas y desventajas, las cuales, según (Rivera, 2005), se encuentran descritas a continuación:

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 se vuelven complejas cuando se trata del aprendizaje de tareas de gran tamaño.
- El tiempo de aprendizaje dependerá del número de patrones a reconocer y de la flexibilidad para reconocer patrones similares.
- Su tiempo de convergencia puede ser elevado, es decir, existen aplicaciones en los cuales requiere de un elevado número de ejemplos para la etapa de entrenamiento.
- No posee la capacidad de interpretación de los resultados, esta tarea recae sobre el programador o la aplicación que emplee la red neuronal artificial.

Las redes neuronales han sido utilizadas para resolver numerosos problemas, entre los que destacan su aplicación en la predicción de series temporales y su capacidad para detectar y explotar la no linealidad de sus datos, aun cuando exista ruido en los datos o que posea la existencia de datos incompletos; también son destacadas por su desempeño en la solución de problemas complejos donde el reconocimiento de patrones o modelos es muy importante (Villada, Muñoz, y García, 2012).

Ian, Yoshua (2015) explican que existe una gran variedad de tipos de redes neuronales artificiales dependiendo de la arquitectura en la cual se inspiraron para su desarrollo. Algunas de estas arquitecturas son:

- Perceptrón Multi Capa (Multi Layer Perceptron, por sus siglas en inglés MLP).
- Red Neuronal Recursiva (Recursive Neuronal Network, también conocido por sus siglas en inglés RNN).
- Memoria a Corto y Largo Plazo (Long Short Term Memory, también conocido por sus siglas en inglés LSTM).

- Red Neuronal Recurrente (Recurrent Neuronal Network, también conocido por sus siglas en inglés RNN).

Así como existen diversos tipos de redes neuronales artificiales, también existe una gran variedad de herramientas y programas para el desarrollo de estas, algunos ejemplos son la librería Keras o la librería Pytorch, ambas desarrolladas para el lenguaje Python.

2.3.3. Sistema Neuro Difuso.

Como es mencionado en el libro “Ingenio Magno” (Hernández-Leal, Duque-Méndez, y Moreno-Cadavid, 2016), un sistema neuro difuso consiste en una combinación de RNA con otra técnica de inteligencia artificial ampliamente difundida: la lógica difusa (LD). Concretamente se trata de un sistema de inferencia neuro difuso adaptativo (ANFIS, por sus siglas en inglés). La primera aproximación realizada sobre la combinación de LD y RNA fue hecha por Jang, que utilizó las RNA para buscar la solución más óptima mediante el cambio de los parámetros de la FL a través de un algoritmo de aprendizaje.

Metodología

Esta investigación será llevada a cabo usando como base la metodología de investigación basada en diseño, pues partiendo de lo mencionado por Confrey (2005), el término diseño hace referencia al diseño instructivo que se elabora, implementa y se somete a escrutinio de investigación.

Los procesos de investigación realizados bajo este enfoque metodológico no están bien definidos. Sin embargo, independientemente del número de etapas en que se divida el proceso, todos ellos incluyen una serie de acciones comunes como son: definición del problema, diseño, desarrollo, implementación y evaluación (de Benito Crosetti y Salinas Ibáñez, 2016).

Esta investigación esta compuesta por 4 etapas o fases principales las cuales son descritas a continuación:

1. Definición y análisis de problema:

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

- c) Identificación de variables climáticas para describir el inicio de la precipitación, mediante un análisis multivariado.

2. Diseño y desarrollo.

- a) Análisis estocástico de series de tiempo para efectos de pronóstico meteorológico (gráfica de las series de tiempo y de los puntos de cruce a diferentes escalas).
- b) Análisis comparativo de diferentes modelos de redes neuronales artificiales.
- c) Desarrollo de modelo inteligente predictivo basado en la red neuronal artificial adecuada para la identificación de patrones y generación de pronóstico de precipitación.

3. Implementación y evaluación.

- a) Implementación y entrenamiento del modelo inteligente predictivo de tipo red neuronal artificial.
- b) Calibración, evaluación y ajuste del modelo empleando datos históricos de las variables climatológicas como datos de entrada.
- c) Validación de los resultados empleando una validación cruzada. Se utilizarán series de tiempo históricas y se reproducirá la condición de precipitación con la red neuronal propuesta y se evaluarán los errores y dispersión del pronóstico.

4. Transferencia de conocimiento y desarrollo tecnológico.

- a) Elaboración de tesis.
- b) Elaboración de artículos.
- c) Participación en congresos.

3.1. Definición y análisis de problema

3.1.1. Adquisición de datos climatológicos

En nuestro país pocas son las redes de monitoreo de variables climatológicas a tiempo real. La heterogeneidad de regiones climáticas representa un problema cuando se trata de estudiar la distribución espacio-temporal de la precipitación (Gutiérrez-López y Barragán-Regalado, 2019). En América Latina y en México las estaciones climatológicas son en general, más abundantes que las estaciones hidrométricas; principalmente porque una estación climatológica es más barata y más fácil de montar que una estación hidrométrica (Shrestha, et al., 2017). A nivel nacional, el Servicio Meteorológico Nacional (SMN) posee a la fecha, mas de 160 Estaciones Meteorológicas Automáticas (EMAS). Las EMAS, además de localizarse en un terreno favorable, se componen de diversos instrumentos y aparatos, que facilitan el registro de las precipitaciones.

se requiere utilizar información registrada, por ejemplo, cada diez minutos, o analizar los registros históricos de las bandas pluviográficas (Jehn, et al., 2019).

3.1.2. Análisis y preprocesamiento de datos

El análisis y preprocesamiento de datos es una de las tareas con mayor importancia y a la que mas tiempo se le debe dedicar dentro de las áreas de la inteligencia artificial y *Data Science*.

¿A qué dedica su tiempo un científico de datos?

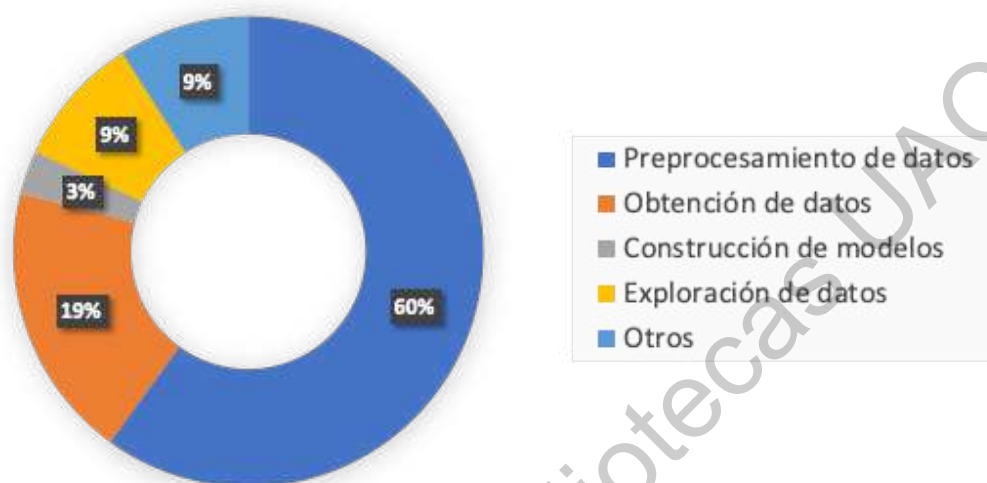


Figura 3.1: *Principal labor de los científicos de datos. Creación propia basado en (Press, 2016)*

Según una encuesta realizada por Press (2016) cuyos resultados pueden ser apreciados en la figura 3.1, la principal actividad de un científico de datos es el preprocesamiento de datos.

La intención de realizar dicho análisis y preprocesamiento es el poder evitar 4 principales problemas que se pueden presentar en cualquier fuente de datos independientemente de su origen, formato o método de adquisición.

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

La ausencia de valores es un problema común cuando se trabaja con grandes registros de información. Esto puede acontecer por infinidad de motivos por ejemplo la

ausencia de medición real, equivocación al almacenar la información, equivocación al recuperarla, entre algunos otros. La principal preocupación de estas ausencias es que impiden que el sistema de aprendizaje automático pueda entrenarse adecuadamente ya que la ausencia de datos no es numéricamente tratable.

Algunas posibles soluciones para esta problemática es el uso de técnicas como la interpolación, la cual puede ser usada cuando se trabaja con datos temporales y consiste en rellenar el dato ausente empleando la información que poseen los datos próximos a estos. Otra posible solución sería rellenar el valor faltante empleando algún valor fijo como la media, la moda o incluso 0. Finalmente otra posible solución para rellenar el valor ausente es el uso de técnicas de regresión con la intención de predecir cual era dicho valor.

La inconsistencia de datos ocurre en muchas ocasiones cuando al procesar datos se detectan errores en el formato o en el tipo de alguno de ellos. Esto puede ser debido a un error de lectura de los datos o a un mal almacenamiento de estos. Por ejemplo: fechas que siempre comienzan por el día del mes y en ciertos registros empiezan por el año, valores que deberían ser numéricos que incluyen otro tipo de caracteres, etc. Existen multitud de validaciones que es necesario comprobar y tratar de solucionar para que los datos sean coherentes.

La existencia de valores duplicados puede afectar en gran medida al emplear dicha información para el entrenamiento de redes neuronales, puesto que de haber valores duplicados, la red neuronal aprendería con ese sesgo en la información.

Los *Outliers* son valores anómalos que sobresalen de los límites máximos o mínimos, pueden deberse a errores de almacenamiento, de medición o de inserción de los datos. Estos valores pueden distorsionar mucho la distribución de los datos haciendo que todo el proceso de aprendizaje se vea afectado.

En adición a lo expuesto anteriormente, la información proporcionada por CONAGUA se encuentra almacenada en archivos en formato XLS (Formato de Microsoft

Excel 97-2004 WorkBook) separados por estación y cada archivo posee la información referente a 1 mes de datos recabados en dicha estación, cada registro se encuentra almacenado con una diferencia de 10 minutos entre cada medición.

De forma adicional, se poseen datos que van desde el año 2000 hasta el 2019, por lo que como era de esperarse, poseen diversidad de formatos y nombres de las variables, haciendo que el uso de dichos datos sea realmente complicado. De forma precisa se detectaron 8 diferentes combinaciones de formatos y nombres de encabezados, las cuales se muestran a continuación:

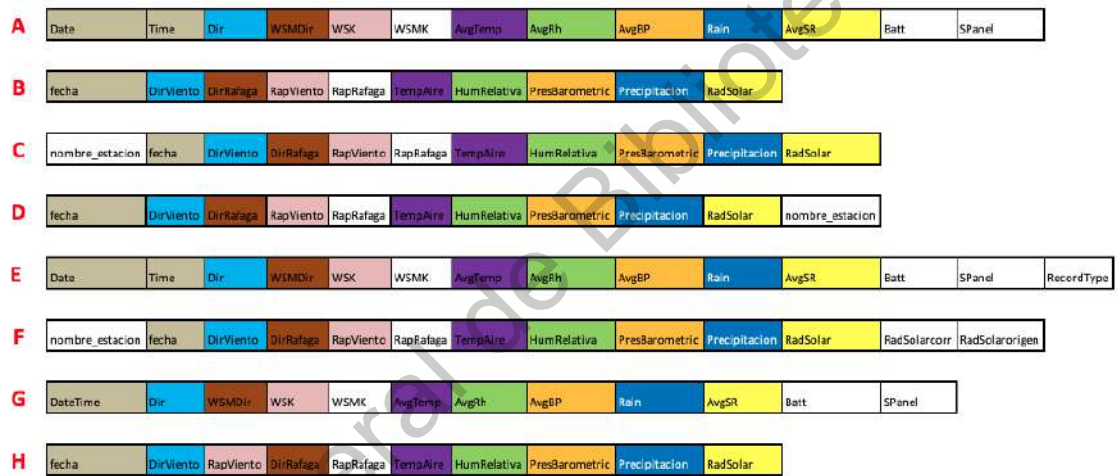


Figura 3.2: Variedad de encabezados dentro de la información proporcionada por CONAGUA de las EMAS. Elaboración propia.

Como se puede apreciar en la figura 3.2, existen varias diferencias que al verlo de forma empírica parecieran no ser la gran cosa. Por ejemplo que en algunos formatos se maneje inglés y español, o que las fechas en algunos estén en un solo campo junto a las horas o que se encuentren divididos en 2 campos diferentes, también en algunos se utiliza el sistema métrico y en otros campos se emplea el sistema imperial.

El problema principal se presenta al querer tratar toda la información como una misma, el hecho de tener tantas discrepancias entre ellos genera que no pueda

usarse directamente para el entrenamiento, razón por la cual se optó por generar una herramienta que permitiera normalizar dicha información y pudiera ser mejor manipulada no solo para entrenamiento de redes neuronales, si no también para cualquier tarea para la cual se necesite acceder a dicha información.

La herramienta desarrollada se le denominó de forma eventual como DATEX (por *Data Extractor*), y consiste en un programa de línea de comandos que se distribuye en formato de archivo ejecutable y es multiplataforma ya que puede ser empleado en sistemas operativos Windows, Linux y Mac OS. Para su utilización es necesario ejecutarlo mediante comandos dentro de CMD o *PowerShell* en entornos Windows o mediante la terminal en entornos Linux y Mac OS.

La herramienta es capaz de realizar 3 principales tareas:

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

La normalización de archivos es la principal tarea por la cual fue desarrollada esta herramienta, con la intención de normalizar los archivos provenientes de las estaciones EMAS de CONAGUA. La herramienta estandariza los valores y nombres de las columnas de un archivo dado y genera un nuevo archivo con extensión XLS con un formato estándar ya definido. Por cuestiones de usabilidad esta funcionalidad se aplica por defecto en las demás tareas, contemplando que un usuario las emplee sin haber normalizado el archivo en una primer instancia.

La concatenación de archivos se desarrolló con el hecho de que generalmente deben emplearse mas de 1 mes de información para entrenamientos, pronósticos y diversidad de cálculos mas. Es por ello que dicha función recibe una serie de archivos con extensión XLS los agrupa en uno solo ya normalizado.

Finalmente la tarea de detección de tormentas se desarrolló con la finalidad de detectar las tormentas cuya lluvia acumulada sea igual o supere un límite proporcionado por el usuario y genera un archivo con extensión XLS por cada una de las tormentas que detecte, almacenando la información desde 24 horas antes del inicio de la tormenta hasta 50 minutos después del último registro de precipitación detectado.

En lo referente a los datos originarios de la Red CIAQ, no es necesario el emplear la herramienta, pues todos se encuentran normalizados en el mismo tipo de dato y con los mismos encabezados, además esta información se encuentra con una separación de 1 minuto entre cada registro.

3.1.3. Identificación de variables

Esta sección se encuentra pendiente de añadir referencias, formato y estilo en formulas y figuras

En la naturaleza, el agua tiene tres estados, pero cada estado depende principalmente de la temperatura y la presión (Hou, Kong, Chen, y Lei, 2013). La transformación del agua de sólida a líquida se conoce como el punto de fusión, y el proceso de transición de sólido a gas se llama sublimación (J. X. Wang y Gaffen, 2001). La presión absoluta a la que el agua pasa de líquido a gas se conoce como presión de vapor, y esta transformación ocurre a una temperatura llamada punto de ebullición (Suparta y cols., 2017).

El término humedad se utiliza mucho cuando se habla de la humedad presente en la atmósfera. La presión parcial ejercida por las moléculas de agua puede aumentar hasta alcanzar un cierto límite, en este umbral, el número de moléculas de agua que se evaporan del líquido es igual al número de las que regresan de la atmósfera, estableciendo un equilibrio dinámico entre la evaporación y la condensación, llamado saturación. El umbral de saturación está determinado por la temperatura T , pero

es independiente de la presión del aire seco (Camuffo, 2014). Si en este proceso completo hay un cambio en la presión P (Pumo y cols., 2019), es porque hay un trabajo, W , que será igual al calor latente C multiplicado por el cambio en temperatura ΔT (Capparelli, 2013). Esto puede mostrarse como

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

Por otra parte, para que se produzca cualquier cambio de estado, es necesario cambiar el volumen específico, es decir, un diferencial de los volúmenes de gas (V_g) a líquido (V_l) (Van der Dussen, De Roode, y Siebesma, 2014; Romps, 2016).

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

igualando las ecuaciones anteriores:

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

Esta formulación para el calor latente de la vaporización fue deducida por Clapeyron de la teoría de Carnot, y fue probada por Clausius. Esta relación se utiliza para calcular la temperatura cuando se conocen los volúmenes específicos y la relación entre el aumento de la presión de saturación y la T , (Camuffo, 2014). Esta ecuación se conoce como la relación Clausius-Clapeyron (C-C), y caracteriza el comportamiento de un sistema cerrado durante un cambio de fase (Agard y Emanuel, 2017; Lorenz y DeWeaver, 2007), donde la temperatura y la presión son constantes por definición (Romps, 2014).

La hipótesis básica de la relación C-C es que a medida que aumenta la temperatura, la humedad relativa se mantiene constante y la humedad específica aumenta tras el incremento de la disponibilidad de humedad en la atmósfera (Park y Min, 2017). En algunas regiones tropicales, el aumento total de las precipitaciones puede ser mayor que el previsto por la relación Clausius-Clapeyron y, por lo tanto, se

requiere una respuesta compensatoria diferente (Chang, Stein, Wang, Kotamarthi, y Moyer, 2016). Ciertamente, la predicción de la precipitación debe basarse en el uso adecuado y detallado de la relación C-C. Este punto de vista es apoyado por (Lenderink, Barbero, Loriaux, y Fowler, 2017), que escribe "las relaciones actuales de escalamiento de precipitación-temperatura indican que los extremos de precipitación horaria pueden tener una respuesta al calentamiento que excede la relación Clausius-Clapeyron". Esto es simple de ver en un gráfico, ya que normalmente se presenta como temperatura ($1/T$) vs. presión ($\ln P$) (Bürger, Heistermann, y Bronstert, 2014; Peleg y cols., 2018). En (Maíllo y Fernández-Pineda, 2011), se pueden proponer 11 reglas diferentes para la deducción de la Ecuación (3). Hasta la fecha, se han desarrollado e introducido varios métodos (estudios) para encontrar la relación C-C entre las precipitaciones extremas subdiarias y la temperatura media diaria (Park y Min, 2017).

La temperatura del punto de rocío, T_d , es la temperatura a la que el aire se satura si se enfría a presión constante (Seidel, Grant, Pszenny, y Allman, 2007). T_d es la temperatura a la que la presión de vapor es igual a la presión de saturación del aire (relación de mezcla de vapor de agua). De la misma manera, el volumen de vapor presentado en la atmósfera puede ser expresado a través de la presión que este vapor genera (Millán, Ghanbarian-Alavijeh, y García-Fornaris, 2010; Chang y cols., 2016). Sin embargo, la presión total en la atmósfera es la suma de la presión causada por el aire seco más la presión producida por el vapor de agua (Harder y Pomeroy, 2013; Holley, Dorling, Steele, y Earl, 2014). Por lo tanto, la mayor presión de vapor que puede estar presente depende de la temperatura de la superficie (Dahm, Bhardwaj, Sperna Weiland, Corzo, y Bouwer, 2019). A medida que la temperatura aumenta, más presión de vapor puede estar contenida en el aire (Mohr y Kunz, 2013; Myoung y Nielsen-Gammon, 2010). Esto puede ser expresado por la relación Clausius-Clapeyron (Gao, Li, y Sorooshian, 2007; Y. Wang y cols., 2018).

Por lo tanto, cuando el aire está saturado de vapor de agua, la presión del vapor

de agua depende sólo de la temperatura (Sim, Lee, y Kim, 2019). Por otro lado, la temperatura de las corrientes ascendentes húmedas inicializadas en la superficie y la mayor profundidad de las nubes son funciones claras del punto de rocío de la superficie (Lenderink y cols., 2017). Por consiguiente, debe proponerse un modelo sencillo que lleve la relación (proporción) entre la presión y el punto de rocío ($\frac{\partial P}{\partial T_d}$). Este modelo también debe permitir confirmar el papel fundamental de la humedad de la superficie en la actividad convectiva. Si se acepta que los parámetros meteorológicos como la presión, la temperatura y la humedad relativa cambian a diferentes alturas (Liu y cols., 2015), se puede implementar un modelo sinóptico de evolución que cruce una serie de humedad, punto de rocío y presión atmosférica.

3.1.4. Modelo CRHUDA

El modelo ARMA (p,q) es un modelo autoregresivo AR de orden (p) en el cual se le agrega el componente MA (q), conocido como promedio móvil. De esta forma a estos modelos se les conoce como modelos de *Box-Jenkins*. La forma general de este modelo es del tipo ($Z_t = \sum_{j=1}^p \alpha_j Z_{t-j}$) donde p es el orden del modelo autorregresivo. Z_t es la variable estandarizada en el tiempo; ε_t es la serie de residuales; y α_j es el coeficiente autorregresivo. Para estimar los parámetros $\alpha_1, \alpha_2, \dots, \alpha_p$ se forma el sistema de p ecuaciones no lineales y se resuelve usando la función de auto-covarianza $r_k = \hat{\alpha}_j r_{k-1} + \hat{\alpha}_j r_{k-2} + \dots + \hat{\alpha}_j r_{k-p}$, donde $k > 0$. El parámetro α_j estimado se denota como $\hat{\alpha}_j$.

Para construir el modelo propuesto llamado CRHUDA, se tienen dos series independientes $S1 \subseteq S2$ de variables climáticas S1 y S2 definidas por un modelo auto-regresivo de primer orden AR(1):

$$\begin{aligned} S1 &= H_t = \phi_1 H_{t-1} + \varepsilon_t \\ S2 &= C_t = \varphi_1 C_{t-1} + \varepsilon_t \end{aligned} \tag{3.4}$$

donde $\phi_1 = r_1$ y $\varphi_1 = r_1$ son los coeficientes de correlación serial $k = 1$ para cada una de las series. Esto significa que hay un coeficiente de proporcionalidad en ambas series que permite escalar la serie para cruzarla en el tiempo t_1 . Sin embargo, hay que recordar que el coeficiente de auto-correlación de retraso k es igual a 1.

La C en la ecuación 3.4 se sustituye por el valor de la presión atmosférica (P), y en el denominador, (T) se sustituye por la temperatura del punto de rocío (T_d), similar a como lo describe Shaw, Royem, y Riha (2011). De esta manera, se grafican dos series temporales diferentes: la primera es el dato de humedad (S_1) y la segunda es la serie (S_2), definida por $(\frac{\partial P}{\partial T_d})$, similar a la relación Clausius-Clapeyron. El cruce de estas dos series mostrará al principio de la alerta ($t_1 : S_1 \subseteq S_2$), y algunas horas después, la serie T_t se cruza una vez mas ($t_2 : S_1 \cap S_2$) y en ese momento la precipitación se inicia (ocurre) $T_t + \Delta t$. La figura 3.3 muestra el esquema conceptual del modelo CRHUDA. Si el modelo predice el comienzo exacto de la precipitación, entonces $\Delta t \rightarrow 0$.

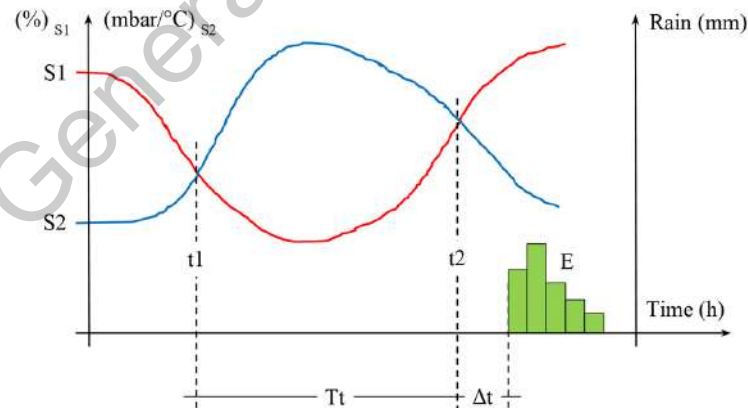


Figura 3.3: Esquema conceptual del modelo CRHUDA (CRossingHUmidity, Dew-point, and Atmospheric pressure). Elaboración propia.

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

Esto quiere decir que para que la precipitación empiece es necesario que ocurra en $t1$ $H_{t1} = C_{t1}$; $(S1 \cap S2)$, y considerando $T_t + \Delta t, H_{t2} = C_{t2}$ en $t2$. Si $\Delta t = 0$, entonces el pronóstico del principio de la precipitación E es preciso. Si $\Delta t \neq 0$, existe un tiempo de retraso en el inicio de la precipitación del evento E.

Para establecer un modelo de predicción preciso de la precipitación, utilizamos el modelo CRHUDA y aplicamos tres variables meteorológicas: 1) humedad, 2) punto de rocío, 3) presión atmosférica (Vincent, van Wijngaarden, y Hopkinson, 2007; Lepore, Allen, y Tippet, 2016). Se ha demostrado que estas tres variables principales están vinculadas entre sí; cuando se utilizan en un modelo cruzado de variables climáticas, se puede predecir el inicio de una tormenta (Rogers, Wang, y Coleman, 2007; Maíllo y Fernández-Pineda, 2011). Por consiguiente, los límites de cada variable deben establecerse primero individualmente y analizarse después para determinar cómo pueden relacionarse entre sí para obtener una respuesta en un escenario adecuado (Egerer, Lin, y Kendal, 2019; Emmanuel, Houngué, Biaou, y Badou, 2019).

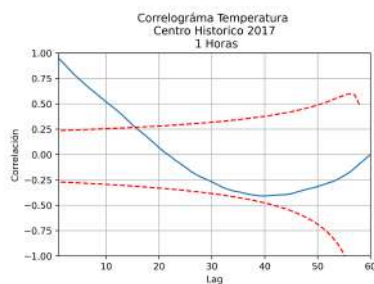
3.2. Diseño y desarrollo

3.2.1. Análisis para pronóstico de series meteorológicas

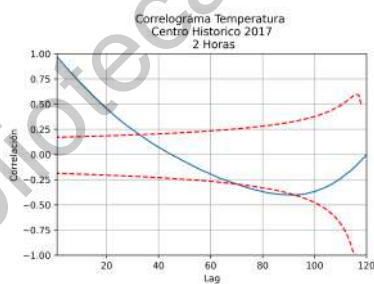
Una serie de tiempo se le denomina a cualquier variable que conste de datos reunidos, registrados u observados sobre incrementos sucesivos de tiempo. Por lo tanto, se concluye que es una secuencia ordenada de observaciones sobre una variable en particular.

Una serie estacionaria es aquella cuyos momentos al origen y a la media no varían a través del tiempo. Estas situaciones se presentan cuando los patrones de demanda que influyen sobre la serie son relativamente estables.

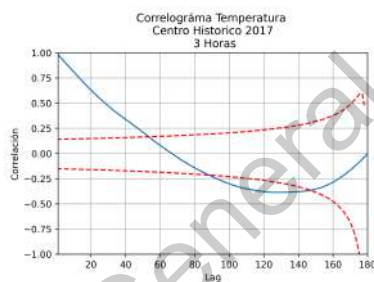
A continuación se muestran los correlogramas de temperatura, humedad, presión atmosférica, punto de rocío y CHUDA, estos correlogramas fueron generados empleando datos de un muestreo de la estación Centro Histórico de Querétaro, perteneciente a la Red CIAQ los datos pertenecen al año 2017.



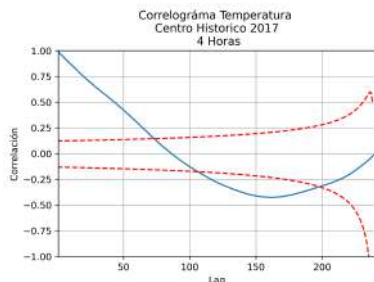
(a) Temperatura 1 Hora



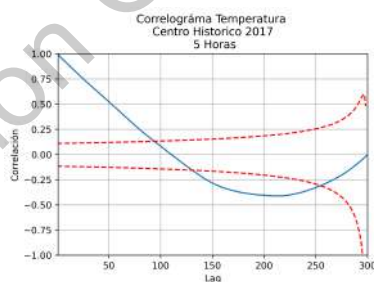
(b) Temperatura 2 Horas



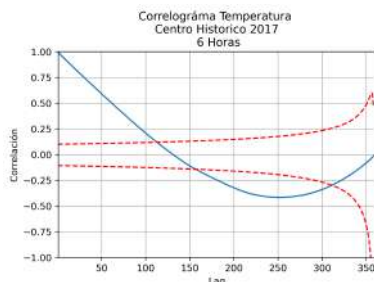
(c) Temperatura 3 Horas



(d) Temperatura 4 Horas

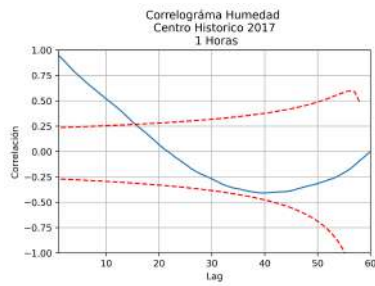


(e) Temperatura 5 Horas

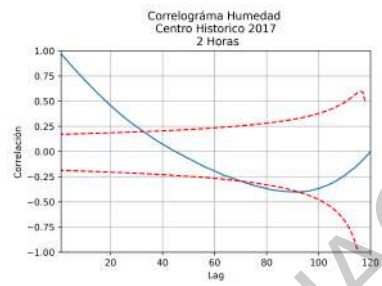


(f) Temperatura 6 Horas

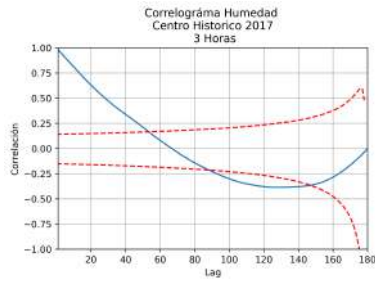
Figura 3.4: Correlogramas de Temperatura. Elaboración propia.



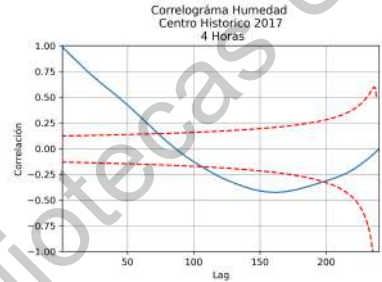
(a) Humedad 1 Hora



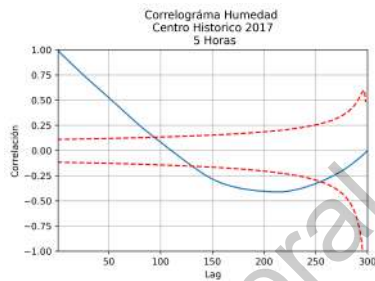
(b) Humedad 2 Horas



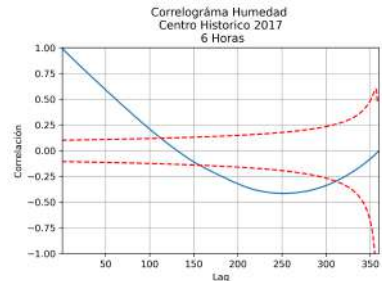
(c) Humedad 3 Horas



(d) Humedad 4 Horas

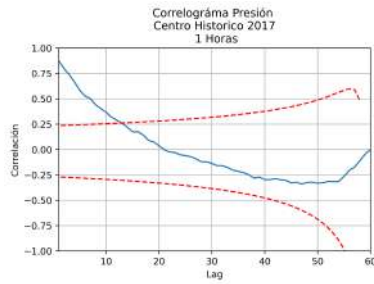


(e) Humedad 5 Horas

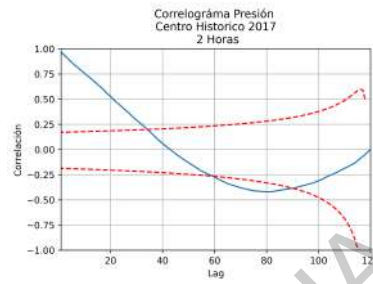


(f) Humedad 6 Horas

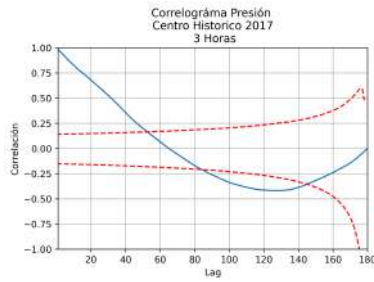
Figura 3.5: Correlogramas de la Humedad. Elaboración propia.



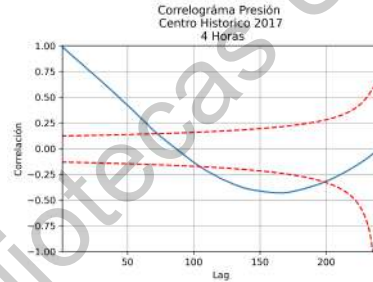
(a) Presión Atmosférica 1 Hora



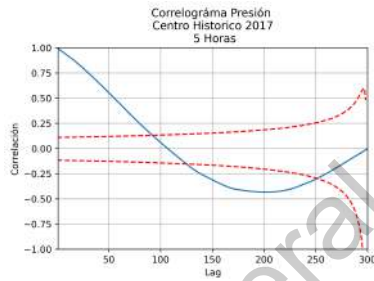
(b) Presión Atmosférica 2 Horas



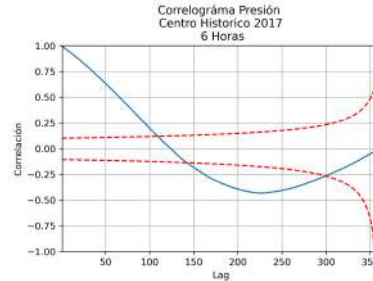
(c) Presión Atmosférica 3 Horas



(d) Presión Atmosférica 4 Horas



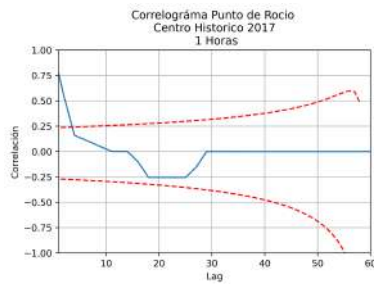
(e) Presión Atmosférica 5 Horas



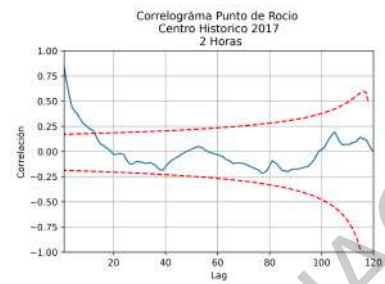
(f) Presión Atmosférica 6 Horas

Figura 3.6: Correlogramas de la Presión Atmosférica. Elaboración propia.

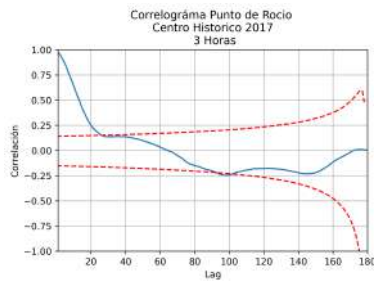
Como se puede apreciar en las gráficas de las figuras 3.4, 3.5, 3.6, la temperatura, la humedad y la presión atmosférica poseen una correlación dentro del límite entre 2 horas y 3 horas, pasando este límite los datos perderán fiabilidad al ser pronosticados mediante modelos estadísticos como lo son los modelos regresivos.



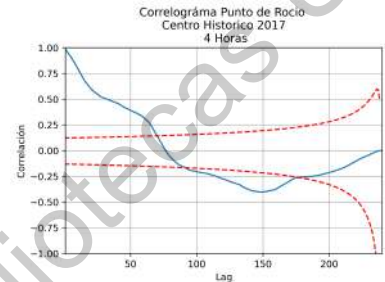
(a) Punto de Roció 1 Hora



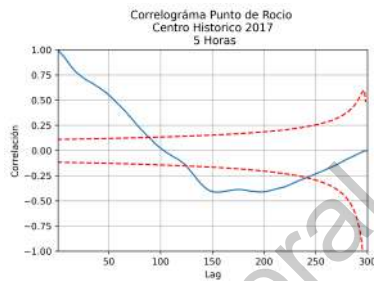
(b) Punto de Roció 2 Horas



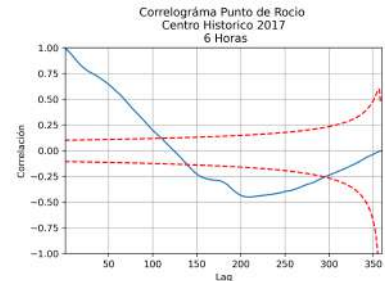
(c) Punto de Roció 3 Horas



(d) Punto de Roció 4 Horas



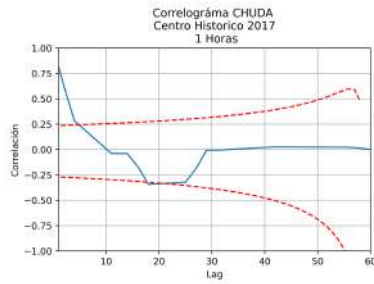
(e) Punto de Roció 5 Horas



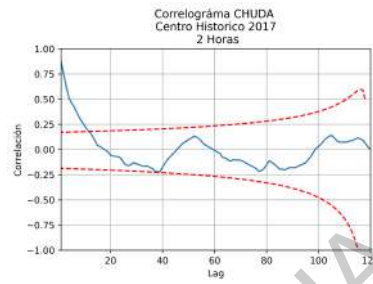
(f) Punto de Roció 6 Horas

Figura 3.7: Correlogramas del Punto de Roció. Elaboración propia.

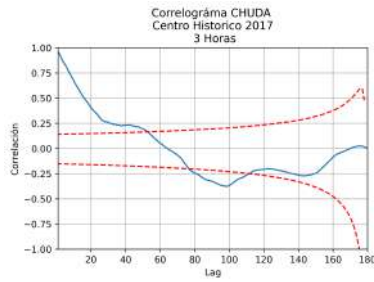
Como se puede apreciar en las gráficas de la figura 3.7, el punto de roció posee una correlación dentro del limite entre 3 horas y 4 horas, pasando este limite los datos perderán fiabilidad al ser pronosticados mediante modelos estadísticos como lo son los modelos regresivos.



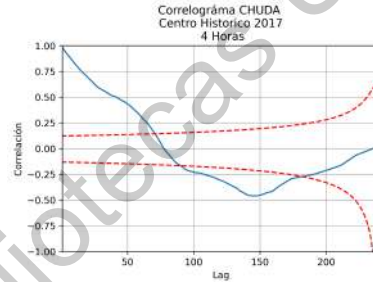
(a) CHUDA 1 Hora



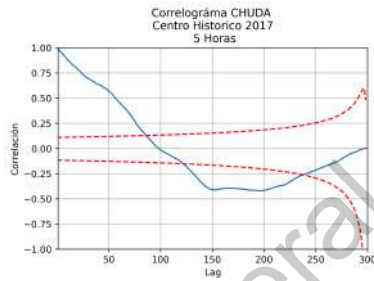
(b) CHUDA 2 Horas



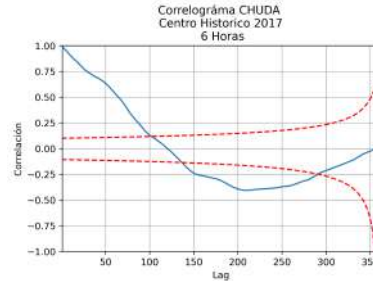
(c) CHUDA 3 Horas



(d) CHUDA 4 Horas



(e) CHUDA 5 Horas



(f) CHUDA 6 Horas

Figura 3.8: Correlogramas de CHUDA. Elaboración propia.

Como se puede apreciar en las gráficas de la figura 3.8, la serie CHUDA posee una correlación dentro del límite entre 2 horas y 3 horas, pasando este límite los datos perderán fiabilidad al ser pronosticados mediante modelos estadísticos como lo son los modelos regresivos.

En cuanto a la lluvia, la lluvia es una serie de tiempo que no posee una correlación pues es completamente aleatoria, depende mas de factores externos que en sus propios valores a lo largo del tiempo.

3.2.2. Análisis Comparativo de redes neuronales artificiales

Existe una gran cantidad de librerías para el desarrollo de redes neuronales, así como también una gran variedad de redes neuronales a partir de las cuales se puede crear un modelo.

Algunas de las librerías mas populares son las siguientes:

- *TensorFlow*
- *PyTorch*
- *Theano*
- *Keras*

Según la comparativa realizada por Gevorkyan, Demidova, Demidova, y Sobolev (2019), el desempeño de las diversas librerías es muy similar, con excepción de la librería PyTorch que es un tanto mas lenta, los resultados de su comparativa pueden apreciarse a continuación en la tabla 3.1.

Tabla 3.1: Resultados del análisis comparativo de las bibliotecas de aprendizaje automático

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

Fuente: elaboración propia, con base Gevorkyan y cols. (2019).

Además del comparativo, se realizo una búsqueda en su documentación, se determino que la opción mas viable es Keras, debido a que puede emplear como base modelos que empleen TensorFlow o Theano.

No se eligió PyTorch debido a que no posee tanta documentación en cuanto a su uso con datos numéricos, esta enfocada principalmente a su uso con el procesamiento de imágenes, ya sea para identificación de objetos o clasificación, en cambio existe una gran variedad de documentación, manuales y libros sobre el uso de Keras empleando datos numéricos.

3.2.3. Modelo inteligente predictivo

El modelo fue construido empleando el lenguaje de python en su versión 3.7 y se empleo como librería Keras con base de TensorFlow.

El modelo planteado usa la red neuronal "Secuencial", el cual permite construir un modelo de red neuronal incorporando diversas capas con diferentes arquitecturas.

Se decidió emplear como tipo de capa oculta al modelo de red neuronal tipo *Long Short Term Memory* (LSTM). Las redes neuronales de tipo LSTM son un tipo especial de redes recurrentes. La característica principal de las redes recurrentes es que la información puede persistir introduciendo bucles en el diagrama de la red, por lo que, básicamente, pueden «recordar» estados previos y utilizar esta información para decidir cuál será el siguiente. Mientras las redes recurrentes estándar pueden modelar dependencias a corto plazo (es decir, relaciones cercanas en la serie cronológica), las redes neuronales tipo LSTM pueden aprender dependencias largas, por lo que se podría decir que tienen una «memoria» a más largo plazo.

El modelo final esta compuesto por 5 capas en total, siendo 1 capa para la entrada de datos, 3 capas ocultas y 1 capa de salida para la generación de los datos pronosticados, en la figura 3.9, se puede apreciar una representación gráfica del modelo empleando el tipo de diagrama Fully Connected Cubic Network (FCCN).

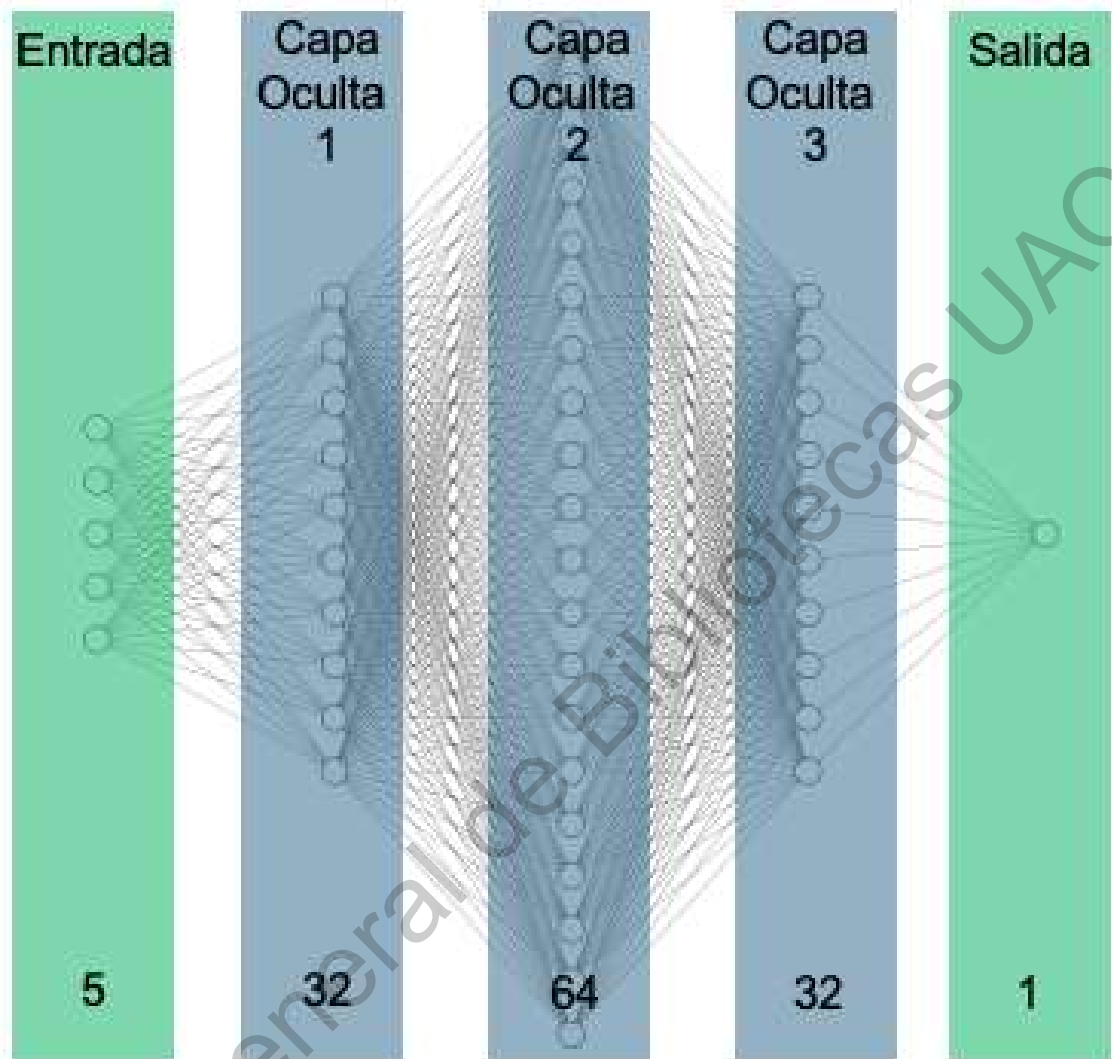


Figura 3.9: Diagrama del modelo inteligente predictivo. Elaboración propia.

En la figura anterior se muestra la cantidad de nodos que posee cada capa, la capa de entrada posee 5 nodos, uno para cada valor de entrada, en cuanto a las capas ocultas estas poseen 32, 64 y 32 respectivamente. La capa de salida esta compuesta por un solo nodo.

Las variables de entrada que fueron empleadas dentro del modelo son la temperatura, la precipitación, la humedad, la presión atmosférica y el punto de rocío, dichas variables de entrada fueron seleccionadas a partir del modelo matemático

CRHUDA

Cada una de las capas ocultas emplea como función de activación la función ELU (*Exponential Linear Unit*), la cual posee una mayor exactitud al trabajar con valores positivos.

Dirección General de Bibliotecas UAQ

Evaluación y Resultados

4.1. Calibración, evaluación y ajuste del modelo inteligente

El modelo comenzó originalmente teniendo aproximadamente de 100 a 150 generaciones para un resultado aceptable en cuanto a la pérdida y cálculo de error y para evitar el sobre entrenamiento se empleaba una capa de *Dropout*. Durante las pruebas de entrenamiento se empezó a modificar las capas y la cantidad de nodos en cada capa hasta que quedo la versión de 3 capas ocultas con 32, 64 y 32 nodos respectivamente, modelo que fue explicado en el capítulo anterior.

De forma adicional, durante las adecuaciones en el entrenamiento se encontró que era más eficiente realizar múltiples entrenamientos con pocas épocas a realizar un entrenamiento con una gran cantidad de épocas requiriendo menos épocas para alcanzar un grado aceptable de precisión. Mediante este ajuste se paso de necesitar 100 o 150 épocas para obtener resultados aceptables a necesitar únicamente de 15 a 40 épocas, dejando la versión final del modelo en 30 épocas exactamente y además ya no fue necesaria la capa de *Dropout* para evitar sobre entrenamiento.

Al obtener las métricas de los entrenamientos, el modelo posee una pérdida de

0.0192. En cuanto al cálculo de error, se empleó la fórmula de MSE, por sus siglas en inglés (*Mean Square Error*), la cual obtuvo un valor de error de 0.00709.

4.2. Validación de resultados

Para entrenamiento y prueba se emplearon los datos de la estación Chulavista con información correspondiente al año de 2018, se usaron un aproximado de 148 mil registros, los resultados de esta prueba se muestran en la figura 4.1

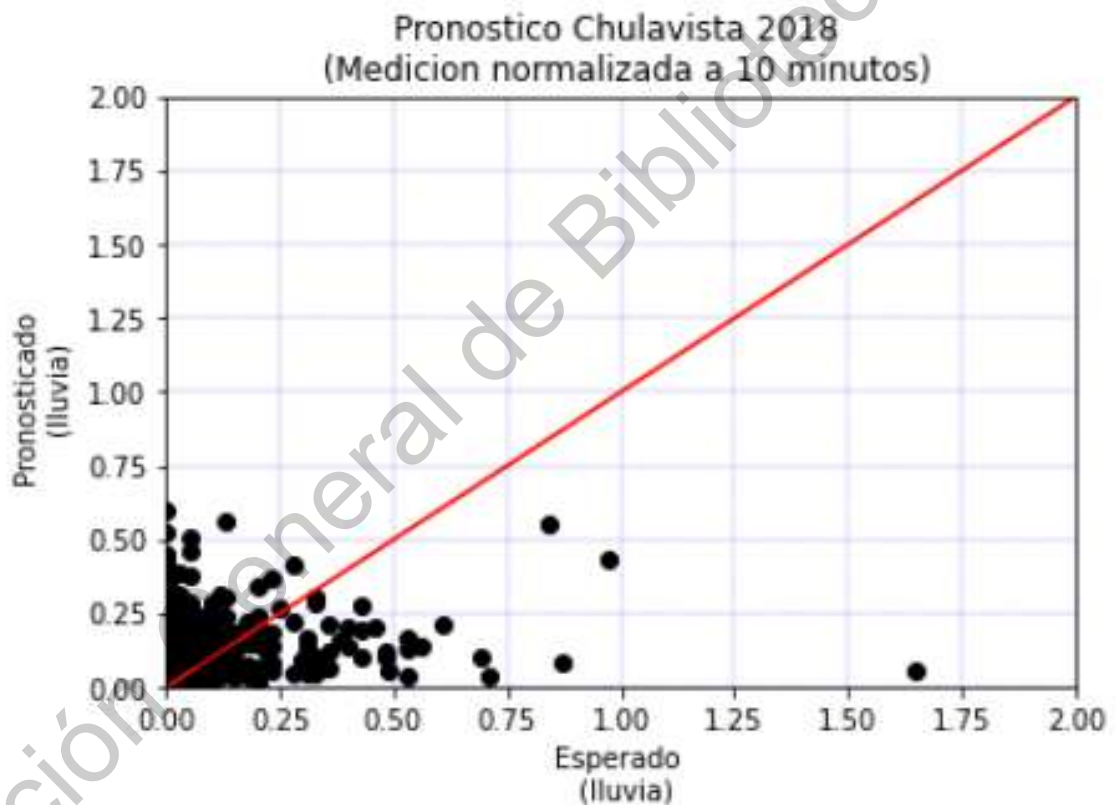


Figura 4.1: Gráfica de validación cruzada de datos pronosticados y datos esperados. Elaboración propia.

La precisión obtenida en la prueba mostrada en la figura anterior corresponde a un 99.9903 %.

Discusión

5.1. Discusión

Como se puede observar en los resultados obtenidos en el capítulo anterior, el modelo es bastante preciso, sin embargo su principal problema es que las variaciones entre lo pronosticado y lo esperado son bastante pequeñas, generando que se le dificulte al modelo optimizar en mejor medida el pronostico que genera, además algo que influye en el buen desempeño del modelo es que al también pronosticar la intensidad de lluvia en los casos en los que no llueve genera que un gran porcentaje de los resultados sean correctos pues se espera 0 de lluvia.

Otro factor a considerar es que los datos con los que se cuenta poseen varios “agujeros”, pues existe una gran cantidad de datos faltantes en la mayoría de las estaciones, ya sea por fallas en los sensores, fallas en la red, fallas de la estación en general o incluso que se deba a la intervención de las personas ya sea por ejemplo tapando los sensores o apagando la red a la que se conecta la estación.

Finalmente considero que es de suma importancia encontrar una forma óptima de realizar el entrenamiento para que el modelo se enfoque en aprender de los casos “positivo” en cuanto a lluvia, es decir en aquellos donde se presentan lluvias

y menos en los casos “negativos”, es decir los que no poseen registros de lluvias. Lo anterior surge ya que la cantidad de datos en los que no se presenta lluvia es considerablemente mayor a aquellos en los que se presentan mediciones de lluvias, pero al ser una serie de tiempo es complicado aumentar la cantidad de casos con medición de lluvias sin afectar en la serie de tiempo o simular los datos de entrenamiento.

5.2. Publicaciones

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

- Un registro de derecho de autor ante INDAUTOR del modelo de pronóstico CRHUDA.
- Participación en el 7 Congreso Internacional Virtual de Innovación, Tecnología y Educación (CIVITEC) como ponente.
- Participación en el 2 Congreso Nacional en Computación y Tecnología Educativa como ponente.
- Publicación en la revista CLIMATE en su volumen 8 del artículo *Algorithm to Predict the Rainfall Starting Point as a Function of Atmospheric Pressure, Humidity, and Dewpoint*, participando como coautor.
- Publicación en la revista NTHE en su volumen 24 del artículo , participando como primer autor.
- Publicación en la revista NTHE en su volumen 24 del artículo , participando como coautor

De forma adicional se tuvo la participación en otros 2 congresos nacionales como coautor y como primer autor, aunque es importante señalar que los temas expuestos

en dichos congresos no corresponden con la presente investigación, sin embargo es importante mencionarlos pues fueron realizados a la par de esta investigación.

Las constancias y artículos mencionados anteriormente se encuentran disponibles en la sección de anexos al final del trabajo.

5.3. Trabajo Futuro

Se pretende dar continuación a la investigación dentro del doctorado, con el enfoque de generar el pronóstico a tiempo real, teniendo la capacidad de implementarlo en algún sistema de alerta temprana. el haber obtenido el modelo matemático CRHUDA es un gran avance en lograr un pronosticador preciso en cuanto a la precipitación se refiere.

Una de las mejoras que se pretende implementar en un futuro es el pronosticar la lluvia empleando una clasificación de intensidad con la intención de poder brindar mayor exactitud al pronóstico y que este pueda ser utilizado en sistemas de alerta temprana y fácilmente interpretado.

Conclusiones

Analizando los resultados obtenidos durante el desarrollo del modelo de pronóstico y el modelo matemático de pronóstico CRHUDA se puede concluir que la hipótesis de esta investigación es correcta, ya que se demostró que las variables climatológicas de punto de rocío, humedad y presión atmosférica están relacionadas con el momento e intensidad de la precipitación. Las preguntas que trataban de ser contestadas mediante esta investigación pudieron ser contestadas y estas se muestran a continuación.

1. ¿Es posible el realizar un pronóstico preciso sobre el momento e intensidad de una precipitación? **Es posible pronosticar el momento exacto mediante el uso del modelo CRHUDA, empleando el modelo inteligente predictivo es capaz de pronosticar de forma precisa la intensidad de precipitación a lo largo del tiempo.**
2. ¿Qué variables afectan directamente en el momento de presentarse una precipitación? **Las variables que influyen en el momento de la precipitación son la humedad, la presión atmosférica y el punto de rocío.**
3. ¿Qué variables afectan directamente en la intensidad de una precipitación? **Las variables que según el modelo inteligente predictivo pueden ser uti-**

lizadas para pronosticar la intensidad de la precipitación son la temperatura, la precipitación, la humedad, la presión atmosférica y el punto de rocío.

4. ¿Qué técnica de inteligencia artificial es la más adecuada para realizar un pronóstico preciso en cuanto al momento e intensidad de una precipitación?
La técnica de redes neuronales fue la mas adecuada para la elaboración del modelo inteligente predictivo propuesto en esta investigación, siendo mas precisos la red neuronal de tipo LSTM debido a que con pocas épocas puede generar un resultado bastante preciso.

El modelo aun requiere de mas investigación debido a que la versión actual no puede ser implementada en un sistema de alerta temprana ya que no es capaz de realizar pronósticos a tiempo real. Como se comento en la sección de trabajo futuro, una forma por la cual podría ser mejorado el modelo es empleando una clasificación para la intensidad de la precipitación, pues es mas sencillo encontrar patrones y poder pronosticar de entre 5 o 6 resultados posibles en comparación a obtener el valor exacto.

Referencias

- Agard, V., y Emanuel, K. (2017). Clausius–clapeyron scaling of peak cape in continental convective storm environments. *Journal of the Atmospheric Sciences*, 74(9), 3043–3054.
- Aguilar, E., Pastor, D., Vázquez, A., y Ibarra, D. (2018). Recolección de datos meteorológicos en tiempo real mediante el uso de funciones asíncronas non-blocking. *Revista NTHE*, 24, 113–117.
- Bagarello, V., Ferro, V., y Giordano, G. (2010). Testing alternative erosivity indices to predict event soil loss from bare plots in Southern Italy. *Hydrological Processes*. doi: 10.1002/hyp.7538
- Berovides Hernández, M. (2013). *Sistema para la revisión y procesamiento básico de datos meteorológicos* (Tesis Doctoral no publicada). Universidad Central “Marta Abreu” de Las Villas.
- Bürger, G., Heistermann, M., y Bronstert, A. (2014). Towards subdaily rainfall disaggregation via clausius–clapeyron. *Journal of Hydrometeorology*, 15(3), 1303–1311.
- Camuffo, D. (2014). Theoretical Grounds for Humidity. En *Microclimate for cultural heritage*. doi: 10.1016/b978-0-444-63296-8.00002-0
- Cano, J. (2017). *Por inundaciones, \$64 mil millones perdidos – AMQueretaro.com*. Descargado 2018-06-26, de <http://amqueretaro.com/queretaro/2017/10/03/por-inundaciones-64-mil-millones-perdidos>
- Capparelli, A. (2013). Fisicoquímica básica. *Series: Libros de Cátedra*.
- Chang, W., Stein, M. L., Wang, J., Kotamarthi, V. R., y Moyer, E. J. (2016). Changes in spatiotemporal precipitation patterns in changing climate conditions. *Journal of Climate*, 29(23), 8355–8376.
- Comisión Estatal de Aguas. (2016). *Pronóstico de tiempo*. Descargado 2019-01-15, de <http://www.ceaqueretaro.gob.mx/clima/>
- Dahm, R., Bhardwaj, A., Sperna Weiland, F., Corzo, G., y Bouwer, L. M. (2019). A

- temperature-scaling approach for projecting changes in short duration rainfall extremes from gcm data. *Water*, 11(2), 313.
- Egerer, M. H., Lin, B. B., y Kendal, D. (2019). Temperature variability differs in urban agroecosystems across two metropolitan regions. *Climate*, 7(4), 50.
- Emmanuel, L. A., Houngouè, N. R., Biaou, C. A., y Badou, D. F. (2019). Statistical analysis of recent and future rainfall and temperature variability in the mono river watershed (benin, togo). *Climate*, 7(1), 8.
- Excelsior. (2017, sep). *Deja lluvia 2 socavones, un muerto e inundaciones, en Querétaro — Excelsior*. Descargado 2018-06-12, de <https://www.excelsior.com.mx/nacional/2017/09/28/1191051>
- Gao, X., Li, J., y 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., y 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.
- Gómez Alma. (2017). *Un deceso y 20 colonias afectadas en la capital por lluvia: MAV — Querétaro*. Descargado 2018-06-26, de <http://www.eluniversalqueretaro.mx/metropoli/27-09-2017/un-deceso-y-20-colonias-afectadas-en-la-capital-por-lluvia-mav#{#}imagen-1>
- Gutiérrez-López, A., y Barragán-Regalado, R. (2019). Ajuste de curvas idf a partir de tormentas de corta duración. *Tecnología y ciencias del agua*, 10(6), 1–24.
- Harder, P., y Pomeroy, J. (2013). Estimating precipitation phase using a psychrometric energy balance method. *Hydrological Processes*, 27(13), 1901–1914.
- Hernández-Leal, E. J., Duque-Méndez, N. D., y 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.
- Holley, D., Dorling, S., Steele, C., y 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., y Lei, H. (2013). Impact of 3dvar data assimilation on the prediction of heavy rainfall over southern china. *Advances in Meteorology, 2013*.
- Huff, F. A. (1967). Time distribution of rainfall in heavy storms. *Water Resources Research*. doi: 10.1029/WR003i004p01007
- Ian, Yoshua, A. C. (2015). Deep Learning. *Deep Learning*. doi: 10.1016/B978-0-12-391420-0.09987-X
- Ibarra Corona, A. M., Cano Iturbe, J. A., Lozano Domínguez, L. J., Vargas Díaz, J., y Alejandro. (2018). *Aplicación móvil para monitoreo de precipitaciones pluviales en Jurica-Querétaro* (n.º 24).
- Jain, H., y Jain, R. (2017, mar). Big data in weather forecasting: Applications and challenges. En *2017 international conference on big data analytics and computational intelligence (icbdac)* (pp. 138–142). IEEE. Descargado de <http://ieeexplore.ieee.org/document/8070824/> doi: 10.1109/ICBDACI.2017.8070824
- Kinnell, P. I. (2014). Applying the RUSLE and the USLE-M on hillslopes where runoff production during an erosion event is spatially variable. *Journal of Hydrology*. doi: 10.1016/j.jhydrol.2014.10.016
- Lenderink, G., Barbero, R., Loriaux, J., y 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.
- Lepore, C., Allen, J. T., y Tippet, M. K. (2016). Relationships between hourly rainfall intensity and atmospheric variables over the contiguous united states. *Journal of Climate, 29*(9), 3181–3197.
- Lin, G. F., Chen, L. H., y Kao, S. C. (2005). Development of regional design hyetographs. *Hydrological Processes*. doi: 10.1002/hyp.5550
- Lin, G.-F., y Jhong, B.-C. (2015, feb). A real-time forecasting model for the spatial distribution of typhoon rainfall. *Journal of Hydrology, 521*, 302–

313. Descargado de <https://www.sciencedirect.com/science/article/pii/S0022169414010117> doi: 10.1016/J.JHYDROL.2014.12.009
- Liu, Z., Chen, B., Chan, S. T., Cao, Y., Gao, Y., Zhang, K., y 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.
- Lorenz, D. J., y DeWeaver, E. T. (2007). The response of the extratropical hydrological cycle to global warming. *Journal of climate*, 20(14), 3470–3484.
- Maíllo, S. V., y Fernández-Pineda, C. (2011). Sobre la obtención de la ecuación de clapeyron-clausius. *Revista Española de Física*, 22(4).
- Makridakis, S. G., Wheelwright, S. C., y Cortés Fregoso, H. (2004). *Métodos de pronósticos*. Limusa. Descargado de <https://uaq-primo.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=52UAQ{ }Aleph000151489{&}context=L{&}vid=UAQR0{&}lang=es{ }CL{&}search{ }scope=uaqro{ }scope{&}adaptor=LocalSearchEngine{&}isFrbr=true{&}tab=uaqro{ }tab{&}query=any ,contains,SpyrosMakridakis{&}sortby>
- Meredith, E. P., Rust, H. W., y Ulbrich, U. (2018). A classification algorithm for selective dynamical downscaling of precipitation extremes. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-22-4183-2018
- Millán, H., Ghanbarian-Alavijeh, B., y García-Fornaris, I. (2010). Nonlinear dynamics of mean daily temperature and dewpoint time series at babolsar, iran, 1961–2005. *Atmospheric Research*, 98(1), 89–101.
- Mohr, S., y 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., y Moon, B. R. (2019). Application of machine learning to an early warning system for very short-term heavy rainfall. *Journal of Hydrology*. doi: 10.1016/j.jhydrol.2018.11.060

- Myoung, B., y Nielsen-Gammon, J. W. (2010). Sensitivity of monthly convective precipitation to environmental conditions. *Journal of Climate*, 23(1), 166–188.
- Nikam, V. B., y Meshram, B. (2013, sep). Modeling Rainfall Prediction Using Data Mining Method: A Bayesian Approach. En *2013 fifth international conference on computational intelligence, modelling and simulation* (pp. 132–136). IEEE. Descargado de <http://ieeexplore.ieee.org/document/6663175/> doi: 10.1109/CIMSim.2013.29
- para la Ciencia y la Tecnología, F. E. (2004). *Meteorología y climatología : unidad didáctica : Semana de la Ciencia y la Tecnología 2004*. Autor. Descargado de <http://www.divulgameteo.es/ampliab/4/31/Meteorologia-y-Climatologia.html>
- Park, I.-H., y Min, S.-K. (2017). Role of convective precipitation in the relationship between subdaily extreme precipitation and temperature. *Journal of Climate*, 30(23), 9527–9537.
- Peleg, N., Marra, F., Fatichi, S., Molnar, P., Morin, E., Sharma, A., y 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.
- Pérez Valencia, I. (2015). *Red CIAQ, el análisis científico de los fenómenos meteorológicos*. Descargado 2018-09-10, de <http://www.conacytprensa.mx/index.php/ciencia/ambiente/2459-red-ciaq>
- 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., y Noto, L. V. (2019). Sensitivity of extreme rainfall to temperature in semi-arid mediterranean regions. *Atmospheric Research*, 225, 30–44.
- Rivas-Asanza, W., Mazon-Olivo, B., y Mejia, F. (2018). Capítulo 1: Generalidades de las redes neuronales artificiales. En (pp. 11–35).
- Rivera, E. (2005). *Introducción a las Redes Neuronales Artificiales*. Editorial

Universidad Don Bosco. Descargado de <http://rd.udb.edu.sv:8080/jspui/handle/11715/238>

- Rogers, J. C., Wang, S.-H., y Coleman, J. S. (2007). Evaluation of a long-term (1882–2005) equivalent temperature time series. *Journal of climate*, 20(17), 4476–4485.
- Romps, D. M. (2014). An analytical model for tropical relative humidity. *Journal of Climate*, 27(19), 7432–7449.
- Romps, D. M. (2016). Clausius–clapeyron scaling of cape from analytical solutions to rce. *Journal of the Atmospheric Sciences*, 73(9), 3719–3737.
- Salman, A. G., Kanigoro, B., y Heryadi, Y. (2015, oct). Weather forecasting using deep learning techniques. En *2015 international conference on advanced computer science and information systems (icacsis)* (pp. 281–285). IEEE. Descargado de <http://ieeexplore.ieee.org/document/7415154/> doi: 10.1109/ICAC SIS.2015.7415154
- Seidel, T. M., Grant, A. N., Pszenny, A. A., y 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., y Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*. doi: 10.1016/j.advwatres.2014.12.013
- Shaw, S. B., Royem, A. A., y Riha, S. J. (2011). 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., y Kim, S. (2019). 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., y Singh, M. (2017). Estimation water vapor content using the mixing ratio method and validated with the anfis pwv model. En *J. phys. conf. ser* (Vol. 852, p. 012041).

- Thiboult, A., Anctil, F., y Ramos, M. H. (2017). How does the quantification of uncertainties affect the quality and value of flood early warning systems? *Journal of Hydrology*. doi: 10.1016/j.jhydrol.2017.05.014
- Van der Dussen, J., De Roode, S., y Siebesma, A. (2014). Factors controlling rapid stratocumulus cloud thinning. *Journal of the Atmospheric Sciences*, 71(2), 655–664.
- van Gerven, M., y Bohte, S. (2017, dec). Editorial: Artificial Neural Networks as Models of Neural Information Processing. *Frontiers in Computational Neuroscience*, 11. Descargado de <http://journal.frontiersin.org/article/10.3389/fncom.2017.00114/full> doi: 10.3389/fncom.2017.00114
- Villada, F., Muñoz, N., y García, E. (2012). Aplicación de las Redes Neuronales al Pronóstico de Precios en el Mercado de Valores. *Información tecnológica*, 23(4), 11–20. Descargado de http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-07642012000400003&lng=en&nrm=iso&tlng=en doi: 10.4067/S0718-07642012000400003
- Vincent, L. A., van Wijngaarden, W. A., y Hopkinson, R. (2007). Surface temperature and humidity trends in canada for 1953–2005. *Journal of Climate*, 20(20), 5100–5113.
- Wang, J. X., y 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., 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.
- World Meteorological Organization. (2018). *Supplement No. 5 - Appendix I-4*. Descargado 2018-11-23, de <http://www.wmo.int/pages/prog/www/DPS/GDPS-Supplement5-AppI-4.html>

Registro de Derecho de Autor

Dirección General de Bibliotecas UNQ

CERTIFICADO

Registro Público del Derecho de Autor

Para los efectos de los artículos 13, 162, 163 fracción I, 164 fracción I, 168, 169, 209 fracción III y demás relativos de la Ley Federal del Derecho de Autor, se hace constar que la **OBRA** cuyas especificaciones aparecen a continuación, ha quedado inscrita en el Registro Público del Derecho de Autor, con los siguientes datos:

AUTORES:

AGUIRRE CARACHEO EDUARDO
GUTIERREZ LOPEZ MARTIN ALFONSO
IBARRA CORONA DIEGO OCTAVIO
IBARRA CORONA MAURICIO ARTURO
MEDEL SAN ELIAS YAZMIN LISSET
MUÑOZ MANDUJANO MARTIN
VARGAS DIAZ JOSE ALEJANDRO

TITULO:

CRHUDA (CROSSING HULMIDITY, DEWPOINT AND, ATMOSPHERIC PRESSURE)

RAMA:

PROGRAMAS DE COMPUTACION

TITULARES:

AGUIRRE CARACHEO EDUARDO
GUTIERREZ LOPEZ MARTIN ALFONSO
IBARRA CORONA DIEGO OCTAVIO
IBARRA CORONA MAURICIO ARTURO
MEDEL SAN ELIAS YAZMIN LISSET
MUÑOZ MANDUJANO MARTIN
VARGAS DIAZ JOSE ALEJANDRO

Con fundamento en lo establecido por el artículo 168 de la Ley Federal del Derecho de Autor, las inscripciones en el registro establecen la presunción de ser ciertos los hechos y actos que en ellas consten, salvo prueba en contrario. Toda inscripción deja a salvo los derechos de terceros. Si surge controversia, los efectos de la inscripción quedarán suspendidos en tanto se pronuncie resolución firme por autoridad competente.

Con fundamento en los artículos 2, 208, 209 fracción III y 211 de la Ley Federal del Derecho de Autor; artículos 64, 103 fracción IV y 104 del Reglamento de la Ley Federal del Derecho de Autor; artículos 1, 3 fracción I, 4, 8 fracción I y 9 del Reglamento Interior del Instituto Nacional del Derecho de Autor, se expide el presente certificado.

Número de Registro: 03-2019-101410131100-01

Ciudad de México, a 16 de octubre de 2019

EL DIRECTOR DEL REGISTRO PÚBLICO DEL DERECHO DE AUTOR

JESUS PARETS GOMEZ



SECRETARÍA DE CULTURA
INSTITUTO NACIONAL DEL
DERECHO DE AUTOR
DIRECCIÓN DE REGISTRO PÚBLICO
DEL DERECHO DE AUTOR

**Constancia Congreso Internacional
CIVITEC**

Dirección General de Bibliotecas UNQ



Asociación Internacional de Investigación en Educación Superior A.C.

Otorga la presente

CONSTANCIA

Martin Muñoz Mandujano

A:

Por su participación como **Ponente** en el
7° Congreso Internacional Virtual de Innovación, Tecnología y Educación "CIVITEC 2019".

Con su ponencia titulada:

"MODELO INTELIGENTE PREDICTIVO BASADO EN REDES NEURONALES ARTIFICIALES PARA EL PRONÓSTICO DE MOMENTO E INTENSIDAD DE PRECIPITACIONES.".

Eje temático:

Tecnologías de la Información y Comunicación

ATENTAMENTE

Dra. María Elizabeth Ojeda Orta
Presidenta **CIVITEC 2019**

Tijuana, Baja California,
7 y 8 de Noviembre de 2019.



**Constancia Congreso Nacional
CONACYTED**

Dirección General de Bibliotecas UNQ



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE INFORMÁTICA

Otorgan el presente reconocimiento a

Martín Muñoz Mandujano

Por su exposición en la ponencia:

**Pronóstico de temperaturas empleando red neuronal
con arquitectura LTSM.**

Durante el Congreso Nacional en Computación y Tecnología Educativa

4, 5 y 6 de noviembre de 2019 Juriquilla, Querétaro

M.I.S.D. Juan Salvador Hernández Valerio
Director de la Facultad de Informática



Dra. Ana Marcela Herrera Navarro
Coordinadora del CONACYTED

Artículo en revista CLIMATE

Dirección General de Bibliotecas UNQ

Article

Algorithm to Predict the Rainfall Starting Point as a Function of Atmospheric Pressure, Humidity, and Dewpoint

Alfonso Gutierrez-Lopez ^{1,*} , Ivonne Cruz-Paz ² and Martin Muñoz Mandujano ² 

¹ Water Research Center, Centro de Investigaciones del Agua-Queretaro (CIAQ), International Flood Initiative, Latin-American and the Caribbean Region (IFI-LAC), International Hydrological Programme (IHP-UNESCO), Universidad Autonoma de Queretaro, Queretaro 76010, Mexico

² Facultad de Informatica, Universidad Autonoma de Queretaro, Juriquilla Queretaro 76230, Mexico; icruz23@alumnos.uaq.mx (I.C.-P.); mmunoz23@alumnos.uaq.mx (M.M.M.)

* Correspondence: alfonso.gutierrez@uaq.mx; Tel.: +52-442-192-1200 (ext. 6401)

Received: 24 September 2019; Accepted: 23 October 2019; Published: 12 November 2019



Abstract: Forecasting extreme precipitations is one of the main priorities of hydrology in Latin America and the Caribbean (LAC). Flood damage in urban areas increases every year, and is mainly caused by convective precipitations and hurricanes. In addition, hydrometeorological monitoring is limited in most countries in this region. Therefore, one of the primary challenges in the LAC region the development of a good rainfall forecasting model that can be used in an early warning system (EWS) or a flood early warning system (FEWS). The aim of this study was to provide an effective forecast of short-term rainfall using a set of climatic variables, based on the Clausius–Clapeyron relationship and taking into account that atmospheric water vapor is one of the variables that determine most meteorological phenomena, particularly regarding precipitation. As a consequence, a simple precipitation forecast model was proposed from data monitored at every minute, such as humidity, surface temperature, atmospheric pressure, and dewpoint. With access to a historical database of 1237 storms, the proposed model allows use of the right combination of these variables to make an accurate forecast of the time of storm onset. The results indicate that the proposed methodology was capable of predicting precipitation onset as a function of the atmospheric pressure, humidity, and dewpoint. The synoptic forecast model was implemented as a hydroinformatics tool in the Extreme Precipitation Monitoring Network of the city of Queretaro, Mexico (RedCIAQ). The improved forecasts provided by the proposed methodology are expected to be useful to support disaster warning systems all over Mexico, mainly during hurricanes and flashfloods.

Keywords: humidity; dewpoint; rainfall; mixing ratio; forecast rainfall model; Clausius–Clapeyron relation; early warning system (EWS); Mexico

1. Introduction

In Mexico, as in most Latin American and Caribbean (LAC) countries, there is a deficit of historical precipitation data measured in time intervals of less than 24 h. Real-time measured values are often required for the implementation of early warning systems. In Mexico, disasters are measured by the economic impact of damage and losses, as well as by the problems caused in the social environment, such as injured and dead people and damaged houses, schools, and hospitals, among other issues. From 2000 to 2014, 2147 million dollars' worth of losses and 186 annual deaths occurred [1]. The year 2013 was very intense in terms of rainfall, especially the month of September, and the historical precipitation depth increased by 60% with respect to the historical mean, registering a monthly mean

of 227.3 mm and exceeding the 1955 record of 212.1 mm [2]. Weather predictions consist of several variables, such as temperature, humidity, wind, dewpoint, among others, trying to provide a suitable and accurate forecast. However, forecasting difficulties arise often because of the small scales involved, the range of responsible physical mechanisms, and the challenge involved in forecasting events of such short duration. The first piece of evidence is the relationship between the natural logarithm of the rainfall with the surface temperature, indicating an exponential behavior [3]. Additionally, it is necessary to point out that precipitation is episodic and does not have continuous values like temperature and other climate variables; therefore, it is noticeable that anticipating the occurrence of precipitation is difficult. Lorenz and Saltzman [4–6] were perhaps the first to perceive that climate is a complex, nonlinear system involving both deterministic and stochastic components. However, the basic idea that the origin of convective precipitation as a function of surface temperature was demonstrated by [7], who used temperature and rain gauge station data of 5-min accumulation intervals. This strong relationship with temperature also implies the relationship with convective available potential energy (CAPE) magnitude. Therefore, it is accepted that the main variables for predicting precipitation are temperature and humidity [8]. Nevertheless, there are other four meteorological parameters—air, dewpoint temperature (or relative humidity), wind speed, and cloud cover—which are strongly correlated with rainfall [9]. Scientific studies are known to show that there is an increase in precipitation intensity due to temperature gradient. In particular, subdaily or hourly precipitation exhibits an improved correlation with thermodynamic constraints, known as the Clausius–Clapeyron (C–C) relation [10]. It has also been confirmed that this phenomenon occurs with different types of precipitation, i.e., with orographic precipitation and convective precipitation [11]. Concerning tropical regions, [12] showed that 30 years ago, the surface temperature was an important meteorological variable to the understanding of tropical weather systems. This was one of the first studies to use temperature gradients to carry out dust haze forecasting. Since 1990, the multivariate statistical studies of 28 different climatic variables in Australia have shown that the main factors for creating a climatic model are a humidity factor, a temperature factor, and a rainfall factor [13].

An acceptable forecast of precipitation intensity is an essential issue in the forecasts of extreme streamflow events [14]. The research to date has tended to focus on hydrometeorological data, such as wind, temperature, humidity, and atmospheric pressure; this information has been used in a daily streamflow forecast during lead times of 5–7 days [15]. However, the purpose of an early warning system (EWS), flood early warning system (FEWS), or an ensemble prediction system (EPS) is to allow warning signals prior to extreme events mainly for very short-term heavy rainfall. Recent developments in machine learning techniques have heightened the need of using the mentioned hydrometeorological data for an effective EWS for very short-term (heavy rain advisory within the next 3–9 h) [16]. Nevertheless, we must accept that short-term quantitative precipitation forecasting (SQPF) is critical for flashflood warning, navigation safety, and many other applications. Zahraei et al. [17] have recently developed a methodology based on the method named PERsiann-ForeCAST. PERCAST predicts the location and rate of rainfall up to 4 h using the most recent storm images to extract the storm features.

Recent evidence suggests that a forecasting model that includes a radar-based nowcasting component is capable of predicting rainfall distribution patterns of 1–2 h for rainstorms of moderate to heavy intensity (30–50 mm/h) [18]. However, difficulties arise when an attempt is made to implement the SQPF, especially when talking about models that can be used to simulate hydrological processes at a daily or hourly time step, because one major drawback of this approach is the large spatial variation in precipitation [19]. The changes experienced by hydroinformatics tools over the past decade remain unprecedented. Many analysts now argue that the strategy of introducing more hourly surface data, including on wind, specific humidity (dewpoint), air temperature, relative vorticity, and moisture divergence flux, has been successful when it is combined with a Bayesian framework [20] or a neural network used to generate rainfall forecast for the next time step [21]. Not only is it important to know the variables that generate rain but, also, it is important to find the right combination of these variables

to make an accurate forecast of the starting time of a storm. From the abovementioned, it is concluded that one of the most important tasks of modern hydrology is to work on simple and effective models in the forecasting of extreme events. The objective of this work is to present a simple forecast model for precipitation onset, called CRHUDA (CRossingHUmidity, Dewpoint, and Atmospheric pressure) based on measurements of humidity, atmospheric pressure, and dewpoint that are taken every minute.

2. Materials and Methods

2.1. The Clausius–Clapeyron Relation

In nature, water has three states, but each state depends mainly on temperature and pressure [22]. The transformation of water from solid to liquid is known as the fusion point, and the process of transition from solid to gas is called sublimation [23]. The absolute pressure at which water passes from liquid to gas is known as vapor pressure, and this transformation occurs at a temperature called the boiling point [24].

The term humidity is used a lot when talking about present moisture in the atmosphere. The partial pressure exerted by water molecules may increase until a certain limit is reached, at this threshold, the number of water molecules evaporating from the liquid equals the number of those returning from the atmosphere, establishing a dynamic balance between evaporation and condensation, called saturation. The saturation threshold is determined by the temperature T , but it is independent of the dry air pressure [25]. If, in this complete process, there is a change in the pressure ΔP [26], it is because there is some work, W , which will be equal to the latent heat C multiplied by the change in temperature ΔT [27]. This can be shown as

$$\Delta W = C \frac{\Delta T}{T} \quad (1)$$

On the other hand, in order for any change of state to take place, it is necessary to change the specific volume, i.e., a differential of the volumes from gas (V_g) to liquid (V_l) [28,29].

$$W = \Delta P (V_g - V_l) \quad (2)$$

By equating Equations (1) and (2),

$$\frac{\Delta P}{\Delta T} = \frac{\partial P}{\partial T} = \frac{C}{T \Delta V} \quad (3)$$

This formulation for the latent heat of vaporization was deduced by Clapeyron from Carnot's theory, and was proved by Clausius. This relation is used to calculate C at any temperature when the specific volumes and the relationship between the increase of saturation pressure and T , are known [25]. This equation is known as the Clausius–Clapeyron relationship (C–C), and it characterizes the behavior of a closed system during a phase change [30,31], where temperature and pressure are constants by definition [32].

The basic hypothesis of the C–C relation is that as the temperature increases, relative humidity remains constant and specific humidity increases after the increase of moisture availability in the atmosphere [10]. In some tropical regions, the total precipitation increase may be greater than that predicted by the Clausius–Clapeyron relation and, thus, a different compensating response is required [33]. Certainly, the prediction of precipitation must be based on the adequate and detailed use of the relationship C–C. This view is supported by [34], who write “*present-day precipitation-temperature scaling relations indicate that hourly precipitation extremes may have a response to warming exceeding the Clausius–Clapeyron relation*”. This is simple to see on a graph, since it is usually presented as temperature ($1/T$) vs. pressure ($\ln P$) [35,36]. In [37], 11 different rules can be proposed for deduction of Equation (3). To date, various methods (studies) have been developed and introduced to find the C–C relation between subdaily extreme precipitation and daily mean temperature [10].

The dewpoint temperature, T_d , is the temperature at which the air is saturated if it is cooled at constant pressure [38]. T_d is the temperature at which the vapor pressure is equal to the saturation pressure of the air (water vapor mixing ratio). In the same way, the vapor volume presented in the atmosphere can be expressed through the pressure that this vapor generates [33,39]. However, the total pressure on the atmosphere is the sum of the pressure caused by dry air plus the pressure produced by water vapor [8,40]. Thus, the greatest vapor pressure that may be present depends of the surface temperature [41]. As the temperature increases, more vapor pressure can be contained in the air [42,43]. This can be expressed by the Clausius–Clapeyron relationship [44,45].

Therefore, when the air is saturated with water vapor, the pressure of the water vapor depends only on the temperature [46]. On the other hand, the temperature of moist updrafts initialized at the surface and the greatest cloud depth are clear functions of surface dewpoint [34]. Consequently, a simple model should be proposed that carries the relationship (proportion) between pressure and dewpoint $\left(\frac{\partial P}{\partial T_d}\right)$. This model must also allow for confirmation of the key role of the surface humidity on convective activity. If it is accepted that meteorological parameters such as pressure, temperature, and relative humidity change at different altitudes [47], a synoptic evolution model crossing a series of humidity, dewpoint, and atmospheric pressure can be implemented.

2.2. The Proposed Model

The ARMA models (p,q) are the autoregressive AR models (p) to which an MA component (q) , called moving averages, has been added, and Box–Jenkins models are formed. The general type of the model is $Z_t = \sum_{j=1}^p \alpha_j Z_{t-j} + \varepsilon_t$ where p is the order of the autoregressive model; Z_t is the standardized variable in time t ; ε_t is the residual series; and α_j is the autoregressive coefficient. To estimate the parameters $\alpha_1, \alpha_2, \dots, \alpha_p$ the system of p nonlinear equations is resolved using the autocovariance function $r_k = \hat{\alpha}_j r_{k-1} + \hat{\alpha}_j r_{k-2} + \dots + \hat{\alpha}_p r_{k-p}$, $k > 0$. The parameters α_j are obtained by $\hat{\alpha}_j$.

To construct the proposed model called CRHUDA, there are two independent time series ($S1 \subseteq S2$) of climatic variables $S1$ and $S2$ defined by an autoregressive model of first-order AR(1):

$$S1 = H_t = \phi_1 H_{t-1} + \varepsilon_t \text{ and } S2 = C_t = \varphi_1 C_{t-1} + \varepsilon_t \quad (4)$$

where $\phi_1 = r_1$ and $\varphi_1 = r_1$ are the serial autocorrelation lag coefficients in time $k = 1$ for each of the series. This means that there is a proportionality coefficient in both series that allows the series to be scaled to cross in time t_1 . However, it should be remembered that the lag autocorrelation coefficient in time $k = 0$, is equal to 1.

The C in Equation (3) is replaced with the value of atmospheric pressure (P), and in the denominator, (T) is replaced with the dewpoint temperature (T_d), similar to [48]. In this way, two different time series are plotted: the first one is the humidity ($S1$) data and the second one is series ($S2$), defined by $\left(\frac{\partial P}{\partial T_d}\right)$, similar to the Clausius–Clapeyron relation. The crossing of these two series will show at the beginning of the alert ($t_1 : S1 \subseteq S2$), and some hours later, the series T_t will again cross ($t_2 : S1 \cap S2$) and, at that moment, precipitation will start $T_t + \Delta t$. Figure 1 shows the conceptual scheme of the model CRHUDA model. If the model predicts the exact start of precipitation, then $\Delta t \rightarrow 0$.

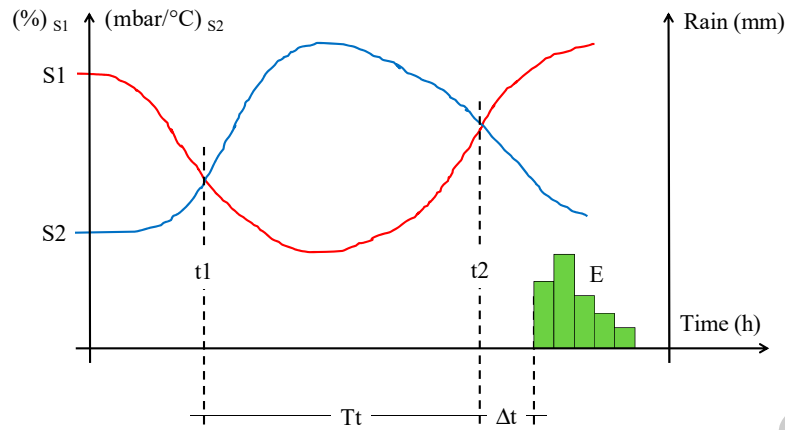


Figure 1. Conceptual scheme of the model CRHUDA (CRossingHUmidity, Dewpoint, and Atmospheric pressure).

$$CRHUDA (S1 \subseteq S2) \rightarrow (S1 \cap S2) \left\{ \begin{array}{l} S1 = \text{humidity} \\ S2 = \frac{\Delta P}{\Delta T} = \frac{\partial P}{\partial Td} = \frac{\text{atmospheric pressure}}{\text{dewpoint}} \end{array} \right.$$

This means that for the precipitation to begin, it must happen on $t1$ $H_{t1} = C_{t1}; (S1 \cap S2)$, and that considering $T_t + \Delta t$, $H_{t2} = C_{t2}$ at $t2$. If $\Delta t = 0$, the forecast of the start of the precipitation event E is precise. If $\Delta t \neq 0$, there is a time delay in the start of the precipitation event E .

2.3. The Precipitation Network RedCIAQ

In the city of Queretaro, in the middle of the Mexican republic, there is a network of 34 weather stations distributed all over the territory of the State and almost 34 stations concentrated in the capital city of Queretaro’s Extreme Precipitation Monitoring Network (RedCIAQ). The collection of data is done minute by minute and in real time such that there is a database of more than 20 million datasets available [24,49]. This climate monitoring network is one of the most advanced systems in Latin America and the Caribbean, and more than 18 hydroinformatics tools have also been developed to allow several analyses to be carried out in real time [50].

The variables monitored by the sensors installed in the automatic meteorological stations (EMA) are rainfall temperature, wind speed and direction, solar radiation, dewpoint, humidity, and atmospheric pressure all transmitted in real time in the portal web (redciaq.uaq.mx) and are extensively consulted by citizens and academic and scientific societies of the state in addition to the municipal and state authorities for the implementation of alerts and support programs for vulnerable sectors. Figure 2 shows the monitoring screen of one of the RedCIAQ stations. When a station is selected, its weather variables are displayed in real time, with a graph of the last 24 h. Figure 3 shows the location of the EMA within the Mexican territory. The real-time details of these stations can be consulted at <https://smn.conagua.gob.mx/es/pronosticos/8-smn-general/38-estaciones-meteorologicas-automaticas-emas>. Appendix A shows the details of the location of these stations. With the data of 523 storms registered from 2012 to 2018, we obtained the time series for the climatic variables of precipitation, humidity, atmospheric pressure, and dewpoint.

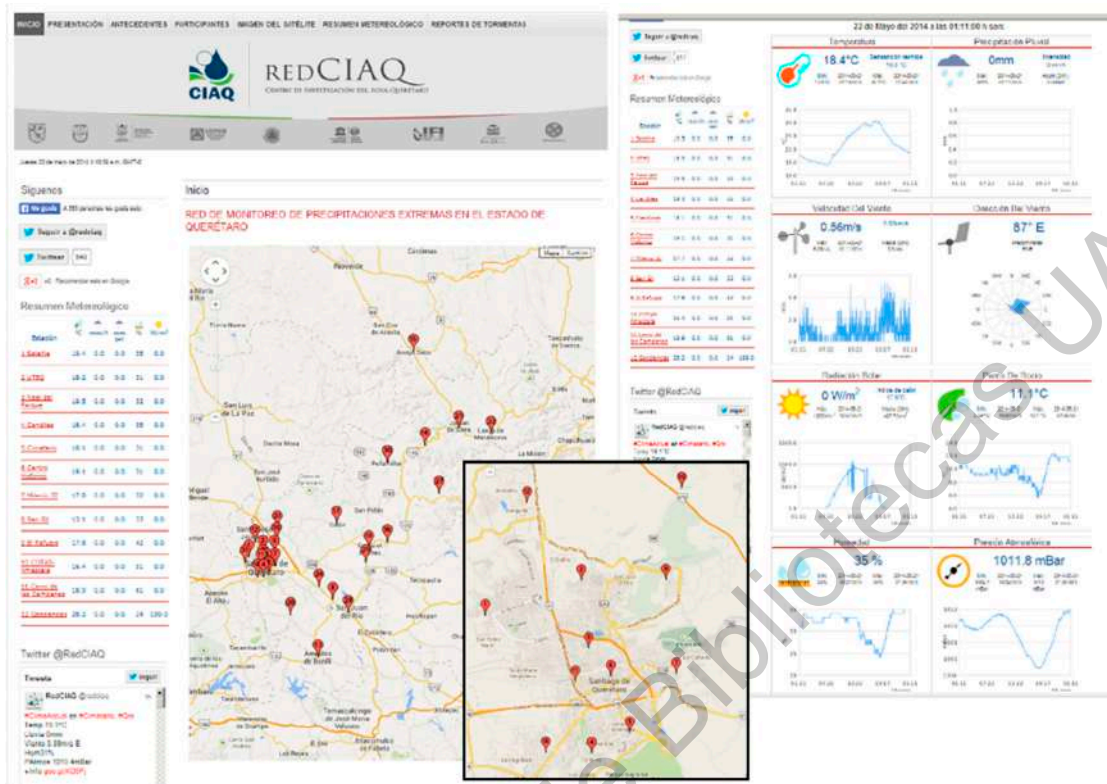


Figure 2. Queretaro’s Extreme Precipitation Monitoring Network real-time transmission website.

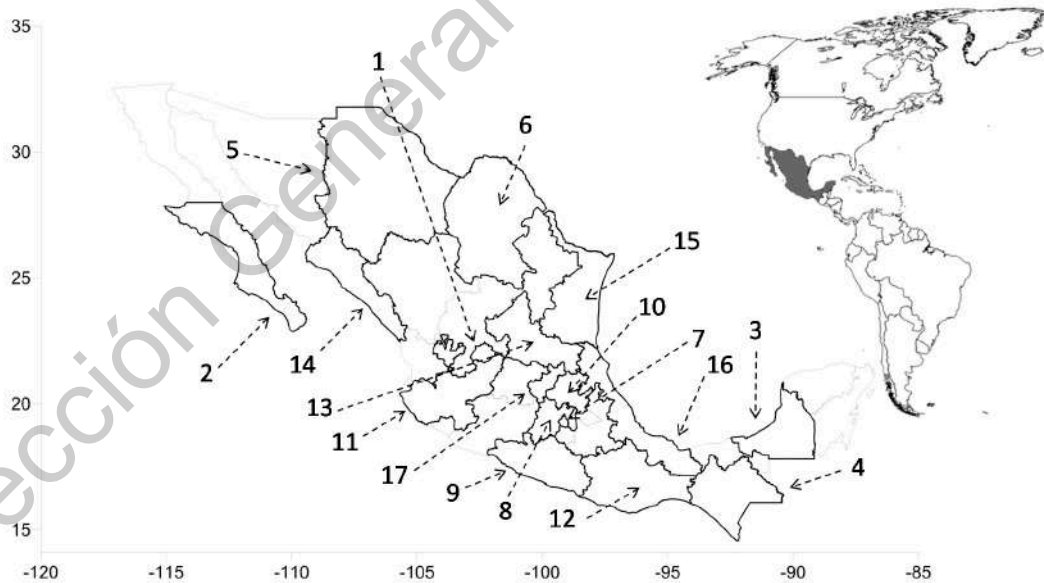


Figure 3. Location of each automatic meteorological station (EMA) and the number of storms by state. The numbers of states are shown in Table 1.

Figure 4 shows an example of time series obtained from pressure and temperature data. In Figure 4, it can be seen that there is a cross relation between these two time series. This same phenomenon is documented in some works when saturation pressure and temperature time series are used. Although there is precipitation, this was not detected in the series of atmospheric pressure and humidity, so it is

necessary to add other climate variables. Figure 5 shows the time series obtained from the data of pressure and humidity at the same time as the occurrence of the precipitation. In this time series, it can be observed that both series are almost parallel over time. Although there are two different phenomena at different scales, the behavior is very important to maintain continuity of the phenomenon. Once this same behavior has been verified in all datasets for all storms at all EMA stations, the CRHUDA ratio is obtained and the results are plotted.

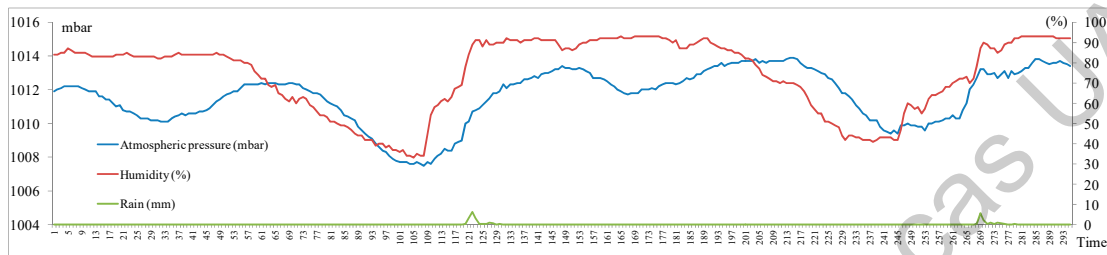


Figure 4. Example of the time series obtained from the data of pressure and humidity in the same time as the occurrence of the precipitation.

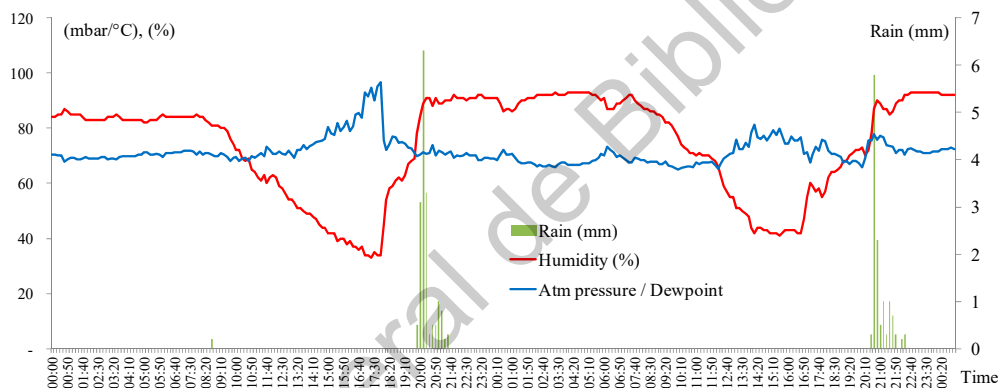


Figure 5. Results of applying the CRHUDA model at the automatic meteorological station Candiles on June 24–25, 2013.

3. Results

To set up a precise prediction model for precipitation, we used the CRHUDA model (Equation (4)) and applied three meteorological variables: 1) humidity, 2) dewpoint, 3) atmospheric pressure [3,51]. These three main variables have been shown to be linked with each other; when used in a crossover model of climate variables, the onset of a storm can be predicted [37,52]. As a consequence, boundaries for each variable should be individually established first, and analyzed afterwards to determine how they are able to relate with each other to obtain a response in a proper scenario [53,54].

The procedure consists of the time series being available at each minute for the three climatic variables referred above. Minute by minute, both series are distributed parallel on the time axis. In a synoptic sense, it is possible to see, surprisingly, that the crossing of these two time series generates a point in time that allows generating an alert between 9 and 10 h in advance of the start of the storm, at about 10:20 in Figure 5. After the first crossing, when the humidity drops quickly, these series cross one more time but in the opposite direction to the start of the precipitation. This procedure was carried out with data from the historical records of the 34 stations of the Queretaro State Extreme Precipitation Monitoring Network (RedCIAQ). During the seven years of minute-by-minute records, the variables used by CRHUDA were measured, then 523 convective events were identified. The calibration of the model was carried out with these 523 storms. To confirm the model, it was necessary to apply it to the data that were monitored every minute, and then apply it to the EMA stations of the

Mexican territory, which take data every 20 min. The calibration results showed a mean time of 10 h ($T_1 + \Delta t = 619.58$ min) between the t_1 alert and the start of precipitation t_2 , with a median of 8.9 h (535 min). The calibration results also showed that the scale factor for the humidity series varied between 0.4 and 2.6 with a mean of 1.784 and a median of 1.84 and, as discussed, the mode was equal to 1. This means that there is, in fact, a stochastic behavior in the time series.

Due to large amount of data that was analyzed, a hydroinformatics tool was created that allows the systematic analysis of all EMA records in the Mexican territory. The tool is called CRHUDA, and is copyrighted. It was developed in C+ language and current work on a second version includes forecasting with ARMA(p,q) models. Some of the Caribbean countries have begun to provide information to verify the usefulness of the proposed model in other regions. Figure 5 shows the results of applying the CRHUDA model at the automatic meteorological station Candiles on June 24–25, 2013. Two storms can be observed, which are then divided into Figures 6 and 7 for detailed analysis.

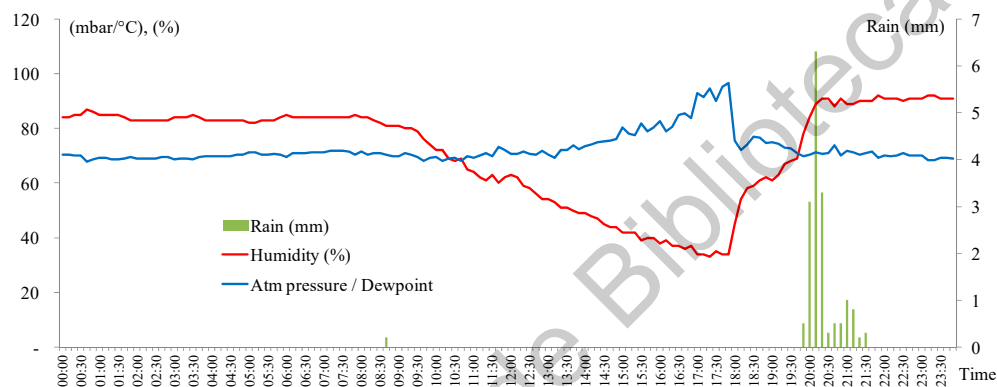


Figure 6. Results of applying the CRHUDA model at the automatic meteorological station Candiles on June 24, 2013.

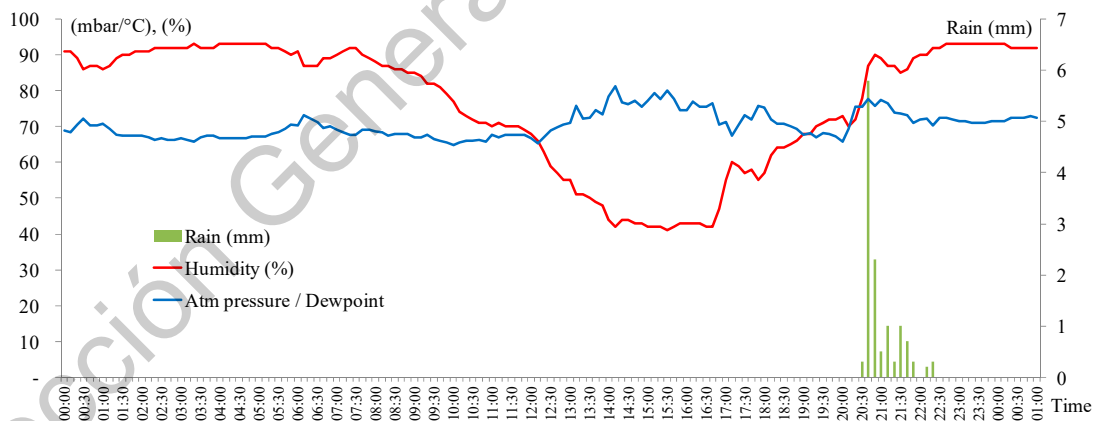


Figure 7. Results of applying the CRHUDA model at the automatic meteorological station Candiles on June 25, 2013.

Tables 1 and 2 show the values of the climatic variables used in the CRHUDA model for time t_1 (warning) and t_2 (start of precipitation), respectively, at the EMA Candiles on June 24, 2013.

Table 1. Climatological values at time t1 $H_{t1} = C_{t1}$; ($S1 \cap S2$) at EMA Candiles on June 24, 2013.

Date	Time	Humidity (%)	Atm Pressure/Dewpoint
24/06/2013	09:50	74	69.240
24/06/2013	10:00	72	69.447
24/06/2013	10:10	72	68.095
24/06/2013	10:20	69	68.961
24/06/2013	10:30	68	69.134
24/06/2013	10:40	69	68.076
24/06/2013	10:50	65	69.758
24/06/2013	11:00	64	69.105

Table 2. Climatological values at time t2 $H_{t2} = C_{t2}$ at EMA Candiles on June 24, 2013.

Date	Time	Humidity (%)	Atm Pressure/Dewpoint	Rain (mm)
24/06/2013	19:00	61	75.069	0
24/06/2013	19:10	63	74.398	0
24/06/2013	19:20	67	72.876	0
24/06/2013	19:30	68	72.679	0
24/06/2013	19:40	69	71.039	0
24/06/2013	19:50	78	69.783	0.5
24/06/2013	20:00	84	70.342	3.1
24/06/2013	20:10	89	71.287	6.3
24/06/2013	20:20	91	70.559	3.3
24/06/2013	20:30	91	71.056	0.3
24/06/2013	20:40	88	73.763	0.5
24/06/2013	20:50	91	70.110	0.5
24/06/2013	21:00	89	71.843	1
24/06/2013	21:10	89	71.365	0.8
24/06/2013	21:20	90	70.502	0.2
24/06/2013	21:30	90	70.997	0.3
24/06/2013	21:40	90	71.520	0

Figures 8–11 show an example of the results after applying the CRHUDA model to the historical data series of the most representative stations of the RedCIAQ. The figures show the model application to convective storm data. For the storms, the total rainfall depth and the return period corresponding to storm duration are provided.

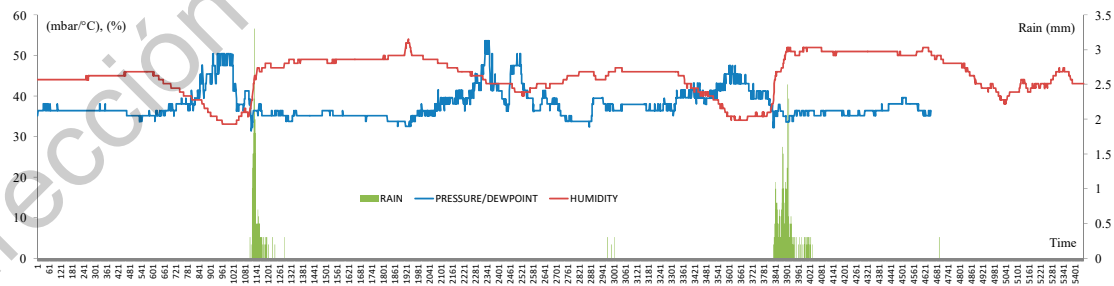


Figure 8. Results of applying the CRHUDA model to Centro Historico data on August 16, 2014 (max depth 10.9 mm in 1 h; max intensity 24.3 mm/h; 17-year return period).

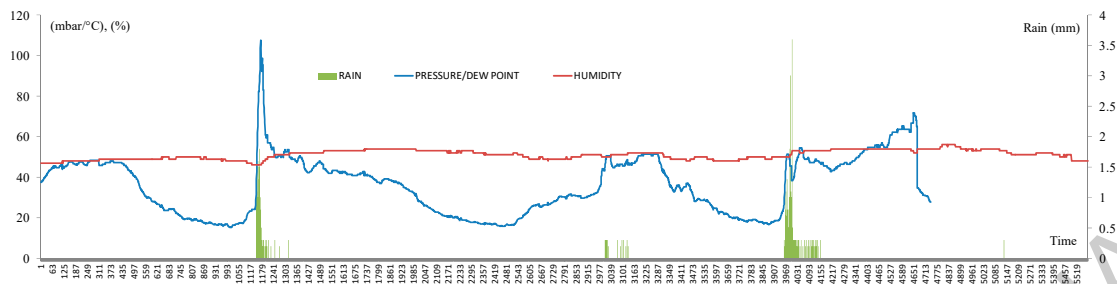


Figure 9. Results of applying the CRHUDA model to Cimatario data on August 16, 2014 (max depth 15.4 mm in 1 h; max intensity 41.2 mm/h; 32-year return period).

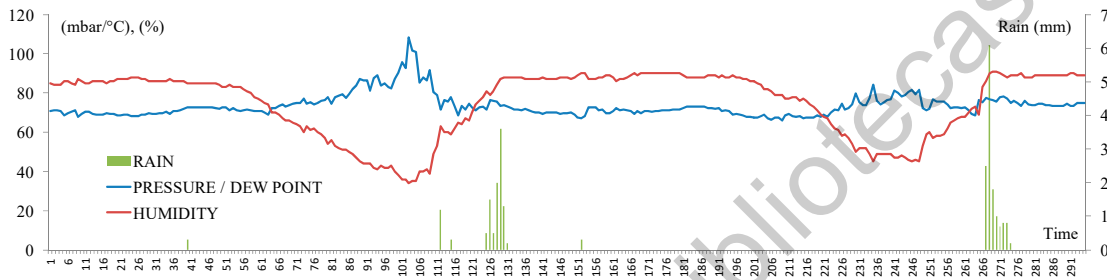


Figure 10. Results of applying the CRHUDA model to Milenio data on August 16, 2014 (max depth 12.1 mm in 1 h; max intensity 42.3 mm/h; 37-year return period).

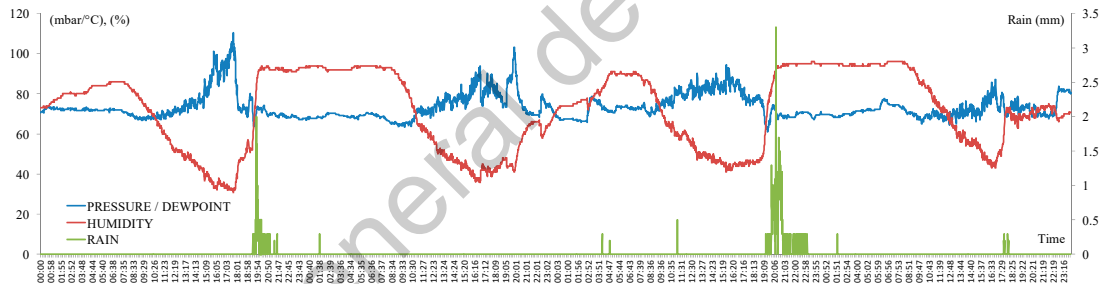


Figure 11. Results of applying the CRHUDA model to UAQ Cerro Campanas data on August 16, 2014 (max depth 55.2 mm in 1 h; max intensity 63.8 mm/h; 53-year return period).

In order to validate the CRHUDA model throughout the territory of Mexico, the model was executed in the same way as 17 out of the 32 states of Mexico. Figures 12–15 show the results of the crossing between the humidity, dewpoint, and atmospheric pressure model. In addition to the 523 storms analyzed in Queretaro, a total of 714 storms were analyzed in 16 Mexican states. Table 3 shows the number of storms and the names of the EMA stations used in each state. Thus, the total number of events analyzed was 1237 convective storms between 1999 and 2018. This database also includes storms caused by hurricanes.

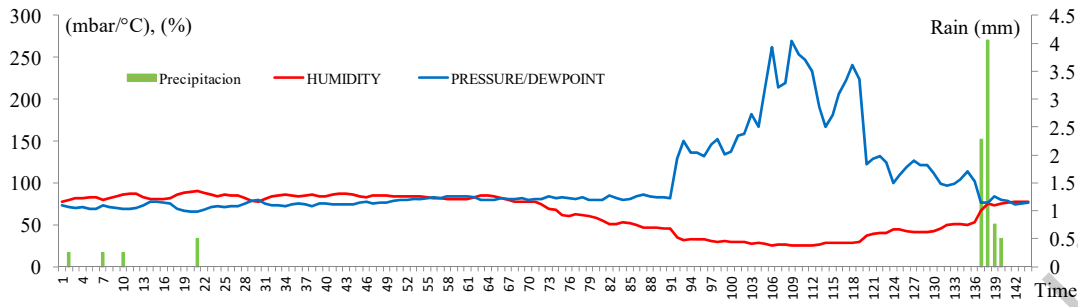


Figure 12. Results of applying the CRHUDA model to Atlacomulco data on April 23, 2018.

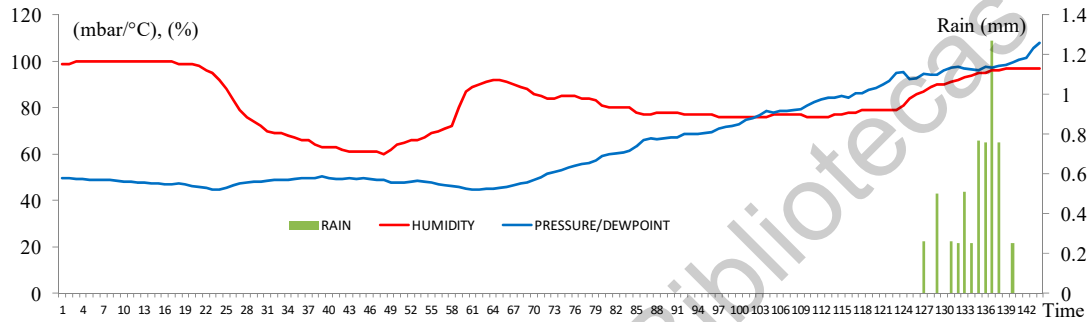


Figure 13. Results of applying the CRHUDA model to B. Del Tordo data on November 12, 2018.

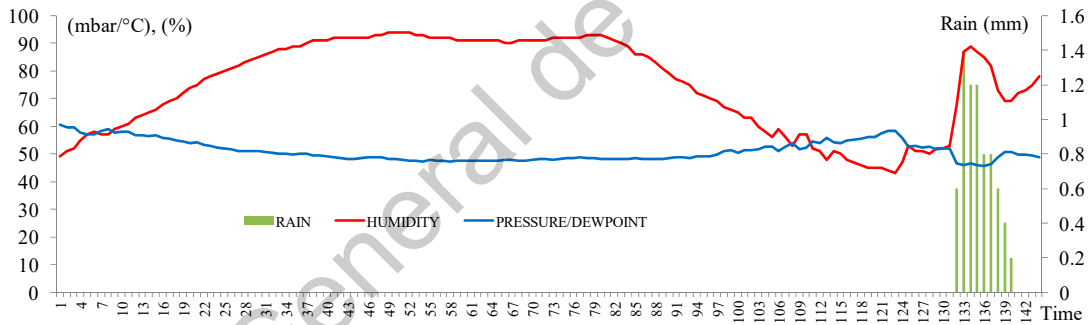


Figure 14. Results of applying the CRHUDA model to Dzilbachen data on February 22, 2018.

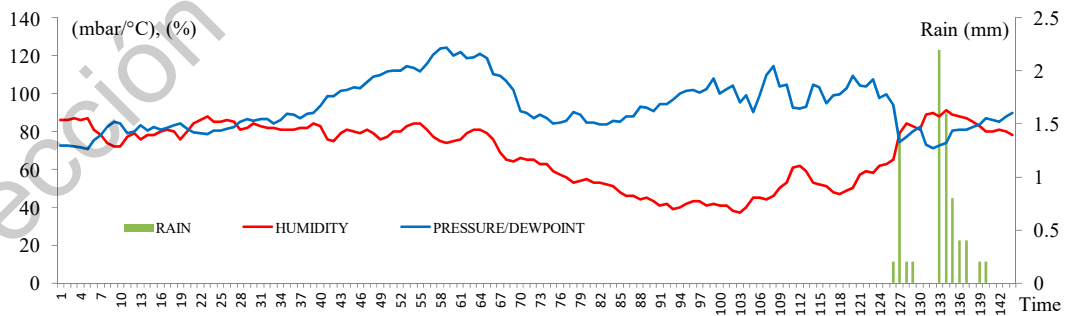


Figure 15. Results of applying the CRHUDA model to Ecoguardas data on April 6, 2018.

Table 3. Location of each automatic meteorological station (EMA) and the number of storms by state.

	State ID	Storms	EMA's Name
1	Aguascalientes	47	Calvillo
2	Baja California Sur	6	Cabo San Lucas
3	Campeche	12	Dzilbachen
4	Chiapas	85	El Triunfo
5	Chihuahua	48	Basaseachic
6	Coahuila	19	Cuatro Cienegas
7	DF	76	Ecoguardas
8	Edo. de Mexico	160	Atlacomulco, Cerro Catedral
9	Guerrero	1	Ciudad Altamirano
10	Hidalgo	51	El Chico
11	Jalisco	18	Chamela-Cuixmala
12	Oaxaca	41	Benito Juarez
13	San Luis Potosi	35	Ciudad Fernandez, Ciudad Valle
14	Sinaloa	24	El Fuerte
15	Tamaulipas	32	B. Del Tordo
16	Veracruz	59	Ciudad Aleman, Coscomatepec
	Subtotal	714	
17	Queretaro	523	See Appendix A, for details (34 EMA)
	Total	1237	

4. Discussion

A strong relationship between surface temperature and precipitation forecast has been reported in the literature, in addition to how these climatic variables impact streamflow ensemble forecasting [55]. According to [56], the potential for rainfall forecasts to be used in hydrological models to predict river flow depends on the response of the basin to earlier events and on the timing of the present event. The question is: What is the forecast time of rainy events? One implication of this is the possibility that a good rainfall forecasting model could be used in an early warning system for floods. Prior studies have noted the importance of short time scales, and that more extreme precipitation is more sensitive to temperature changes. Understanding how precipitation characteristics change in response to climatic elements provides new insight into convective organization and the structure of short-duration storms [33]. The present study was designed to find the effect of the Clausius–Clapeyron relationship and combination of atmospheric pressure, dewpoint, and humidity as the variables that cause most meteorological phenomena, in particular, precipitation. The results of this study show that it is possible to combine the climatic variables mentioned in two series that can be synoptically plotted to determine where both series cross, and this usually occurs, on average, between 9 and 10 h before the start of precipitation. It can therefore be assumed that the CRHUDA model includes the C–C relationship and, additionally, it allows combining of the aforementioned variables into a simple model for forecasting the onset of precipitation. These results are consistent with those of other studies that suggest that the dependency on surface dewpoint temperature follows two times the C–C relation, supported by the simple physical argument that this 2C–C behavior arises from the physics of convective clouds [34]. It is somewhat surprising that CRHUDA, as a simple synoptic model, can predict the onset of precipitation in a trustworthy way. Between 9 and 10 h beforehand, the forecast seems acceptable; this is also in accordance with our earlier observations which showed that it is possible to use the minute-by-minute hydrometeorological real-time dataset. Hence, it can be suggested that a daily rainfall database allows the correct spatial–temporal disaggregation in spatially distributed hourly rainfall [57]. Although these results differ from some published studies [58], models such as GEFSRv2 and SREF tend to overforecast light to moderate precipitation and underforecast heavy precipitation. Hence, a rainfall disaggregation model to which the temperature has been added as a driver should offer a more realistic assessment of future precipitation extremes [59]. According to [35], in a graph of rain intensity versus temperature, the crossover point between the two series always occurs at the onset of precipitation. This finding

has important implications for the development of a precipitation forecasting model for providing forecasts 9–10 h in advance, as it is only necessary to find the right combination of climate variables. Therefore, the results of the CRHUDA model suggest its applicability in the Mexican territory.

In most cases, the start of precipitation was precise. In other words, in almost all cases, $\Delta t = 0$. Although there are some events that anticipated crossings and a delay in precipitation starting point happened as shown in Figures 13 and 14, the proposed model is able to provide estimates between 9 and 10 h in advance. Figure 16 shows the frequency histogram of the warning times for Queretaro stations in 2018. Figure 17 shows the frequency histogram of the proportionality coefficient ϕ_1 for all of the Queretaro EMAs in 2018.

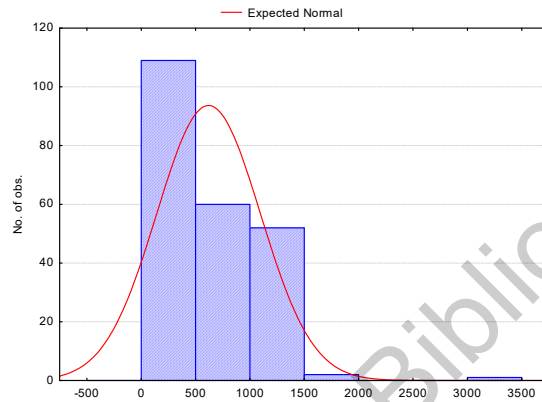


Figure 16. Frequency histogram of the warning times (in minutes) $T_t + \Delta t$ for all of the Queretaro EMAs in 2018.

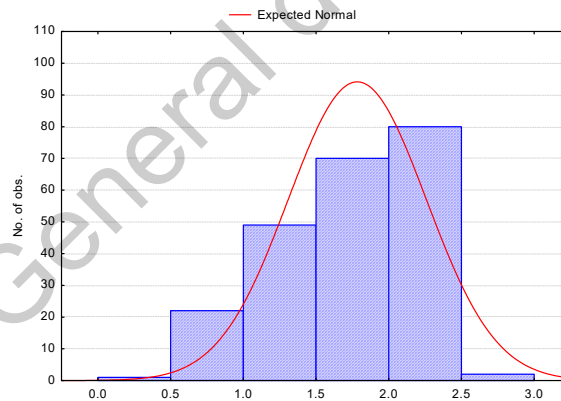


Figure 17. Frequency histogram of proportionality coefficient ϕ_1 for all of the Queretaro EMAs in 2018.

Regarding the proportionality coefficient or the parameter of an AR(1) model, it is important to mention that these coefficients were obtained in all cases. However, the hydroinformatics tool works, by default, with a value equal to 1. This is because it was shown that the coefficient of the AR(1) model only allows moving the humidity series within the graph, as shown in Figure 18.

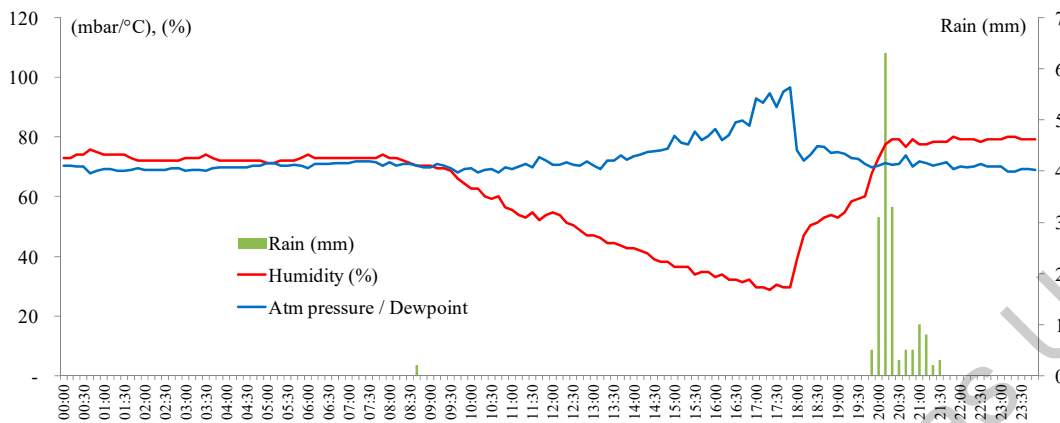


Figure 18. Results of applying the CRHUDA model at the automatic meteorological station Candiles on June 24, 2013 with $r_1 = 0.8746$.

It is interesting to compare Figures 6 and 18, since they concern the same EMA station during the same storm, but the parameter of the AR(1) model has been included. It is observed that the values of forecast time and precipitation occurrence do not vary. However, when scaling the series, the crossing time is clearer. This procedure is carried out in real time on days when the crossing of series is not clear. That is to say, for greater precision, the stochastic parameters of the time series are used.

To finish in terms of propositional logic, the truth table construction is shown below (Table 4). It starts with the two main premises: If there is a crossing of the two proposed series in time ($t1 : S1 \subseteq S2$) then, some hours later, the T_t series will cross again ($t2 : S1 \cap S2$) and, at that moment, $T_t + \Delta t$, precipitation will begin. If there is no crossing, then there will be no rain.

Table 4. Propositional logic truth table of the CRHUDA model.

$S1 \subseteq S2$	$S1 \cap S2$	$(S1 \subseteq S2) \rightarrow (S1 \cap S2)$	$\sim(S1 \subseteq S2) \rightarrow \sim(S1 \cap S2)$	$(S1 \subseteq S2) \rightarrow (S1 \cap S2) \wedge \dots$ $\dots \sim(S1 \subseteq S2) \rightarrow \sim(S1 \cap S2)$
True	True	True	True	True
True	False	False	True	False
False	True	True	False	False
False	False	True	True	True

It is concluded that it is, in fact, a valid model of the type contingency; a compound proposition which is sometimes true and sometimes false. That is to say, contingent truth, or truth, in fact, is understood as that proposition that can be true or false, according to the values of its constituent propositions.

5. Conclusions

Mexico’s National Center for Disaster Prevention establishes that one of the mechanisms used by the Mexican government to protect against and mitigate damage caused by disasters of different types is an early warning system. To guarantee its proper functioning, the coordinated participation of scientific groups, technical agencies, those responsible for communication and dissemination, as well as the population itself is required. The National Center for Disaster Prevention emphasizes warnings: “you must keep in mind that a clear and timely warning, along with knowledge of what is expected and how to react, makes a great difference for people and their communities”. Therefore, it is essential to inform the population in an accessible way and in sufficient time to mitigate the effects of natural phenomena [60]. The proposed model is a simple model that can surprisingly predict the starting point of precipitation 9–10 h in advance.

This paper has highlighted the relevance of having a simple forecast precipitation model and the reasons for the widespread use Clausius–Clapeyron relation. This paper has argued that a combination of crossing humidity, dewpoint, and atmospheric pressure is the best instrument for making a synoptic forecast of precipitation onset. Two sets of series were considered: (i) humidity and (ii) the relationships between atmospheric pressure and dewpoint. This phenomenon of crossing variables was verified through the analysis of 1237 storms. Recent studies show that the time series of temperature and humidity, when properly combined, can generate a trusted precipitation forecasting model [61].

Taken together, these results suggest that current early warning systems could be based on measurements of rainfall intensity. Those systems could be combined with monitoring of water levels [62,63]. These findings suggest several courses of action to record, send, and monitor floods using a new generation of smart water level gauges [64]. Another important practical implication is, for example, as occurred in Tabasco, Mexico, that the climatic information from a flood early warning system is transmitted in real time and published using a social network (Twitter) by radio frequency (915 MHz) using LoRa modulation [65]. An implication of this is the possibility that in the LAC region, a simple extreme rainfall warning system can be implemented at low cost, due to the fact that the application of the CRHUDA model only requires three measured climatic variables. To date, only precipitation forecast models based on the principle of analogue prediction are capable of producing accurate forecasts with a 6 to 8 h lead time in forecasting [59]. Currently, in other countries, the simple implementation of extreme rainfall warning systems combined with flood simulation models is a priority. Many developing countries are working to cut the number of fatalities due to flash floods, improve the efficiency of disaster risk reduction efforts, and play an important role in strengthening the resilience to climate change [66]. In these cases, CRHUDA could be an algorithm that is easy to implement.

At present, we are working on characterizing the spatial–temporal relationship of the parameters of the AR(1) models in the used climatological series. It is expected that it will soon be possible to demonstrate that the parameters of these AR(1) models vary with some differences in spatial patterns in the Mexican territory. For now, it was important to present the model and its application to other regions.

Returning to the suggestion posed at the beginning of this study, we presented a simple forecasting model for precipitation onset based on crossing humidity, dewpoint, and atmospheric pressure. Therefore, there is a definite need to continue research into estimating the uncertainty of precipitation onset and many further improvements are required, including the inputs of EWS [67]. The CRHUDA model was already tested during the current rainy season (2019), achieving great precision in forecasting the start of precipitation. The authors invite researchers to apply the CRHUDA model to their time series to validate this forecasting model.

Author Contributions: Conceptualization, formal analysis, methodology and validation A.G.-L.; software, data curation, and writing—original draft preparation, I.C.-P. M.M.M. contributed to the systematization of the calculations of the model CRHUDA, resources, software.

Funding: This research received funding of Secretaria de Educacion Publica (SEP-Mexico). Programa para el Desarrollo Profesional Docente (PRODEP) and was supported by the Fund for the Reinforcement of Research UAQ-2018 Research and Postgraduate Division (FOFIUAQ/FIN201911).

Acknowledgments: The authors are grateful to the Risk Management Unit of the UNESCO Regional Office of Science for Latin America and the Caribbean. We express our sincere gratitude to Josue Elizondo Gomez, Miguel Angel Sanchez Quijano and Bruno Miguel Paz Aviña for discussions on the topic of this work. We acknowledge the constructive comments of the three anonymous reviewers.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. RedCIAQ Weather Stations.

Id	Name	Long	Lat	Id	Name	Long	Lat
1	Chulavista	−100.47	20.63	18	Ezequiel Montes	−99.90	20.67
2	Belén	−100.41	20.65	19	Huimilpan	−100.27	20.37
3	Real del Parque	−100.40	20.61	20	Landa de Matamoros	−99.32	21.18
4	Candiles	−100.40	20.55	21	Pedro Escobedo	−100.14	20.50
5	Cimatario	−100.38	20.56	22	San Joaquín	−100.01	20.38
6	Centro histórico	−100.39	20.59	23	San Juan del Rio	−99.97	20.39
7	Milenio III	−100.35	20.59	24	Tequisquiapan	−99.91	20.61
8	San gil	−100.44	20.70	25	Tolimán	−99.93	20.90
9	El refugio	−100.35	20.65	26	Viñedos	−100.49	20.61
10	COTAS Amazcala	−100.34	20.71	27	El esparrago	−100.01	20.38
11	Cerro de las Campanas	−100.41	20.59	28	Santa Rosa Jáuregui	−100.45	20.74
12	Amealco de Bonfil	−100.14	20.19	29	Unión de Ejidos	−100.23	20.65
13	Pinal de Amoles	−99.63	21.14	30	Joaquín Herrera	−99.57	20.92
14	Arroyo seco	−99.69	21.55	31	Juriquilla	−100.45	20.72
15	Cadereyta de Montes	−99.81	20.70	32	UAQ Aeropuerto	−100.37	20.62
16	Colón	−100.05	20.78	33	Pasteur y 57	−100.38	20.58
17	Corregidora	−100.43	20.55	34	CICATA QRO IPN	−100.37	20.57

References

1. CENAPRED. *Disasters in Mexico: Social and Economic Impacts (1980–2014)*; Centro Nacional de Prevención de Desastres: Secretaría de Gobernación, Mexico, 2016.
2. Gutierrez-Lopez, A.; Fortanell Trejo, M.; Albuquerque Gonzalez, N.; Bravo Prado, F. Análisis de la variabilidad espacial en la precipitación en la zona metropolitana de Querétaro empleando ecuaciones de anisotropía. In *Investigaciones Geográficas*; Instituto de Geografía, Universidad Autónoma de México: Ciudad de México, Mexico, 2019. [\[CrossRef\]](#)
3. Lepore, C.; Allen, J.; Tippet, M. Relationships between Hourly Rainfall Intensity and Atmospheric Variables over the Contiguous United States. *J. Clim.* **2016**, *29*, 3181–3197. [\[CrossRef\]](#)
4. Saltzman, B. On the Maintenance of the Large-Scale Quasi-Permanent Disturbances in the Atmosphere. *Tellus* **1959**, *11*, 425–431. [\[CrossRef\]](#)
5. Saltzman, B. *Dynamical Paleoclimatology*, 1st ed.; Academic Press: New York, NY, USA, 2002.
6. Lorenz, E. Deterministic Nonperiodic Flow. *J. Atmos. Sci.* **1963**, *20*, 130–141. [\[CrossRef\]](#)
7. Berg, P.; Haerter, J. Unexpected increase in precipitation intensity with temperature—A result of mixing of precipitation types? *Atmos. Res.* **2013**, *119*, 56–61. [\[CrossRef\]](#)
8. Holley, D.; Dorling, S.; Steele, C.; Earl, N. A climatology of convective available potential energy in Great Britain. *Int. J. Climatol.* **2014**, *34*, 3811–3824. [\[CrossRef\]](#)
9. Lekouch, I.; Lekouch, K.; Muselli, M.; Mongruel, A.; Kabbachi, B.; Beysens, D. Rooftop dew, fog and rain collection in southwest Morocco and predictive dew modeling using neural networks. *J. Hydrol. (Amst.)* **2012**, *448*, 60–72. [\[CrossRef\]](#)
10. Park, I.; Min, S. Role of Convective Precipitation in the Relationship between Subdaily Extreme Precipitation and Temperature. *J. Clim.* **2017**, *30*, 9527–9537. [\[CrossRef\]](#)
11. Dyson, L.; van Heerden, J.; Sumner, P. A baseline climatology of sounding-derived parameters associated with heavy rainfall over Gauteng, South Africa. *Int. J. Climatol.* **2014**, *35*, 114–127. [\[CrossRef\]](#)
12. Omotosho, J. Equivalent potential temperature and dust haze forecasting at Kano, Nigeria. *Atmos. Res.* **1989**, *23*, 163–178. [\[CrossRef\]](#)
13. Puvaneswaran, M. Climatic classification for Queensland using multivariate statistical techniques. *Int. J. Climatol.* **1990**, *10*, 591–608. [\[CrossRef\]](#)
14. Damrath, U.; Doms, G.; Frühwald, D.; Heise, E.; Richter, B.; Steppeler, J. Operational quantitative precipitation forecasting at the German Weather Service. *J. Hydrol. (Amst.)* **2000**, *239*, 260–285. [\[CrossRef\]](#)
15. Rasouli, K.; Hsieh, W.; Cannon, A. Daily streamflow forecasting by machine learning methods with weather and climate inputs. *J. Hydrol. (Amst.)* **2012**, *414*, 284–293. [\[CrossRef\]](#)

16. Moon, S.; Kim, Y.; Lee, Y.; Moon, B. Application of machine learning to an early warning system for very short-term heavy rainfall. *J. Hydrol. (Amst.)* **2019**, *568*, 1042–1054. [[CrossRef](#)]
17. Zahraei, A.; Hsu, K.; Sorooshian, S.; Gourley, J.; Hong, Y.; Behrangi, A. Short-term quantitative precipitation forecasting using an object-based approach. *J. Hydrol. (Amst.)* **2013**, *483*, 1–15. [[CrossRef](#)]
18. Li, P.; Lai, E. Short-range quantitative precipitation forecasting in Hong Kong. *J. Hydrol. (Amst.)* **2004**, *288*, 189–209. [[CrossRef](#)]
19. Carrera-Hernández, J.; Gaskin, S. Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico. *J. Hydrol. (Amst.)* **2007**, *336*, 231–249. [[CrossRef](#)]
20. Carter, M.; Elsner, J.; Bennett, S. A quantitative precipitation forecast experiment for Puerto Rico. *J. Hydrol. (Amst.)* **2000**, *239*, 162–178. [[CrossRef](#)]
21. Valverde Ramírez, M.; de Campos Velho, H.; Ferreira, N. Artificial neural network technique for rainfall forecasting applied to the São Paulo region. *J. Hydrol. (Amst.)* **2005**, *301*, 146–162. [[CrossRef](#)]
22. Hou, T.; Kong, F.; Chen, X.; Lei, H. Impact of 3DVAR Data Assimilation on the Prediction of Heavy Rainfall over Southern China. *Adv. Meteorol.* **2013**, *2013*, 129642. [[CrossRef](#)]
23. Wang, J.; Gaffen, D. Late-Twentieth-Century Climatology and Trends of Surface Humidity and Temperature in China. *J. Clim.* **2001**, *14*, 2833–2845. [[CrossRef](#)]
24. Suparta, W.; Alhasa, K.; Singh, M. Estimation water vapor content using the mixing ratio method and validated with the ANFIS PWV model. *J. Phys. Conf. Ser.* **2017**, *852*, 012041. [[CrossRef](#)]
25. Camuffo, D. Theoretical Grounds for Humidity. In *Microclimate for Cultural Heritage*; Elsevier: Amsterdam, The Netherlands, 2014; Chapter 2A; pp. 49–76. ISBN 9780444632982. [[CrossRef](#)]
26. Pumo, D.; Carlino, G.; Blenkinsop, S.; Arnone, E.; Fowler, H.; Noto, L. Sensitivity of extreme rainfall to temperature in semi-arid Mediterranean regions. *Atmos. Res.* **2019**, *225*, 30–44. [[CrossRef](#)]
27. Capparelli, A. *Fisicoquímica Básica*, 1st ed.; Universidad Nacional La Plata Argentina: La Plata, Argentina, 2013.
28. Van der Dussen, J.; de Roode, S.; Siebesma, A. Factors Controlling Rapid Stratocumulus Cloud Thinning. *J. Atmos. Sci.* **2014**, *71*, 655–664. [[CrossRef](#)]
29. Romps, D. Clausius–Clapeyron Scaling of CAPE from Analytical Solutions to RCE. *J. Atmos. Sci.* **2016**, *73*, 3719–3737. [[CrossRef](#)]
30. Agard, V.; Emanuel, K. Clausius–Clapeyron Scaling of Peak CAPE in Continental Convective Storm Environments. *J. Atmos. Sci.* **2017**, *74*, 3043–3054. [[CrossRef](#)]
31. Lorenz, D.; DeWeaver, E. The Response of the Extratropical Hydrological Cycle to Global Warming. *J. Clim.* **2007**, *20*, 3470–3484. [[CrossRef](#)]
32. Romps, D. An Analytical Model for Tropical Relative Humidity. *J. Clim.* **2014**, *27*, 7432–7449. [[CrossRef](#)]
33. Chang, W.; Stein, M.; Wang, J.; Kotamarthi, V.; Moyer, E. Changes in Spatiotemporal Precipitation Patterns in Changing Climate Conditions. *J. Clim.* **2016**, *29*, 8355–8376. [[CrossRef](#)]
34. Lenderink, G.; Barbero, R.; Loriaux, J.; Fowler, H. Super-Clausius–Clapeyron Scaling of Extreme Hourly Convective Precipitation and Its Relation to Large-Scale Atmospheric Conditions. *J. Clim.* **2017**, *30*, 6037–6052. [[CrossRef](#)]
35. Bürger, G.; Heistermann, M.; Bronstert, A. Towards Subdaily Rainfall Disaggregation via Clausius–Clapeyron. *J. Hydrometeorol.* **2014**, *15*, 1303–1311. [[CrossRef](#)]
36. Peleg, N.; Marra, F.; Fatichi, S.; Molnar, P.; Morin, E.; Sharma, A.; Burlando, P. Intensification of Convective Rain Cells at Warmer Temperatures Observed from High-Resolution Weather Radar Data. *J. Hydrometeorol.* **2018**, *19*, 715–726. [[CrossRef](#)]
37. Velasco, S.; Fernández-Pineda, C. Sobre la obtención de la ecuación de Clapeyron-Clausius. *Rev. Española Física* **2008**, *22*, 7–14.
38. Seidel, T.; Grant, A.; Pszenny, A.; Allman, D. Dewpoint and Humidity Measurements and Trends at the Summit of Mount Washington, New Hampshire, 1935–2004. *J. Clim.* **2007**, *20*, 5629–5641. [[CrossRef](#)]
39. Millán, H.; Ghanbarian-Alavijeh, B.; García-Fornaris, I. Nonlinear dynamics of mean daily temperature and dewpoint time series at Babolsar, Iran, 1961–2005. *Atmos. Res.* **2010**, *98*, 89–101. [[CrossRef](#)]
40. Harder, P.; Pomeroy, J. Estimating precipitation phase using a psychrometric energy balance method. *Hydrol. Process.* **2013**, *27*, 1901–1914. [[CrossRef](#)]

41. Dahm, R.; Bhardwaj, A.; Sperna Weiland, F.; Corzo, G.; Bouwer, L. A Temperature-Scaling Approach for Projecting Changes in Short Duration Rainfall Extremes from GCM Data. *Water (Basel)* **2019**, *11*, 313. [[CrossRef](#)]
42. Mohr, S.; Kunz, M. Recent trends and variabilities of convective parameters relevant for hail events in Germany and Europe. *Atmos. Res.* **2013**, *123*, 211–228. [[CrossRef](#)]
43. Myoung, B.; Nielsen-Gammon, J. Sensitivity of Monthly Convective Precipitation to Environmental Conditions. *J. Clim.* **2010**, *23*, 166–188. [[CrossRef](#)]
44. Gao, X.; Li, J.; Sorooshian, S. Modeling Intraseasonal Features of 2004 North American Monsoon Precipitation. *J. Clim.* **2007**, *20*, 1882–1896. [[CrossRef](#)]
45. Wang, Y.; Tang, L.; Zhang, J.; Gao, T.; Wang, Q.; Song, Y.; Hua, D. Investigation of Precipitable Water Vapor Obtained by Raman Lidar and Comprehensive Analyses with Meteorological Parameters in Xi'an. *Remote Sens. (Basel)* **2018**, *10*, 967. [[CrossRef](#)]
46. Sim, I.; Lee, O.; Kim, S. Sensitivity Analysis of Extreme Daily Rainfall Depth in Summer Season on Surface Air Temperature and Dew-Point Temperature. *Water (Basel)* **2019**, *11*, 771. [[CrossRef](#)]
47. Liu, Z.; Chen, B.; Chan, S.; Cao, Y.; Gao, Y.; Zhang, K.; Nichol, J. Analysis and modelling of water vapour and temperature changes in Hong Kong using a 40-year radiosonde record: 1973–2012. *Int. J. Climatol.* **2014**, *35*, 462–474. [[CrossRef](#)]
48. Shaw, S.; Royem, A.; Riha, S. The Relationship between Extreme Hourly Precipitation and Surface Temperature in Different Hydroclimatic Regions of the United States. *J. Hydrometeorol.* **2011**, *12*, 319–325. [[CrossRef](#)]
49. Aguilar, E.; Pastor, D.; Vázquez, A.Y.; Ibarra, D. Recolección de datos meteorológicos en tiempo real mediante el uso de funciones asíncronas non-blocking. *Rev. NTHE* **2018**, *24*, 113–117.
50. Gil, S.; Ramírez, G.; Muñoz, M.Y.; González, S. Implementación de un modelo de datos para el almacenamiento de información climatológica en el estado de Querétaro. *Rev. NTHE* **2018**, *24*, 16–19.
51. Vincent, L.; van Wijngaarden, W.; Hopkinson, R. Surface Temperature and Humidity Trends in Canada for 1953–2005. *J. Clim.* **2007**, *20*, 5100–5113. [[CrossRef](#)]
52. Rogers, J.; Wang, S.; Coleman, J. Evaluation of a Long-Term (1882–2005) Equivalent Temperature Time Series. *J. Clim.* **2007**, *20*, 4476–4485. [[CrossRef](#)]
53. Egerer, M.; Lin, B.; Kendal, D. Temperature Variability Differs in Urban Agroecosystems across Two Metropolitan Regions. *Climate* **2019**, *7*, 50. [[CrossRef](#)]
54. Emmanuel, L.; Houngouè, N.; Biaou, C.; Badou, D. Statistical Analysis of Recent and Future Rainfall and Temperature Variability in the Mono River Watershed (Benin, Togo). *Climate* **2019**, *7*, 8. [[CrossRef](#)]
55. Verkade, J.; Brown, J.; Reggiani, P.; Weerts, A. Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *J. Hydrol. (Amst.)* **2013**, *501*, 73–91. [[CrossRef](#)]
56. Yucel, I.; Onen, A.; Yilmaz, K.; Gochis, D. Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall. *J. Hydrol. (Amst.)* **2015**, *523*, 49–66. [[CrossRef](#)]
57. Segond, M.; Onof, C.; Wheeler, H. Spatial-temporal disaggregation of daily rainfall from a generalized linear model. *J. Hydrol. (Amst.)* **2006**, *331*, 674–689. [[CrossRef](#)]
58. Siddique, R.; Mejia, A.; Brown, J.; Reed, S.; Ahnert, P. Verification of precipitation forecasts from two numerical weather prediction models in the Middle Atlantic Region of the USA: A precursory analysis to hydrologic forecasting. *J. Hydrol. (Amst.)* **2015**, *529*, 1390–1406. [[CrossRef](#)]
59. Wu, M.; Lin, G. The very short-term rainfall forecasting for a mountainous watershed by means of an ensemble numerical weather prediction system in Taiwan. *J. Hydrol. (Amst.)* **2017**, *546*, 60–70. [[CrossRef](#)]
60. Bentley, M.; Stallins, J. Synoptic evolution of Midwestern US extreme dew point events. *Int. J. Climatol.* **2008**, *28*, 1213–1225. [[CrossRef](#)]
61. Danladi, A.; Stephen, M.; Aliyu, B.; Gaya, G.; Silikwa, N.; Machael, Y. Assessing the influence of weather parameters on rainfall to forecast river discharge based on short-term. *Alex. Eng. J.* **2018**, *57*, 1157–1162. [[CrossRef](#)]
62. Hofmann, J.; Schüttrumpf, H. Risk-Based Early Warning System for Pluvial Flash Floods: Approaches and Foundations. *Geosciences (Basel)* **2019**, *9*, 127. [[CrossRef](#)]
63. Borsch, S.; Khristoforov, A.; Krovotyntsev, V.; Leontieva, E.; Simonov, Y.; Zatyagalova, V. A Basin Approach to a Hydrological Service Delivery System in the Amur River Basin. *Geosciences (Basel)* **2018**, *8*, 93. [[CrossRef](#)]

64. Chang, C.; Chung, M.; Yang, S.; Hsu, C.; Wu, S. A Case Study for the Application of an Operational Two-Dimensional Real-Time Flooding Forecasting System and Smart Water Level Gauges on Roads in Tainan City, Taiwan. *Water (Basel)* **2018**, *10*, 574. [[CrossRef](#)]
65. Leon, E.; Alberoni, C.; Wister, M.; Hernández-Nolasco, J. Flood Early Warning System by Twitter Using LoRa. *Proceedings* **2018**, *2*, 1213. [[CrossRef](#)]
66. Hoedjes, J.; Kooiman, A.; Maathuis, B.; Said, M.; Becht, R.; Limo, A.; Mumo, M.; Nduhiu-Mathenge, J.; Shaka, A.; Su, B. A Conceptual Flash Flood Early Warning System for Africa, Based on Terrestrial Microwave Links and Flash Flood Guidance. *ISPRS Int. J. Geo Inf.* **2014**, *3*, 584–598. [[CrossRef](#)]
67. Cloke, H.; Pappenberger, F. Ensemble flood forecasting: A review. *J. Hydrol. (Amst.)* **2009**, *375*, 613–626. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Dirección General de Bibliotecas UAQ

Artículo en revista NTHE

Dirección General de Bibliotecas UNQ

Implementación de un modelo de datos para el almacenamiento de información climatológica en el estado de Querétaro

Sarai de Jesús Gil Ramos¹, Gabriela Xicoténcatl Ramírez²,
Martín Muñoz Mandujano³ y Sharol Rubí González Serrano⁴

Resumen

El presente artículo explica la implementación de un modelo de datos para el almacenamiento de información climatológica en el estado de Querétaro; se trata de un modelo de datos relacional. De la misma forma, se implementó una tabla transaccional, así como una tabla histórica, donde se almacena toda la serie de tiempo de los datos climatológicos. Como resultado, se obtiene una base de datos que almacena una gran cantidad de información sobre el comportamiento de las distintas variables meteorológicas, que son posteriormente analizadas para dar a conocer a la población lo que sucede con las condiciones climatológicas de nuestra región.

Palabras clave: almacenamiento, base de datos, clima, implementación, MariaDB

¹sgil13@alumnos.uaq.mx

²gabyxico@uaq.mx

Facultad de Informática, Universidad Autónoma de Querétaro

³mmunoz23@alumnos.uaq.mx

⁴sharol9813@gmail.com

^{1,3 y 4}Centro de Desarrollo de la Facultad de Informática

Universidad Autónoma de Querétaro

Av. de las Ciencias s/n, Juriquilla, C.P. 76230

Santiago de Querétaro, Querétaro, México

Fecha de recepción:
28 de septiembre de 2018

Fecha de aceptación:
05 de diciembre de 2018

Introducción

Diferentes fenómenos meteorológicos, tales como lluvia, nieve, huracanes, entre otros, son los que se presentan en nuestro entorno diariamente. Tener conocimiento de ellos, así como de las razones por las cuales se originan y qué elementos climatológicos influyen en su aparición, puede resultar útil, más aún si nos encontramos en zonas donde dichos fenómenos causan, frecuentemente, estragos catastróficos. Para poder empezar a comprender estos conceptos, es necesario comenzar desde términos básicos, por ejemplo, el clima, sus variables meteorológicas y las repercusiones que estas últimas tienen al momento de analizarlas o manejarlas.

La Real Academia Española menciona que el clima es “un conjunto de características y condiciones atmosféricas que una región tiene” (RAE, 2018). Por otro lado, es importante resaltar que las variables meteorológicas ayudan en diversas ocasiones a realizar distintas mediciones, análisis, investigaciones, instrumentos de estudio y un sinnúmero de actividades que permiten dar a conocer el estado climatológico de una zona específica.

Para el caso de nuestra entidad, uno de los usos que se le puede otorgar a las variables meteorológicas es el del análisis, es decir, recolectar estos datos para, primeramente, identificar los valores que generaron, descubrir por medio de ellos por qué se generó ese estado y, posteriormente, realizar algún pronóstico sobre futuros estados del clima que pudieran desencadenar algún fenómeno meteorológico.

Debido a lo anterior, se considera necesario contar con una herramienta que permita el monitoreo del comportamiento meteorológico en nuestra entidad, de manera que pueda difundirse en una plataforma web que sea accesible para el público; así, se podrá dar a conocer información relevante para que se puedan tomar medidas precautorias y para realizar análisis. Además, todo lo anterior podría utilizarse como instrumento para futuras investigaciones.

El nacimiento de la Red de Monitoreo de Precipitaciones Extremas en el Estado de Querétaro

(RedCIAQ, 2017) se debe a esa importante necesidad. Es un proyecto académico que provee un pronóstico del clima y, además, de genera diariamente una cantidad considerable de información que se almacena en una base de datos, para análisis de ciertos fenómenos meteorológicos que se presentan en el estado de Querétaro (Pérez, 2015).

Al ser un volumen alto de información, nos enfrentamos a los siguientes retos: primeramente, un problema de elección del proceso más adecuado para almacenar todos estos datos meteorológicos; en consecuencia, tomando en cuenta la cantidad de datos que se recaban, almacenan y muestran, ¿qué opción se tiene para que, cuando el sistema esté siendo usado por varios usuarios, el rendimiento y el tiempo de respuesta sean los adecuados? Se debe hacer un completo análisis de la estructura necesaria para el almacenamiento, para que, con ello, la transacción de cada dato sea lo más precisa posible, el tiempo de muestra de resultados sea menor y, además, se cumpla con las necesidades primordiales de RedCIAQ, que son la recolección y el análisis de datos climáticos en el estado de Querétaro, para con ellos poder proporcionar un pronóstico sobre probables fenómenos meteorológicos.

Por lo anterior, en este artículo se presenta el proceso de implementación de este modelo de datos del sistema de monitoreo de Red CIAQ, así como la forma en que se almacenan.

Fundamentación teórica

Una base de datos es considerada una colección de datos persistentes que pueden compartirse e interrelacionarse (Silberschatz, 2014). Algunas de sus propiedades más relevantes son:

- Persistencia
- Compartidas
- Interrelación

Las base de datos se pueden dividir en diferentes tipos; sin embargo, en este artículo se tendrán dos en consideración: 1) la base de datos relacional, que es un repositorio en el que se comparten datos, es decir,

se crean relaciones entre las tablas y a cada relación se le asigna un nombre único; 2) base de datos transaccional, que tiene como objetivo principal la recepción y el envío de datos a grandes velocidades. Es importante tener claro el concepto de transacción en el lenguaje de las bases de datos, que es el conjunto de comandos ejecutados al mismo tiempo, a fin de formar una unidad lógica de trabajo (Silberschatz, 2014), es decir, sentencias o comandos que se ejecutan en su totalidad o no y que hacen que la base de datos pase de un estado consistente a otro.

Otro concepto que debemos tener en cuenta es el de una “tabla transaccional”, que contiene los identificadores de datos estructurados que siguen abiertos, sin haberse cerrado aún, por cada transacción a la que tienen acceso los almacenamientos estructurados.

Metodología

Metodología empleada

Para el modelado de la base de datos relacional, se decidió utilizar como metodología una iterativa, que en este caso comienza con la creación del modelo *entidad relación*, mediante el cual se pasa de los requerimientos de la información a un esquema conceptual o lógico; de este último, se pasa a la elaboración de la base de datos (Gina, 2003). Lo anterior se explica mejor en la figura 1.

Identificación de requerimientos

El modelo de datos que requiere el sistema de la RedCIAQ debe tomar en cuenta, para su análisis, el almacén de varios de los valores climáticos de las diferentes estaciones meteorológicas en el estado de Querétaro, por ejemplo temperatura, humedad, dirección de viento, grado de dirección de viento, velocidad de viento, barómetro, lluvia, intensidad de lluvia, radiación solar, sensación térmica, índice de calor, punto de rocío.

Otro punto importante es que estos datos deben mostrarse en el sistema web, para actualizarse cada determinado tiempo y, aunque el sistema esté siendo utilizado de manera simultánea por múltiples usuarios, el tiempo de respuesta sea el menor posible.

Es necesario contar con un registro de las estaciones que supervisa el sistema y de quién dependen, así como su ubicación en coordenadas.

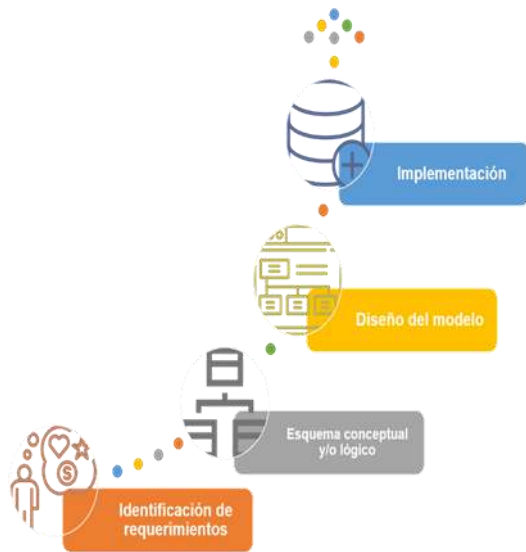


Figura 1. Metodología utilizada. Fuente: elaboración propia.

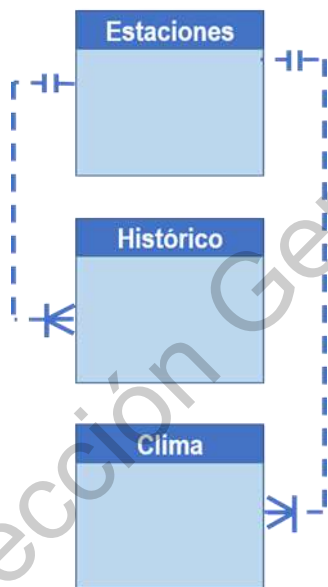


Figura 2. Modelo ER de la base de datos para el sistema RedCIAQ. Fuente: elaboración propia.

Elección de infraestructura para la BD

Se decidió emplear bases de datos relacionales, debido a sus características y a la facilidad para hacer modificaciones en la etapa de pruebas y desarrollo, así como a su rendimiento para procesar solicitudes provenientes de un sitio web.

MariaDB es un servidor de base de datos de código abierto (MariaDB Foundation, 2018a). En esencia, SQL MariaDB y MySQL —su competencia— son las dos mejores opciones para la base de datos relacional; son muy similares y, de hecho, son compatibles en conectores, sintaxis e instalación (MariaDB Foundation, 2018b).

Se eligió el servicio de SQL MariaDB sobre MySQL, debido a que la implementación que ofrece MariaDB posee una mayor gama de opciones de configuración, administración y rendimiento (Bartholomew, 2012).

Diseño del modelo

Como se muestra en la figura 2, se tienen 3 tablas en el modelo. Una de ellas (figura 2, tabla “Estaciones”) almacena los datos de cada una de las estaciones meteorológicas que se toman en cuenta para monitoreo de sus valores climáticos. En otra (figura 2, tabla “Clima”) se almacenan los valores climáticos que se generan cada día, recibidos de la tabla transaccional.

En la tabla transaccional (figura 2, tabla “Histórico”) se almacenan de manera temporal los valores climáticos de cada estación a cada minuto, para después vaciarlos y enviarlos a la tabla de “Histórico”, cada día.

Implementación

Como primer paso, se ha definido que, de las tres tablas que se tienen, dos de ellas serán relacionales y la última será transaccional; esto como parte del mecanismo para un mejor rendimiento al momento de mostrar los datos.

Tabla transaccional

La finalidad primordial de este tipo de tablas es recolectar y recuperar datos en un tiempo mínimo.

No existe la necesidad de preocuparse por redundancia o duplicación de información; además, una de las características que tiene y podemos aprovechar en gran medida es que permite conectar con tablas relacionales (Hernandez, 2015).

Tabla histórica

Una tabla de “Históricos” almacena y proporciona información de distintos periodos y mantiene automáticamente los datos históricos, es decir, pasados (Sarda, 1990).

Crear una tabla para datos históricos, separada de una tabla relacional, es una buena solución, ya que posee una complejidad simple al momento de su diseño e implementación, puesto que no requiere de cambios en las tablas relacionales originales.

Fundamentación de la elección de modelo de datos transaccional y con históricos

Para este modelo de datos, es necesaria la recolección de los valores climáticos de cada estación meteorológica por minuto, así como el almacenamiento y la renovación de datos por día. La decisión de ocupar tablas relacionales y transaccionales radica en dos puntos importantes:

1. Es una cantidad importante de datos los que se recolectan y muestran.
2. Se necesita un rendimiento óptimo al momento de mostrar todos esos datos, sin que lleguen a afectarse totalmente en caso de que se presente algún error.

Se elige la tabla transaccional porque cumple con estas funciones, incluidas en sus características, las cuales son: atomicidad, consistencia, aislamiento y durabilidad (Silberschatz, 2014). Además, se optó por crear una tabla que funcione como “Histórico” para los datos insertados en la tabla transaccional, de tal forma que permita generar consultas transaccionales a una mayor velocidad, así como limitar las consultas históricas solo cuando sea necesario.

Función para transacción de datos

Para que los datos de nuestra tabla transaccional se envíen a la tabla de histórico, se ha definido la creación de un evento SQL, en el cual se indica que se

almacenen, por día, los valores climáticos reunidos cada minuto, que se envíen a la tabla histórica y, finalmente, se vacíe la tabla transaccional. Ese proceso se repite diariamente a media noche.

Resultados

Actualmente, la base de datos posee un aproximado de 25 millones de datos; por lo tanto, al implementar la tabla transaccional, el rendimiento en el tiempo de respuesta ha sido optimizado; con ello, es posible acceder a la información actual con suma facilidad, ya que se evita cargar todos los datos históricos para consultas actuales. El tiempo concreto de respuesta al realizar una consulta a una estación específica es de 0.01 segundos. El proceso para acceder a la información histórica es lento, aunque son pocas las peticiones de ese tipo; el tiempo que tarda el servidor en responder a esta consulta es de 4.05 minutos.

Discusión y conclusiones

La base de datos posee ahora una velocidad de respuesta adecuada y se han hecho pruebas con esquemas no relacionales, como mongoddb; también se tiene como resultado un tiempo de respuesta similar al actual. Aunado a esto, su implementación posee una mayor complejidad, por lo que, de momento, al no proporcionar mejoras significativas, no es necesaria la migración hacia ese esquema no relacional.

Referencias

- Bartholomew D (2012). MariaDB vs. MySQL: similarities between MariaDB and MySQL. *Eandbsoftware*, 1-6.
- Gina MA (2003). Análisis, diseño e implementación de un sistema de información como soporte a la gestión académica para la escuela tecnológica de la Universidad Nacional De Piura. *Universidad Nacional de Piura*.
- Hernandez CH (2015). Tipos de base de datos. Tomado de: <https://www.mindmeister.com/es/596069815/tipos-de-base-de-datos>
- MariaDB Foundation (2018a). About MariaDB. Tomado de: <https://mariadb.org/about/>
- MariaDB Foundation (2018b). MariaDB vs. MySQL. Compatibility. MariaDB knowledge base.
- Pérez I (2015). Red CIAQ, el análisis científico de los fenómenos meteorológicos.
- Real Academia Española. (2018). Clima | Definición de clima. *Diccionario de la Lengua Española*. Edición del Tricentenario.
- RedCIAQ. (2017). Nosotros | RedCIAQ.
- Sarda, N. L. (1990). Extensions to SQL for Historical Databases. *IEEE Transactions on Knowledge and Data Engineering*, 2(2), 220–230. <https://doi.org/10.1109/69.54721>
- Silberschatz, A. (2014). *Fundamento de bases de datos*. España: The McGraw-Hill Companies.

Artículo en revista NTHE

Dirección General de Bibliotecas UNQ

Web scraping para la recopilación de datos meteorológicos

Martín Muñoz Mandujano¹, Juan Salvador Hernández Valerio², Sharol Rubí González Serrano³ y Aixa Pérez Liévana⁴

Resumen

El presente artículo explica el proceso por el cual fue desarrollado un programa en lenguaje Python para la obtención y almacenamiento de datos meteorológicos provenientes de diversas estaciones climáticas que se muestran en diferentes sitios web. Estos datos son almacenados dentro de una base de datos relacional para su análisis y procesamiento posterior dentro del sistema de monitoreo de la RedCIAQ. Actualmente, se poseen aproximadamente 25 millones de registros, provenientes de más de 30 estaciones.

¹mmunoz23@alumnos.uaq.mx

³sharol9813@gmail.com

⁴aixa.lievana@gmail.com

²valerio@uaq.mx

Facultad de Informática
Universidad Autónoma de Querétaro

^{1,3 y 4}Centro de Desarrollo de la Facultad de Informática
Universidad Autónoma de Querétaro
Universidad Autónoma de Querétaro
Av. de las Ciencias s/n, Juriquilla, C.P. 76230 Santiago de Querétaro, Querétaro, México

Fecha de recepción:
28 de septiembre de 2018

Fecha de aceptación:
05 de diciembre de 2018

Abstract

This article explains the process by which a program was developed in Python language for obtaining and storing meteorological data from several weather stations that are displayed on various websites. This data is stored within a relational database for analysis and further processing within the CIAQ Network monitoring system. Currently there are approximately 25 million records from more than 30 stations.

Palabras clave: clima, hidroinformática, monitoreo, *scraping*

Introducción

Actualmente, debido al cambio climático, realizar un monitoreo constante sobre las condiciones del tiempo ha tomado cada vez más importancia; por ello, existe una amplia gama de sistemas o redes de monitoreo y pronóstico climático, tanto de gobiernos como de empresas. Monitorear las condiciones climatológicas es de suma relevancia para algunas actividades productivas, como agricultura, ganadería y turismo, y afecta la vida diaria de las personas. Algunas de las redes de monitoreo existentes son el sistema del INIFAP para monitoreo y pronósticos, perteneciente a la dependencia gubernamental de SAGARPA (INIFAP, 2018), la plataforma de IBM Weather Underground —un sistema que muestra la información de una amplia gama de estaciones climatológicas— (The Weather Company, 2018) y el sistema WeatherLink, de Davis Instruments (Davis Instruments, 2018).

Un gran número de estaciones meteorológicas despliega su información mediante diversos sistemas web, uno de los cuales es el sistema de monitoreo RedCIAQ, el cual obtiene una amplia gama de información de múltiples estaciones para poder realizar cálculos meteorológicos. Si la obtención de la información de estos sistemas se realizara de manera manual, sería una labor prácticamente imposible, debido a la gran cantidad de datos que se

procesan y a que el tiempo requerido para evaluarlas (un minuto) es limitado. Por este motivo, fue necesario implementar un método que permitiera automatizar la recolección de datos y, en este sentido, *web scraping* fue la mejor opción. Dicho método es la práctica de la recopilación de datos que se muestran en una página web, diferente a la utilización de funciones propias de la página o a una persona empleando un navegador web; comúnmente, es un programa automático (Mitchell, 2015). Es decir, el *web scraping* consiste en enviar solicitudes a una determinada página web para obtener la información y realizar algún tratamiento en ella.

Metodología

Desarrollo

El programa que realiza el método *web scraping* se desarrolló con base en la metodología de desarrollo iterativo. Según Castro (2004):

El desarrollo iterativo es un método de construcción de productos cuyo ciclo de vida está compuesto por un conjunto de iteraciones, las cuales tienen como objetivo entregar versiones del *software*. Cada iteración se considera un subproyecto que genera productos de *software* y no solo documentación, permitiendo al usuario tener puntos de verificación y control más rápidos e induciendo un proceso continuo de pruebas y de integración desde las primeras iteraciones.

Durante el análisis inicial, se definieron aspectos primordiales del sistema. Primeramente, se realizó una toma de requerimientos con el personal de CIAQ y se estableció que la obtención de datos se debía realizar cada minuto en todas las estaciones que forman parte del sistema RedCIAQ.

Se decidió crear un programa desde cero debido a que el programa original, que se basaba en otro lenguaje de programación, requería demasiado

trabajo de mantenimiento para que pudiese alcanzar un óptimo funcionamiento y fácil mantenimiento.

El lenguaje que se eligió fue Python, debido a que la sencillez de su sintaxis permite desarrollar algoritmos complejos en un corto periodo de tiempo (Borrell, 2008); asimismo, permite que el mantenimiento del código sea más simple.

Como lo muestra la tabla 1, dentro de Python existen dos principales formas de realizar *web scraping*, utilizando la librería BeautifulSoup o la API Scrapy. Scrapy obtiene y devuelve los estilos utilizados en la página web, así como expresiones específicas encontradas, lo que dificulta la identificación del valor que corresponde a una métrica. BeautifulSoup busca y devuelve información en forma de código HTML, justo como se muestra en la página (González Peña, Lourenço, López-Fernández, Reboiro-Jato y Ferndáñez-Riverola, 2013). Con base en lo anterior, se decidió utilizar BeautifulSoup, debido a que este método permite realizar, de forma más autónoma, la detección de la información contenida en el sitio web; además, esta información posee un formato más simple para su procesamiento y posterior almacenamiento en una base de datos.

Herramienta	Tipo				Facilidades de Extracción
	C: Cliente HTTP	Domain-specific language	API/stand alone	Lenguaje	R: Expresiones Regulares H: HTML parsed tree X: XPath C: CSS selectors
UNIX shell (curl/wget, grep, sed, cut, paste, awk)	CP	No	SA	bash	R
Curl/libcurl	C	No	Ambos	C + bindings	Ninguna
Web-Harvest	F	Si	Ambos	Java	RX
Jsoup	CP	No	API	Java	HC
HttpClient	C	No	API	Java	Ninguna
JARVEST	F	Si	Ambos	JRuby/Java	RXC
WWW::Mechanize	CP	No	API	Perl	RX
Scrapy	F	No	Ambos	Python	RX
BeautifulSoup	P	No	No	Python	H

Tabla 1. Comparativa de herramientas de scraping en diferentes lenguajes de programación. Fuente: Elaboración propia con base a Glez-Peña (2013).

Siguiendo algunos ejemplos dentro del libro de Mitchell (2015), se realizó la primera versión del programa. Se emplearon otras librerías para que el programa pudiera hacer todo lo que se le pedía. Se utilizó la librería “requests” para realizar la petición HTTP a los sitios en que se muestra la información actual de cada estación meteorológica; la librería de “threading” para poder realizar el análisis de múltiples estaciones de manera simultánea; la librería de “time” para poder ejecutar de manera cíclica el análisis, y la librería de “MySQLdb” para obtener la dirección web o URL de las estaciones, así como para vaciar los datos obtenidos de cada estación.

En esta primera fase, se implementó el modelo de base de datos, que ya se encontraba en operación con el programa anterior.

La página de WeatherLink posee dos vistas para la información de una estación meteorológica; una de ellas solo posee información general del clima (figura 1), mientras que la otra es una vista completa de los datos del clima (figura 2).

Después del análisis inicial y su primera iteración, el programa realizaba un *scraping* óptimo en cuanto a tiempo de ejecución; sin embargo, con respecto de la cantidad de estaciones de las cuales se extraía información de manera correcta, oscilaba entre 8 y 10 por minuto de un total de 13 estaciones disponibles para pruebas, debido a que el programa requería ser modificado para filtrar datos nulos y normalizarlos en una sola unidad de medida para cada métrica.

El programa, estando en etapa inicial de desarrollo, se mostró al personal de CIAQ, que brindó una retroalimentación en cuanto a los datos almacenados y a la forma en que se requerían para su posterior despliegue en el sistema de RedCIAQ.

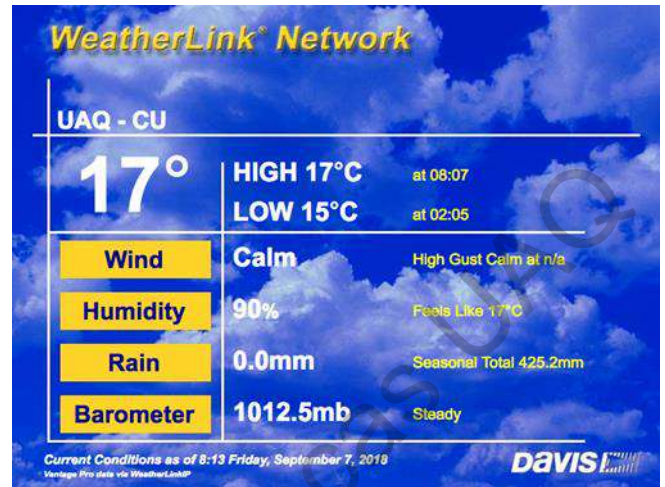
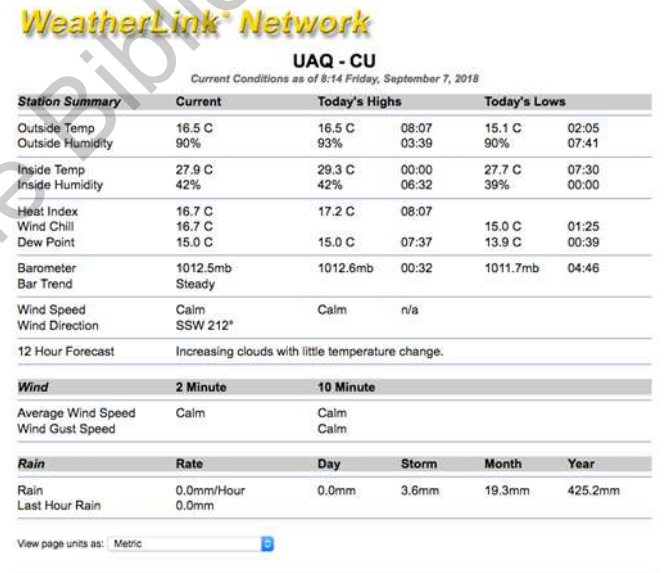


Figura 1. Vista simple de datos meteorológicos.
Fuente: weatherlink.com.



Fuente: weatherlink.com.

En una segunda iteración se filtraron los valores nulos. En el caso de “velocidad del viento”, en ocasiones se devolvía el parámetro “Calm”, el cual indica que la velocidad de viento es muy baja como para ser detectada por la estación. También fue necesario normalizar las unidades de medida de las métricas, puesto que, al provenir de diferentes estaciones, algunas de ellas proporcionaban información en diferentes escalas o diferentes unidades de medida; de esta forma, los datos

almacenados de las estaciones estarían dentro de la misma escala.

Se decidió añadir nuevas métricas meteorológicas que ya eran desplegadas por los sistemas web, pero no eran almacenadas en el modelo de datos original. El número de estaciones en las cuales se realizaba el proceso de *web scraping* incrementó a 34; algunas fueron estaciones nuevas y otras no estaban registradas en el sistema anterior.

Algoritmo

El proceso que emplea el programa es el siguiente. Primero, obtiene la dirección web (URL) de cada una de las estaciones disponibles en la base de datos, de manera posterior, a través del uso de hilos; según Stallings (2014), esto se refiere a dividir un proceso en múltiples y simultáneos procesos más pequeños, para aumentar la productividad del proceso. Es decir, en lugar de realizar el *web scraping* estación por estación, se realiza para todas las estaciones al mismo tiempo, con lo cual se obtiene toda la información meteorológica que muestra la página web en ese momento; esta información es filtrada, procesada y normalizada a un tipo de dato y unidad de medida en específico. Una vez procesados los datos, estos son insertados en la base de datos en dos tablas diferentes.

Para obtener los elementos deseados, estos son indicados empleando un nombre de clase específico que posee el elemento que deseamos extraer de la página.

La base de datos fue modificada para almacenar los nuevos campos y consta de tres tablas: en una, se almacenan datos de las estaciones, tales como nombre, ubicación en latitud y longitud, datos del responsable de cada estación y dependencia, entre otros. En la segunda tabla, se almacenan los datos históricos recabados por *web scraping* y se

almacenan otros datos como temperatura, humedad, radiación solar, máximos y mínimos, por mencionar algunos. Se almacena hora y fecha de los datos que fueron medidos por el sistema web, así como hora y fecha de consulta por el programa; esto para poder saber cuando una estación falla o cuando deja de actualizar sus datos. La tercera tabla fue creada para que las consultas a la base de datos sean más ágiles, debido al enorme número de datos que se almacenan en la tabla de datos históricos; en esta nueva tabla, se almacenan únicamente los valores del día actual y se borran los del día anterior, de manera que todas las consultas que requieran los valores actuales pueden obtenerse sin que se vea afectado el tiempo de respuesta por el número de registros. Si se desea acceder a datos más antiguos, se utiliza la tabla histórica, con la consideración de que puede tardar en mostrar los resultados.

Resultados

Este programa permite tener una base de datos que registre las condiciones climatológicas del estado de Querétaro en tiempo real; de esta manera, los datos pueden ser utilizados para investigación o divulgación entre los ciudadanos.

El desarrollo iterativo permitió que el programa fuera perfeccionando la implementación del método de *web scraping*, a fin de que cada iteración pudiese registrar más información y minimizar los fallos.

El estado actual del programa permite recolectar el 100% de la información mostrada por más de 30 estaciones meteorológicas a cada minuto, siempre y cuando el sistema web desde donde se realice la consulta muestre actualizada la información.

Actualmente, se han registrado alrededor de 25 millones de registros, que equivalen a 10 GB de información. El rendimiento del programa del *scraping*, así como la cantidad de estaciones que

puede consultar de manera simultánea, depende de las características del equipo que lo ejecute.

Algunas de las ventajas que ofrece el programa sobre su versión anterior son: facilidad de mantenimiento en el código, la cantidad de filtros que posee para almacenar los datos de una forma homogénea en cuanto a formato y la exigencia de recursos en el equipo que lo ejecuta.

Discusión y conclusiones

Almacenar datos climatológicos permitirá tener un compilatorio en forma de serie de tiempo, algo muy importante para poder analizar fenómenos como lluvias atemporales, sequías o inundaciones mediante cálculos estadísticos o empleando diversos algoritmos. Asimismo, esta serie de tiempo puede ser empleada para generar pronósticos mediante modelos estadísticos, así como entrenamiento para inteligencias artificiales que permitan generar pronósticos, o realizar análisis de las tormentas sucedidas.

Referencias bibliográficas

- Borrell G (2008). Python como entorno de desarrollo científico. Recuperado de: <http://torroja.dmt.upm.es/media/files/paper.pdf>
- Castro RA (2004). *Estructura básica del proceso unificado de desarrollo de software. Sistemas y telemática*. Recuperado de: http://www.icesi.edu.co/contenido/pdfs/rcastro_estructura-bas-puds.pdf
- Davis Instruments (2018). Davis | WeatherLink - WeatherLink Station Map. Recuperado de <https://www.weatherlink.com/map.php> (consultado el el 29 de agosto de 2018)
- Glez-Peña D, Lourenço A, López-Fernández H, Reboiro-Jato M y Fdez-Riverola F (2013). Web scraping technologies in an API world. *Briefings in Bioinformatics*. Recuperado de: <https://doi.org/10.1093/bib/bbt026>
- INIFAP (2018). Laboratorio Nacional de Modelaje y Sensores Remotos. Recuperado de: <http://clima.inifap.gob.mx/lnmysr> (consultado el 7 de septiembre de 2018)
- Mitchell R (2015). *Web scraping with Python. The effects of brief mindfulness intervention on acute pain experience: an examination of individual difference*. Recuperado de: <https://doi.org/10.1017/CBO9781107415324.004>
- Stallings W (2014). *Operating systems: internals and design principles* (8ª ed.). Pearson Education. Recuperado de: <https://books.google.com.mx/books?id=omA4AwAAQBAJ>
- The Weather Company (2018). Weather Forecast & Reports - Long Range & Local | Weather Underground. Recuperado de: <https://www.wunderground.com> (consultado el 7 de septiembre de 2018)