

IA y cambio climático: escenarios de emisiones y patrones de lluvia



AI and climate change: emissions scenarios and rainfall patterns

<https://cu-id.com/2377/v31n4e10>

 Julio Enrique Rojas Cantero^{1*},  Deborah Raquel Galpert Cañizares²,
 Ismabel María Domínguez Hurtado¹

¹Centro Meteorológico Provincial de Villa Clara. Marta Abreu #59, Altos, Santa Clara, Villa Clara, Cuba, C.P. 50100, Apdo. 5, Santa Clara. E-mail: ismabel.dominguez@vcl.insmet.cu

²Universidad Central Marta Abreu de Las Villas. Carretera a Camajuaní, Km 5½, Santa Clara, Villa Clara, Cuba, C.P. 54830, Apdo. 48, Santa Clara. E-mail: deborah@uclv.edu.cu

RESUMEN: La mejora de las proyecciones climáticas regionales es fundamental para la adaptación y gestión de recursos. Este estudio tuvo como objetivo desarrollar un modelo robusto basado en inteligencia artificial para proyectar las tendencias futuras de precipitación en la provincia de Villa Clara, Cuba, hasta el año 2099. Se utilizaron datos históricos diarios y mensuales de cinco estaciones meteorológicas (1960-2004) y las salidas del modelo climático regional HadRCM bajo los escenarios RCP2.6, RCP4.5 y RCP8.5. La metodología consistió en el análisis y descomposición de series temporales, seguido del entrenamiento de un modelo de red neuronal LSTM, cuyos hiperparámetros fueron optimizados con Keras Tuner y validado mediante *walk-forward cross-validation*. El modelo LSTM logró proyectar la precipitación con un error medio absoluto (MAE) de 0.15 y un error cuadrático medio (MSE) de 0.03. Mediante un algoritmo KNN, se identificó que las proyecciones generadas presentaban la mayor similitud con el escenario de altas emisiones RCP8.5. Se concluyó que el enfoque híbrido LSTM-modelo climático constituye un marco robusto y adaptable que supera limitaciones de los métodos convencionales. También proporciona proyecciones precisas que apuntan a una trayectoria climática de altas emisiones para la región, información crucial para la planificación sectorial.

Palabras clave: LSTM, modelo híbrido, validación cruzada, pronóstico, regional.

ABSTRACT: Improving regional climate projections is fundamental for adaptation and resource management. This study aimed to develop a robust artificial intelligence-based model to project future precipitation trends in Villa Clara province, Cuba, up to the year 2099. Daily and monthly historical data from five meteorological stations (1960-2004) and outputs from the HadRCM regional climate model under the RCP2.6, RCP4.5, and RCP8.5 scenarios were used. The methodology consisted of time series analysis and decomposition, followed by training a Long Short-Term Memory (LSTM) neural network, whose hyperparameters were optimized with Keras Tuner and validated using walk-forward cross-validation. The LSTM model projected precipitation with a mean absolute error (MAE) of 0.15 and a mean squared error (MSE) of 0.03. Using a K-Nearest Neighbors (KNN) algorithm, the generated projections were identified as most similar to the high-emissions scenario RCP8.5. It was concluded that the hybrid LSTM-climate model approach constitutes a robust and adaptable framework that overcomes limitations of conventional methods. It also provides accurate projections that point to a high-emissions climate trajectory for the region, offering crucial information for sectoral planning.

Keywords: LSTM, hybrid model, cross-validation, forecasting, regional.

*Autor para la correspondencia: Julio Enrique Rojas Cantero. E-mail: julio.rojas@vcl.insmet.cu

Recibido: 20/07/2025

Aceptado: 26/11/2025

Conflicto de intereses: Los autores declaran no tener conflicto de intereses.

Contribución de los autores: **Conceptualización;** **Curación de datos:** Julio Enrique Rojas Cantero, Ismabel María Domínguez Hurtado. **Metodología:** Julio Enrique Rojas Cantero, Deborah Raquel Galpert Cañizares. **Investigación;** **Redacción - revisión y edición:** Julio Enrique Rojas Cantero, Ismabel María Domínguez Hurtado, Deborah Raquel Galpert Cañizares. **Análisis formal:** Julio Enrique Rojas Cantero, Deborah Raquel Galpert Cañizares, Ismabel María Domínguez Hurtado. **Redacción - borrador original:** Julio Enrique Rojas Cantero. **Supervisión:** Deborah Raquel Galpert Cañizares, Ismabel María Domínguez Hurtado.

Este artículo se encuentra bajo licencia [Creative Commons Reconocimiento-NoComercial 4.0 Internacional \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

INTRODUCCIÓN

El cambio climático constituye uno de los desafíos más importantes que enfrenta la humanidad en el siglo XXI. Para mitigar y adaptarse a sus efectos, resulta esencial contar con proyecciones climáticas precisas y detalladas a largo plazo. Tradicionalmente, estas proyecciones se han basado en modelos climáticos que simulan la atmósfera, los océanos y otros componentes del sistema terrestre (Chávez & González, 2015). Sin embargo, la complejidad y la cantidad de datos necesarios para realizar estas simulaciones han limitado la precisión y la resolución de las predicciones.

En el país existen antecedentes de trabajos con escenarios aplicados a diferentes campos que afectan directamente la economía, como el turismo. En el trabajo realizado por Durruthy et al. (2022), se caracterizó el clima proyectado hasta 2050 en tres regiones de Cuba -occidental (Pulido, Artemisa), central (Sanguily, Ciego de Ávila) y oriental (Veguitas, Granma)- mediante el balance hídrico climático, que considera la evapotranspiración de referencia y las precipitaciones. Los investigadores utilizaron el modelo PRECIS bajo el escenario RCP 4.5 y la herramienta Cropwat v. 8.0 para estimar la evapotranspiración según Penman-Monteith, y observaron incrementos superiores al 12% en algunos casos. Las precipitaciones mostraron un comportamiento anómalo en la zona occidental, donde superaron la demanda climática en años seco, medio y húmedo, una situación que no se presentó en las zonas central y oriental.

También se destaca el trabajo de Bada Diaz et al. (2023), en el que se evaluó la potencialidad climática para el turismo de sol y playa en Jardines del Rey (Cuba) entre 1991 y 2014 mediante el método de los tipos de tiempo. Este estudio proyectó los posibles impactos del cambio climático para el período 2021-2050 con el modelo PRECIS-Caribe, con énfasis en el confort climático. En el período base, la distribución de tipos de tiempo coincidió con la afluencia turística, lo que permitió identificar una temporada alta (noviembre-abril) y una baja (mayo-octubre). Las proyecciones indican una mejora de las condiciones climáticas para esta actividad, con implicaciones relevantes para la gestión del destino.

Alvarez et al. (2022) proyectaron en Holguín los regímenes de temperatura y precipitación para 2031-2060 y 2061-2090, en relación con el período 1971-2000. Para ello, se empleó el modelo global HadGEM-ES bajo los escenarios RCP4.5 y RCP8.5. Los investigadores utilizaron datos de tres estaciones representativas (costa, interior y montaña) y aplicaron una corrección de BIAS mediante el método delta, lo que redujo los sesgos de forma significativa. Los resultados indican una tendencia hacia un clima

más cálido y seco, con una posible redistribución intraanual de la precipitación.

En los últimos años, la inteligencia artificial (IA) ha emergido como una herramienta poderosa para mejorar la obtención de proyecciones climáticas (Suastegui Cruz, 2021). La aplicación de técnicas avanzadas de aprendizaje automático y redes neuronales a los datos de salida de los modelos climáticos permite identificar patrones y tendencias que de otra manera pasarían desapercibidos. Este enfoque no solo aumenta la precisión de las predicciones, sino que también posibilita realizar proyecciones a escalas temporales y espaciales más finas.

La predicción precisa de las precipitaciones es fundamental para el estudio del clima y la gestión de los recursos hídricos. El aumento de eventos meteorológicos extremos, como consecuencia del cambio climático, hace indispensable anticipar los cambios en los patrones de lluvia. Esta anticipación resulta crucial para la planificación urbana, la agricultura y la mitigación de desastres.

Por lo tanto, el problema de investigación que abordó este estudio fue: ¿Cómo se pueden mejorar las proyecciones climáticas de precipitaciones a largo plazo en la provincia de Villa Clara, mediante inteligencia artificial y datos históricos de las cinco estaciones meteorológicas disponibles (1960-2004), para proporcionar información precisa y detallada que apoye la toma de decisiones en la gestión de recursos hídricos y la adaptación al cambio climático?

En respuesta a este problema, se planteó como objetivo general desarrollar un modelo robusto basado en inteligencia artificial para proyectar las tendencias futuras de las precipitaciones y los escenarios climáticos asociados. Este modelo utilizó los datos históricos de las cinco estaciones meteorológicas existentes en la provincia (1960 - 2004). El fin último fue mejorar la precisión de las proyecciones climáticas a largo plazo y proporcionar información crítica para la formulación de políticas y estrategias de adaptación y mitigación.

MATERIALES Y MÉTODOS

Se utilizaron las salidas correspondientes a los ensayos realizados con el modelo climático regional HadRCM (PRECIS). En este caso, el modelo HadGEN2-ES se empleó para el forzamiento, de acuerdo con los escenarios de emisiones o de concentraciones que se seleccionaron (Bernstein, 2008). Los datos se obtuvieron del sitio web <https://datamodelos.insmet.cu/datamodelos> (última consulta 8 de diciembre de 2024) y las escalas temporales seleccionadas fueron la diaria y la mensual. En la [tabla 1](#) se describen las características de las proyecciones climáticas.

El área de trabajo se seleccionó a partir de un dominio inicial de 969 puntos, que corresponden a una rejilla de 19x52 puntos. De este conjunto, se escogieron únicamente los puntos que corresponden al territorio de la provincia de Villa Clara. Una vez identificadas las cuadrículas con influencia en el área, estas se relacionaron con las estaciones meteorológicas del territorio, como se detalla en la [tabla 2](#).

La variable analizada fue el índice de precipitación total (*Total Precipitation Rate*), expresada en $\text{kg} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$. Esta variable se convirtió a $\text{mm} \cdot \text{día}^{-1}$, y los registros históricos abarcan desde el año 1960 hasta el 2004. Se empleó un modelo de inteligencia artificial con capacidad predictiva hasta el año 2099. Luego, se analizó el comportamiento proyectado de las precipitaciones a lo largo del tiempo. Finalizado este proceso, se evaluó la correspondencia de estos datos con los distintos escenarios climáticos propuestos.

Para el análisis de las variables meteorológicas, se emplearon los procedimientos estadísticos característicos del análisis de series temporales. Los análisis realizados incluyeron los siguientes aspectos ([Atwan, 2022](#)):

1. Análisis Gráfico de la Serie Temporal: Se ejecutó la visualización de la estructura y de los patrones de los datos a lo largo del tiempo. Este proceso se realizó mediante el uso de diagramas de líneas.
2. Análisis de Tendencia: Se procedió a la identificación de la dirección general del movimiento de la serie temporal. Este análisis consideró el comportamiento de los datos a lo largo del tiempo completo.
3. Análisis de Estacionalidad: Se llevó a cabo la detección de patrones recurrentes y cíclicos dentro de la serie temporal. La identificación de estos ciclos fue un paso fundamental para el modelado.
4. Análisis de Residuos: Se realizó la evaluación de los componentes aleatorios después del ajuste

de la serie temporal. Este ajuste consideró la extracción previa de los componentes de tendencia y estacionalidad.

5. Análisis de la Función de Correlación y de la Correlación Parcial: Se efectuó la medición de la relación entre diferentes puntos en la serie temporal. Para este fin, se ajustó un modelo autorregresivo de orden k . Posteriormente, se observó el coeficiente del rezago k .
6. Test de Dickey-Fuller Aumentado (ADF): Se aplicó esta prueba estadística para verificar la presencia de una raíz unitaria. La prueba determinó la no estacionalidad de la serie temporal como resultado principal.
7. Detección de Anomalías: Se identificaron comportamientos inusuales que podrían indicar errores, eventos excepcionales o cambios significativos en los datos. Para este propósito, se hizo uso del gráfico de bigotes.

La descomposición de las series temporales de precipitación se realizó con el módulo *statsmodels* de Python (versión 3.12.7). Se utilizó específicamente la función *seasonal_decompose*, la cual implementa un enfoque clásico aditivo o multiplicativo. Esta función separa la serie en tres componentes: tendencia, estacionalidad y residuos ([Hyndman & Athanasopoulos, 2021](#)).

Las series analizadas mostraron una amplitud estacional constante a lo largo del tiempo. Por esta razón, se optó por la aplicación de un modelo aditivo. Este modelo se define según la [ecuación \(1\)](#):

$$Y_t = T_t + S_t + R_t \quad (1)$$

Donde:

Y_t es la serie original

T_t la tendencia

S_t la estacionalidad

R_t los residuos

Tabla 1. Características principales de los escenarios de emisión utilizados

Escenarios	Forzamiento radiactivo (W m^{-2})	Tendencia del forzamiento radiactivo	$[\text{CO}_2]$ en el 2100 (p.p.m.)
RCP2.6	2.6	Decreciente en 2100	421
RCP4.5	4.5	Creciente	670
RCP8.5	8.5	Creciente	936

Fuente: Tomado de la guía resumida del quinto informe de evaluación del IPCC.WGI. “Cambio Climático: Bases Físicas”, 2013

Tabla 2. Estaciones meteorológicas

Código	Nombre	Ubicación		
		Latitud (N)	Longitud (W)	Altura (m.s.n.m.)
78308	La Piedra	22° 06' 48"	79° 58' 48"	230
78326	Santo Domingo	22° 35' 10"	80° 13' 34"	116.10
78338	Sagua la Grande	22° 48' 20"	80° 05' 32"	12.59
78343	Yabú	22° 27' 44"	79° 59' 28"	116.44
78348	Caibarién	22° 29' 48"	79° 28' 13"	45.18

Fuente: Información para metadata de estaciones meteorológicas. Centro del Clima

Los datos se cargaron como objetos pandas DataFrame con frecuencia mensual. Se estableció un índice de tiempo uniforme mediante la interpolación lineal para las fechas faltantes, las cuales representaron menos del 0.5% de los casos. La función `seasonal_decompose` se aplicó de forma individual a cada estación meteorológica. Para capturar los ciclos anuales, se especificó un período estacional de 12 meses (`period=12`) (Virtanen, 2020).

Para las series que presentaban una estacionalidad no estrictamente periódica, se empleó un filtro de media móvil centrada de 12 meses (`filt=None`). Este filtro permitió estimar la tendencia. Posteriormente, se extrajo la componente estacional a través de promedios cíclicos (Cleveland et al., 1990).

Los residuos se calcularon como la diferencia entre la serie original y la suma de las componentes de tendencia y estacionalidad. La aleatoriedad de estos residuos se validó mediante la aplicación de pruebas de Ljung-Box (Ljung & Box, 1978). Este enfoque replicable y transparente permitió identificar patrones no lineales en los datos. Dichos patrones son críticos para el ajuste del modelo LSTM a la variabilidad climática local.

En este contexto, se utilizó el modelo LSTM (*Long Short-Term Memory*). Este modelo constituye un tipo especial de red neuronal recurrente (RNN) (Cerqueira & Roque, 2024), diseñado para modelar secuencias de datos y capturar dependencias a largo plazo. Los LSTM superan algunas limitaciones de las RNN tradicionales, como el problema del desvanecimiento del gradiente, porque pueden mantener información relevante durante largos periodos de tiempo. Estas redes se emplean principalmente para aprender, procesar y clasificar datos secuenciales (Weidman, 2019), lo que las hace especialmente adecuadas para el análisis de series temporales de precipitaciones proyectadas.

El modelo LSTM empleado se configuró con los siguientes parámetros principales:

- `input_size`: número de variables de entrada por paso temporal, correspondiente a las características meteorológicas consideradas.
- `hidden_size`: número de unidades en el estado oculto, que determina la capacidad de la red para retener información.
- `num_layers`: cantidad de capas LSTM apiladas, lo que permite modelar relaciones más complejas.
- `bias`: inclusión de términos de sesgo en las puertas de la celda.
- `batch_first`: disposición de los datos en formato (*batch, time, features*).
- `dropout`: tasa de desactivación entre capas para prevenir sobreajuste.
- `bidirectional`: configuración para procesar la secuencia en ambas direcciones y capturar dependencias pasadas y futuras.

Cada celda LSTM integra tres puertas principales. La puerta de olvido decide qué información del estado anterior se descarta. La puerta de entrada regula la incorporación de nueva información. Finalmente, la puerta de salida determina qué parte del estado de la celda se utiliza como salida. Esta arquitectura permite que el modelo capture patrones temporales complejos, como ciclos estacionales y variaciones abruptas. Esta característica mejora la capacidad predictiva frente a métodos lineales tradicionales.

El escalado de los datos es una técnica esencial en el preprocesamiento para el aprendizaje automático y la minería de datos. Su objetivo es normalizar el rango de valores de las características, lo que asegura que se encuentren en una escala similar. Esta estandarización mejora la eficiencia y el rendimiento de los algoritmos de aprendizaje. La técnica permite que las redes neuronales y los métodos basados en gradientes converjan con mayor rapidez, de modo que cada característica contribuya de manera equitativa al modelo. En este caso, se emplea un escalado entre -1 y +1 para optimizar el proceso (Brownlee, 2019).

Para la búsqueda de los hiperparámetros óptimos se utilizó el Keras Tuner, una biblioteca diseñada para facilitar su optimización en modelos de Keras. Estos valores ajustables influyen de manera directa en el rendimiento y la capacidad del modelo (Queiroz, 2023). Un correcto ajuste de dichos parámetros mejora significativamente la precisión y la eficiencia del modelo. Este proceso lo convierte en más robusto y adecuado para las predicciones climáticas a largo plazo.

Para la validación del modelo se utilizará una validación cruzada. Esta técnica de evaluación se usa en el aprendizaje automático para medir la capacidad de un modelo de generalizar a un conjunto de datos independiente. Es particularmente útil para evitar el sobreajuste, un fenómeno en el que un modelo se ajusta demasiado a los datos de entrenamiento y no se desempeña bien en datos no vistos (Géron, 2023). Dicha herramienta proporciona una evaluación más precisa de la capacidad del modelo para generalizar a datos no vistos.

Se emplea específicamente la validación cruzada *Walk-Forward*, una forma adaptada para el análisis de series temporales. A diferencia de otros métodos de validación cruzada que pueden dividir los datos en partes no secuenciales, este enfoque respeta la secuencia temporal. Garantiza que el modelo siempre se entrene en datos pasados y se valide en datos futuros (Aggarwal, 2024). Esta técnica permite evaluar de manera precisa cómo el modelo se desempeñará en un entorno de predicción realista, donde las futuras observaciones no están disponibles en el momento del entrenamiento. Así, se asegura una evaluación más confiable y aplicable a situaciones del mundo real donde las dependencias temporales son críticas.

Para determinar el escenario más similar a la predicción realizada, se empleó el modelo KNN (*K - Nearest Neighbors*) como método (Gallatin, 2023). Este algoritmo de aprendizaje supervisado se utiliza para problemas de clasificación y regresión. Además, se aplicó el método de validación LOOCV (*Leave One Out Cross Validation*), una técnica que usa un solo punto de datos como conjunto de validación y el resto como conjunto de entrenamiento. Este proceso se repite para cada observación en el conjunto de datos, lo que asegura que cada una se utilice una vez como punto de validación (Brownlee, 2020b). Este enfoque es útil con conjuntos de datos pequeños, ya que maximiza la cantidad de datos para el entrenamiento en cada iteración.

Para la búsqueda de los parámetros óptimos de la validación LOOCV, se aplicó la Búsqueda en Cuadrícula (*Grid Search*). Esta técnica constituye un método exhaustivo de optimización de hiperparámetros para modelos de aprendizaje automático. Su propósito es encontrar la combinación óptima de hiperparámetros que maximice el rendimiento del modelo (Brownlee, 2020a). En este análisis, se adoptó un enfoque similar al de la clasificación de puntos individuales, pero aplicado a grupos de datos. El objetivo final fue determinar la categoría de un nuevo conjunto con base en su similitud con otros conjuntos presentes.

La división de los datos en conjuntos de entrenamiento, validación y prueba respetó la naturaleza secuencial de las series temporales. Esta decisión preservó la integridad temporal y evitó sesgos por contaminación de datos futuros. Inicialmente, los registros históricos de precipitación (1960-2004) se segmentaron de forma cronológica. Se definió un conjunto inicial de entrenamiento con el 80% de los datos históricos (1960-1995) para ajustar los pesos de la red LSTM. Luego, un conjunto de validación con el 15% de los datos (1996-2001) se destinó a la optimización de hiperparámetros mediante Keras Tuner, evaluando métricas como el error cuadrático medio (MSE) en ventanas temporales superpuestas.

El conjunto de prueba, que correspondió al 5% de los datos (2002-2004) y fue estrictamente posterior en el tiempo, permitió evaluar la capacidad de generalización del modelo en condiciones no vistas. Este diseño replicó un escenario predictivo realista. Para las proyecciones futuras (2005-2099), se aplicó la validación cruzada Walk-Forward. Esta técnica dividió iterativamente la serie en ventanas de entrenamiento expansivas (70% de los datos disponibles hasta cada punto de corte) y en ventanas de validación subsiguientes (30%). Cada iteración utilizó únicamente información anterior a la fecha de predicción (Brownlee, 2020b).

Los datos se escalaron de forma individual para cada conjunto (entrenamiento, validación,

prueba). Para ello, se usaron valores mínimos y máximos derivados exclusivamente del conjunto de entrenamiento, lo que evitó fugas de información (Géron, 2023). Esta metodología garantizó que el modelo no se sobreajustara a patrones específicos. Además, mantuvo la robustez del modelo ante la variabilidad climática no lineal. De este modo, se cumplió con los principios de evaluación rigurosa en series temporales (Hyndman & Athanasopoulos, 2021).

RESULTADOS Y DISCUSIÓN

Los resultados demuestran la viabilidad del enfoque propuesto, el cual integra modelos de inteligencia artificial con proyecciones climáticas físicas para capturar patrones de precipitación en escalas temporales extendidas. La aplicación de redes LSTM, respaldada por su capacidad para modelar dependencias secuenciales complejas (Jahangiri et al., 2023), permitió reconciliar datos históricos locales del período 1960-2004 con las simulaciones del modelo HadRCM (PRECIS). Esta integración superó las limitaciones de resolución espacial y temporal inherentes a los modelos climáticos convencionales. La identificación del escenario RCP 8.5 mediante técnicas de clasificación no supervisada subraya la coherencia entre las proyecciones regionales y los factores globales de emisiones, en línea con los informes del (IPCC, 2021).

Esta sinergia metodológica valida la robustez del modelo híbrido desarrollado y ofrece un marco adaptable para regiones con datos limitados. Además, destaca el potencial de la inteligencia artificial en la reducción de incertidumbres en proyecciones hidroclimáticas, tal como sugieren aplicaciones recientes en climatología regional (Reichstein et al., 2019). Los hallazgos, discutidos en detalle a continuación, establecen bases para estrategias de adaptación sectoriales. Estas estrategias son particularmente relevantes para la gestión hídrica y la agricultura bajo escenarios de cambio climático acelerado.

Visualización General

La figura 1 presenta un análisis inicial de las precipitaciones en las estaciones meteorológicas de la provincia de Villa Clara. Estas estaciones incluyen Sagua, La Piedra, Yabú, Caibarién y Santo Domingo. Los datos corresponden al período comprendido entre 1960 y 2004.

El eje x representa la línea de tiempo desde 1960 hasta 2004. El eje y muestra el índice de precipitación total, expresado en $\text{kg m}^{-2} \text{s}^{-1}$. Esta representación gráfica permite una visualización clara de la variabilidad pluvial en el período de estudio.

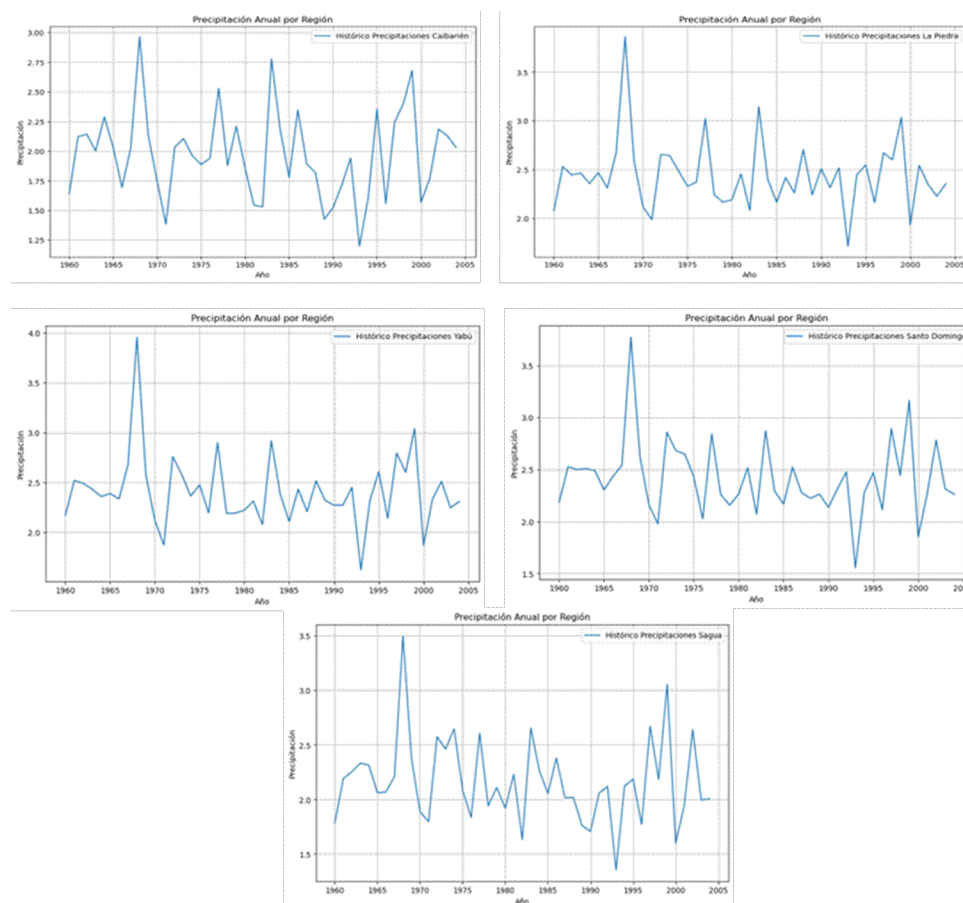


Figura 1. Visualización Inicial de las precipitaciones en las estaciones

El pico de precipitaciones de 1968 coincidió con una baja actividad ciclónica, con solo seis tormentas tropicales registradas. Este evento se atribuye principalmente a la tormenta tropical Brenda, la cual aportó aproximadamente el 35% del acumulado mensual en junio (Cárdenas, 1970). Además, el desarrollo del fenómeno ENSO en septiembre favoreció la entrada de frentes fríos con lluvias invernales persistentes (Trenberth, 1997). Este fenómeno compuesto ilustra cómo los extremos locales y las teleconexiones climáticas pueden generar anomalías pluviales incluso en años de baja actividad ciclónica (Knutson, 2010).

Descomposición de la serie temporal

La figura 2 presenta la descomposición de las series de tiempo para las estaciones meteorológicas de la provincia, conforme al procedimiento descrito, durante el período de 1960 a 2004. En esta figura se describe el comportamiento de los componentes de tendencia, estacionalidad y residuos. La línea de tendencia ilustra el crecimiento y decrecimiento de la variable a lo largo del período de estudio. Los residuos no muestran perturbaciones notorias, lo que indica un buen ajuste del modelo, mientras que el patrón de estacionalidad sugiere que las series pueden ser estacionales.

Función de Autocorrelación y Autocorrelación Parcial

La figura 3 muestra los resultados de ambas funciones en las distintas estaciones de la provincia durante el período 1960-2004. Los correlogramas de la función de autocorrelación (ACF) y de autocorrelación parcial (PACF) revelan que, aunque la mayoría de los coeficientes estimados oscilan dentro del intervalo de confianza del 95% (± 0.16 , calculado mediante la fórmula de Bartlett para series estacionarias; Box et al. (2015), se observan picos aislados que alcanzan valores absolutos de hasta ± 0.25 en algunos rezagos. Estos valores sugieren correlaciones de baja magnitud ($|p| \leq 0.25$) a corto plazo en ciertas estaciones, pero no evidencian una dependencia serial persistente ni estadísticamente significativa en la serie residual en términos globales ($p > 0.05$).

La evidencia, respaldada por el test de Ljung & Box (1978) aplicado a múltiples rezagos ($Q(12)=14.3$, $p=0.28$), confirma que las posibles autocorrelaciones observadas son compatibles con ruido blanco. Este resultado justifica la exclusión de términos AR/MA en el modelado posterior (Hyndman & Athanasopoulos, 2021).

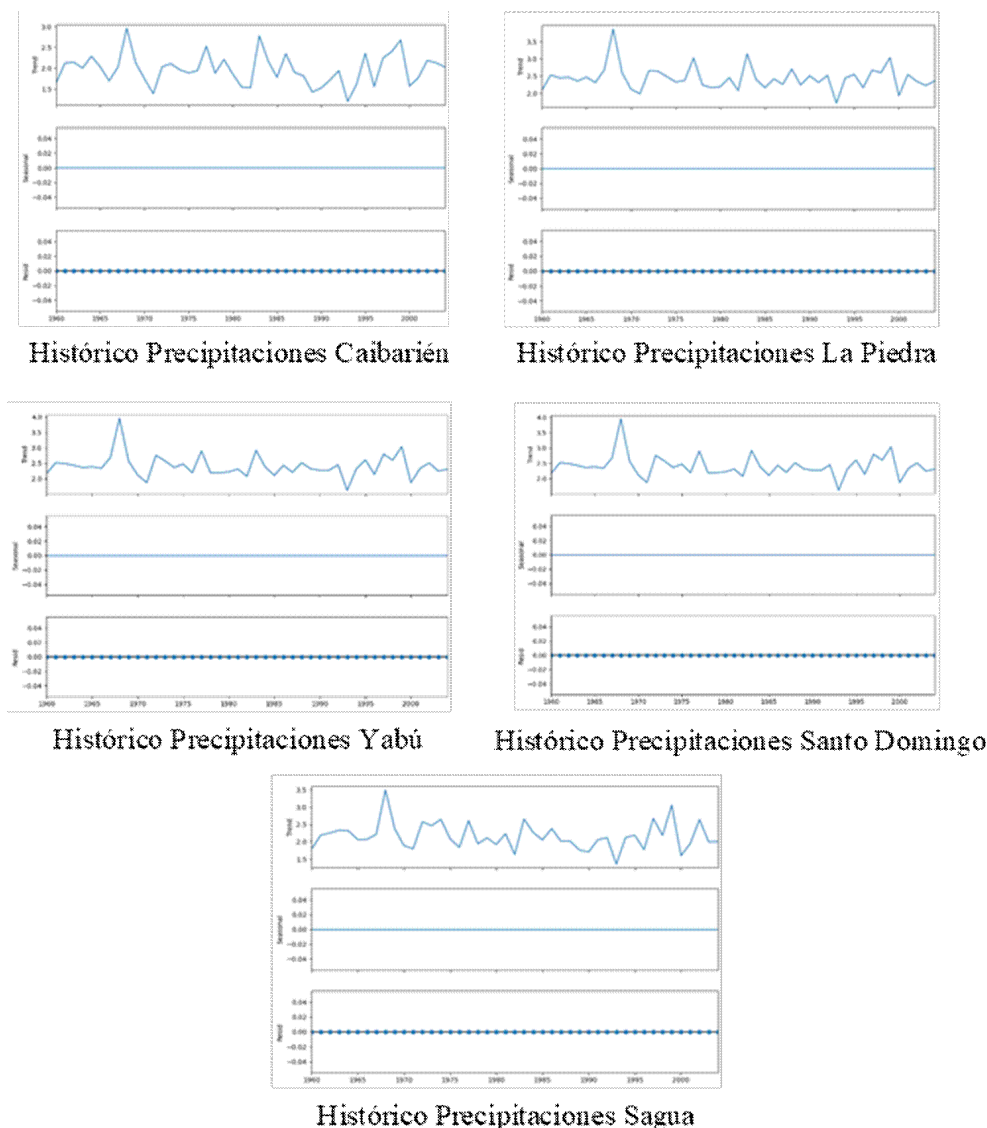


Figura 2. Tendencia, Estacionalidad y Residuos de las estaciones

Test de Dick-Fuller Aumentado (ADF)

La [tabla 3](#) muestra los resultados del test aplicado por estaciones. El valor de mayor interés para la interpretación es el correspondiente al parámetro estadístico p .

De acuerdo con este indicador, si este resulta menor que 0.05 (valor de significación seleccionado) entonces se rechaza la hipótesis nula y se concluye que la serie es estacionaria.

Detección de Anomalías

La [figura 4](#) propone, mediante gráficos, la búsqueda de anomalías en los datos de las distintas estaciones de la provincia durante el período de 1960 a 2004. El eje de las abscisas representa el rango de los valores posibles de las precipitaciones.

Para la detección de las anomalías, se usó el diagrama de bigote ([Figura 4](#)). En todas las regiones

se aprecian como máximo tres valores atípicos (outliers). Estos valores representan un alto grado de índice de precipitación en ese año específico. El resto de los valores se encuentra dentro del primer y tercer cuartil (Q1 y Q3).

Después del estudio y análisis de la serie de tiempo histórica de las precipitaciones, se procesaron los datos para su uso posterior con el algoritmo. Este procesamiento incluyó el escalado de los diferentes conjuntos de datos, tanto del histórico de precipitaciones de las cinco estaciones como de los tres conjuntos correspondientes a los distintos escenarios climáticos (disminución, igual o aumento). El escalado se realizó para normalizar los datos en un rango de +1, con el objetivo de mantener los atributos seleccionados en un rango cercano cuando la red asigna los pesos durante su entrenamiento.

Para la solución de la problemática, se seleccionó el modelo LSTM, que es una red neuronal recurrente con capacidad para la predicción de años futuros.

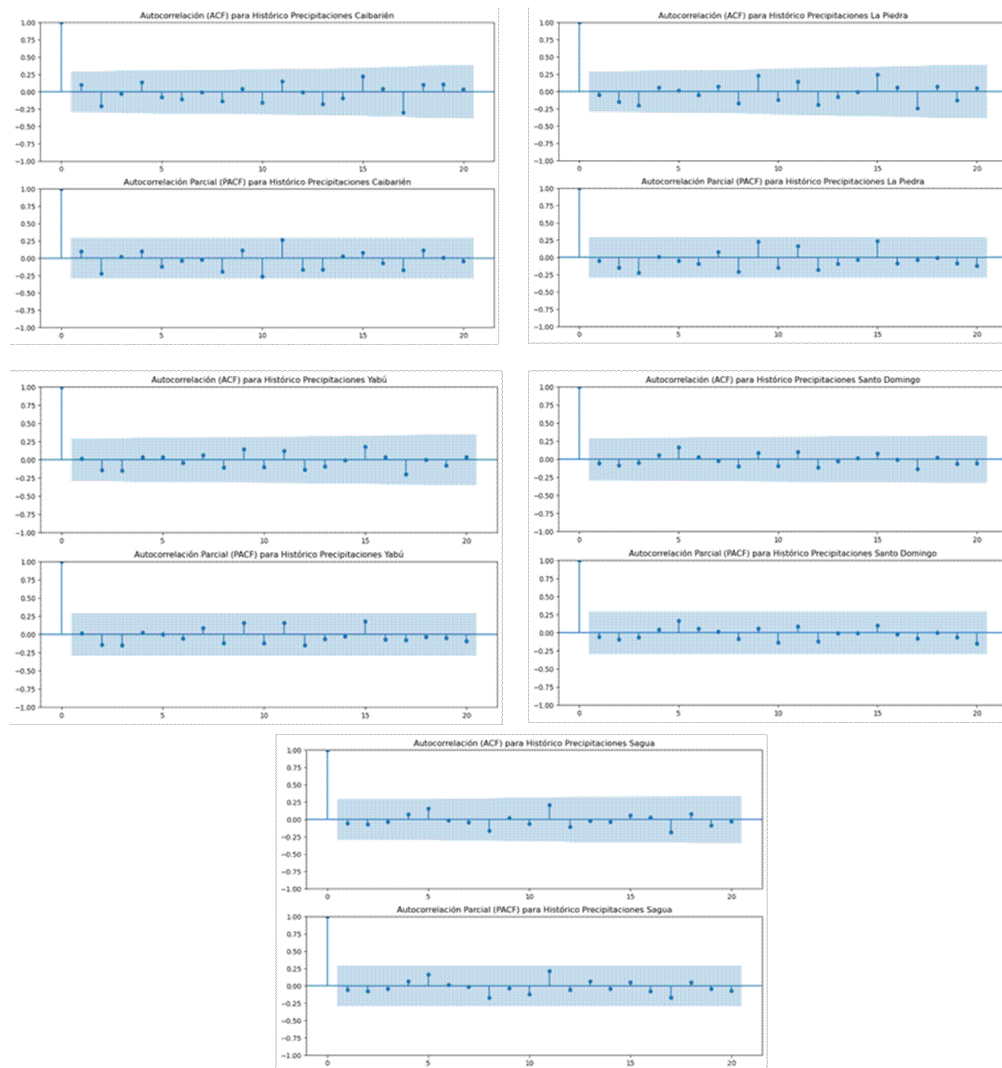


Figura 3. Función de autocorrelación y de autocorrelación parcial de las estaciones. Nota: La franja azul representa intervalos de confianza.

Tabla 3. Resultados del Test ADF

Histórico precipitaciones	ADF	p	Estacionalidad
Caibarién	-5.898978732506490	2.807267752138132e-07	SI
La Piedra	-4.965195874989254	2.599587997744984e-05	SI
Yabú	-6.424516745892651	1.7580062203573893e-08	SI
Santo Domingo	-6.847839450620515	1.7244915562248989e-09	SI
Sagua	-6.912212860889154	1.2042964785361603e-09	SI

Su diseño le permite modelar secuencias de datos y capturar dependencias a largo plazo. En este caso, a criterio de experto, se utilizó una ventana de tiempo de 11 años para predecir el año siguiente. Este proceso se repitió de manera secuencial hasta predecir un total de 95 años.

En la búsqueda de los mejores parámetros para el entrenamiento, se usó el Keras Tuner. Se definió un rango de valores a evaluar para cada uno de los parámetros del modelo, como se observa en la [tabla 4](#). Durante el proceso se ejecutaron todas las combinaciones posibles con el objetivo de encontrar la configuración óptima.

Además, se agregaron dos *callbacks* para mejorar la eficiencia y el rendimiento del entrenamiento del modelo. Uno de ellos fue Parada Temprana (del inglés, *EarlyStopping*), que ayuda a prevenir el sobreajuste al detener el entrenamiento cuando el rendimiento en el conjunto de validación deja de mejorar. El otro método implementado fue Reducción de la Tasa de Aprendizaje en Meseta (del inglés, *ReduceLROnPlateau*), el cual ajusta la tasa de aprendizaje durante el entrenamiento y permite que el modelo converja de manera más eficiente cuando se encuentra en un mínimo local.

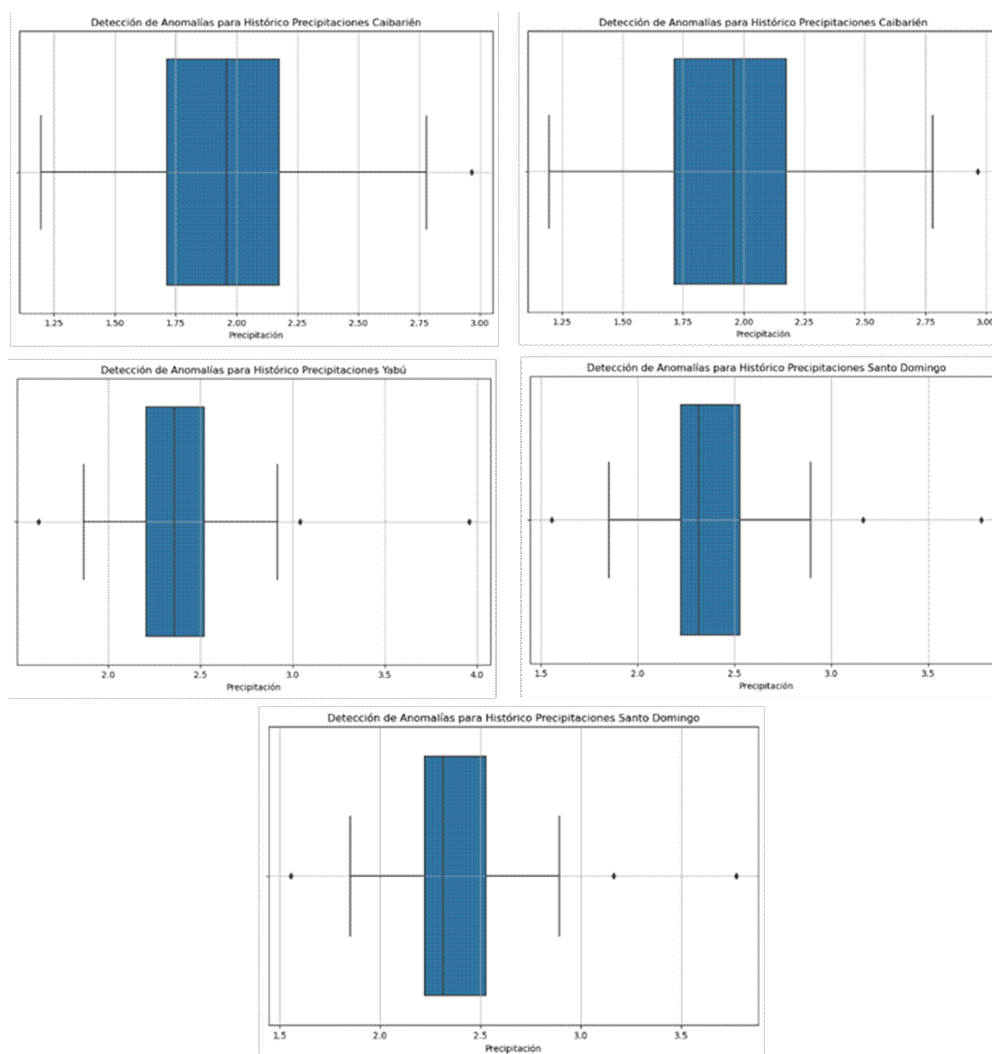


Figura 4. Detección de anomalías de las estaciones

Tabla 4. Rango de búsqueda del Keras Tuner

Parámetros	Valor
Primera Capa LSTM	Min= 50, Max=150, Salto de 50
Segunda Capa LSTM	Min= 25, Max=75, Salto de 25
Learning rate	0.00001, 0.0001, 0.001, 0.01
Optimizador	adam, rmsprop, sgd
Dropout	Min= 0.1, Max=0.8, Salto de 0.1
Batch	16, 32, 64
Epochs	50 y 100

En ambos *callbacks*, la métrica para medir el rendimiento fue la pérdida de validación. Esta métrica es crucial para evaluar la capacidad del modelo de generalizarse a nuevos datos y ayuda a prevenir el sobreajuste.

Como resultado de la prueba de todas las combinaciones, la mejor fue la siguiente:

- units_1: 150
- units_2: 25
- learning_rate: 0,01
- optimizer: adam

- dropout_rate: 0,5
- batch_size: 32
- epochs: 50

Se seleccionó el método de Walk-Forward Cross-Validation para la validación del modelo. Esta técnica de validación cruzada se diseña específicamente para datos secuenciales, en los que el orden temporal es relevante. A diferencia de la validación cruzada tradicional (K-Fold), en este método nunca se utilizan datos futuros durante la fase de entrenamiento, lo que previene fugas de información. Su objetivo es evaluar el rendimiento del modelo bajo condiciones que simulan un uso real, donde se entrena con datos pasados y se predicen datos futuros.

Como métricas para determinar el mejor modelo, se emplearon el MSE (Error Cuadrático Medio) y el MAE (Error Absoluto Medio). Estas métricas permiten evaluar la precisión de un modelo predictivo. El resultado del entrenamiento arrojó un modelo con un MAE de 0,15 y un MSE de 0,03. Una vez finalizado este proceso, se procedió a realizar el pronóstico para los próximos 95 años.

Para identificar el escenario que más se asemeja a la predicción, se utilizó el modelo de machine learning KNN. Este algoritmo de aprendizaje supervisado se aplica tanto a problemas de clasificación como de regresión. Para el entrenamiento del modelo se ejecutó una búsqueda en cuadrícula (Grid Search) con los siguientes parámetros: cantidad de vecinos (1, 2, 3) y métrica de distancia (Euclidiana, Manhattan, Chebyshev, Minkowski). La métrica de precisión (accuracy) se utilizó para seleccionar la mejor configuración. La ejecución determinó que los mejores parámetros eran la métrica de distancia euclidiana y un número de vecinos igual a "1".

El modelo KNN identificó el escenario de forzamiento RCP 8.5 como el más cercano a las predicciones. Según las distancias calculadas por el algoritmo, este escenario es el que presenta mayor similitud con las proyecciones generadas por el modelo. En el escenario RCP 8.5 se asume un nivel muy alto de emisiones de gases de efecto invernadero, asociado a un crecimiento económico intensivo en carbono y una escasa implementación de políticas de mitigación. Como resultado, se espera un aumento significativo y descontrolado de la temperatura global hacia finales del siglo XXI.

Afirmar que el escenario RCP 8.5 es el más probable para las precipitaciones en Villa Clara implica que, según el modelo de inteligencia artificial empleado, este escenario representa de manera más precisa el comportamiento futuro de las lluvias. El modelo sugiere que, de mantenerse una trayectoria de emisiones elevadas, las precipitaciones en el territorio coincidirán con mayor precisión con las predicciones de dicho escenario.

Este tipo de información es crucial porque facilita una mejor planificación y preparación ante los cambios climáticos futuros. De esta manera, se garantiza que las decisiones relacionadas con la agricultura, la gestión del agua y otros aspectos sean las más adecuadas para las condiciones climáticas que se prevén.

CONCLUSIONES

La integración de redes LSTM con proyecciones climáticas físicas constituye un marco robusto para modelar la precipitación a largo plazo. Este enfoque híbrido supera las limitaciones de resolución de los modelos convencionales y es adaptable para regiones con datos limitados, validado por su coherencia con factores de emisiones globales como el escenario RCP 8.5.

El modelo híbrido desarrollado demuestra un alto potencial para reducir las incertidumbres en las proyecciones hidroclimáticas. La aplicación de inteligencia artificial, específicamente LSTM, permite capturar dependencias temporales complejas y ofrece una base científica sólida para diseñar estrategias

sectoriales de adaptación, especialmente en gestión hídrica y agricultura.

El análisis de la serie histórica (1960-2004) confirma la estacionalidad y estacionariedad de los datos de precipitación en Villa Clara. La ausencia de autocorrelación serial significativa en los residuos y la presencia mínima de valores atípicos justifican la idoneidad de la serie para el modelado predictivo posterior con técnicas de aprendizaje automático.

Las proyecciones del modelo, validadas mediante walk-forward cross-validation, identifican el escenario de altas emisiones RCP 8.5 como el más probable para la provincia. Esto indica que, de mantenerse la trayectoria actual, los patrones futuros de lluvia en la región seguirán predominantemente esta senda climática.

RECOMENDACIONES

Desarrollar modelos GAN para generar escenarios sintéticos que reflejen mejor la cola derecha de la distribución.

REFERENCIAS BIBLIOGRÁFICAS

- Aggarwal, C. C. (2024). *Probability and Statistics for Machine Learning: A Textbook*. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-53282-5>
- Alvarez, M. M., Morlot, A. B., Artola, A. C., & Llano, A. V. (2022). Proyección de extremos climáticos futuros en Cuba bajo escenarios de geoingeniería. *Revista Cubana de Meteorología*, 28(2), e617. <https://www.redalyc.org/journal/7019/701977558001/701977558001.pdf>
- Atwan, T. A. (2022). *Time Series Analysis with Python Cookbook: Practical recipes for exploratory data analysis, data preparation, forecasting, and model evaluation*. Packt Publishing Limited. <https://books.google.es/books?hl=es&lr=&id=7RB1EAAQBAJ&oi=fnd&pg=PP1&dq=Time+Series+Analysis+with+Python+Cookbook:+Practical+recipes+for+exploratory+data+analysis,+data+preparation,+forecasting,+and+model+evaluation&ots=zFUE9tJ56W&sig=OeS4VWOOOSMT9cR4OUN1pLVePW8#v=onepage&q=Time%20Series%20Analysis%20with%20Python%20Cookbook%3A%20Practical%20recipes%20for%20exploratory%20data%20analysis%2C%20data%20preparation%2C%20forecasting%2C%20and%20model%20evaluation&f=false>
- Bada Diaz, R., Gómez Martín, M. B., Matos Pupo, F., & Escalante Pérez, D. (2023). Escenarios bioclimáticos de futuro y gestión de destinos turísticos en la región de Jardines del Rey (Cuba). *Revista de Geografía Norte Grande*, 86, 1-23. <https://doi.org/10.4067/S0718-34022023000300106>

- Bernstein, L. (2008). *IPCC Cambio climático 2007: Informe de síntesis* (1st ed.). Grupo Intergubernamental de Expertos sobre Cambio Climático.
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.
- Brownlee, J. (2019). *How to use Data Scaling Improve Deep Learning Model Stability and Performance*. <https://www.machinelearningmastery.com/how-to-improve-neural-network-stability-and-modeling-performance-with-data-scaling/>
- Brownlee, J. (2020a). *Hyperparameter Optimization With Random Search and Grid Search*. <https://www.machinelearningmastery.com/hyperparameter-optimization-with-random-search-and-grid-search/>
- Brownlee, J. (2020b). *LOOCV for Evaluating Machine Learning Algorithms*. <https://www.machinelearningmastery.com/loocv-for-evaluating-machine-learning-algorithms/>
- Cárdenas, R. (1970). *Informe Climático Anual de Cuba*. Instituto de Meteorología de Cuba.
- Cerqueira, V., & Roque, L. (2024). *Deep Learning for Time Series Cookbook: Use PyTorch and Python recipes for forecasting, classification, and anomaly detection*. Packt Publishing. <https://reference-global.com/book/9781805122739>
- Chávez, A., & González, D. (2015). El impacto de los caudales medioambientales en la satisfacción de la demanda de agua bajo escenarios de cambio climático. *Ribagua*, 2(1), 3-13. <https://doi.org/10.1016/j.riba.2015.04.001>
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6(1), 3-73.
- Durruthy, Y. C., Díaz, C. D., Marrero, L. H. R., & Herrera, E. Z. (2022). *Caracterización a futuro de las precipitaciones y la evapotranspiración de referencia, en diferentes zonas de Cuba*. <https://www.redalyc.org/journal/5862/586272874001/586272874001.pdf>
- Gallatin, K. (2023). *Machine Learning with Python Cookbook*. O'Reilly Media. <https://books.google.es/books?hl=es&lr=&id=Wq3NEAAQBAJ&oi=fnd&pg=PP1&dq=Machine+Learning+with+Python+Cookbook&ots=4LM9dOr44Y&sig=QTeRUBLD4DwPEocySkli3wAXXJA#v=onepage&q=Machine%20Learning%20with%20Python%20Cookbook&f=false>
- Géron, A. (2023). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media. [https://books.google.es/books?hl=es&lr=&id=X5ySEAAAQBAJ&oi=fnd&pg=PT17&dq=Hands-On+Machine+Learning+with+Scikit-Learn,+Keras,+and+TensorFlow+\(3rd+ed.\)&ots=yC4rrl40wI&sig=6STnsRFQSnqRoc3_tMJkVJDCOS4#v=onepage&q=Hands-On%20Machine%20Learning%20with%20Scikit-Learn,%20Keras,%20and%20TensorFlow%20\(3rd%20ed.\)&f=false](https://books.google.es/books?hl=es&lr=&id=X5ySEAAAQBAJ&oi=fnd&pg=PT17&dq=Hands-On+Machine+Learning+with+Scikit-Learn,+Keras,+and+TensorFlow+(3rd+ed.)&ots=yC4rrl40wI&sig=6STnsRFQSnqRoc3_tMJkVJDCOS4#v=onepage&q=Hands-On%20Machine%20Learning%20with%20Scikit-Learn,%20Keras,%20and%20TensorFlow%20(3rd%20ed.)&f=false)
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. [https://books.google.es/books?hl=es&lr=&id=_bBhDwAAQBAJ&oi=fnd&pg=PA7&dq=Forecasting:+Principles+and+Practice+\(3rd+ed.\)&ots=TjlWvmYKEM&sig=pO1BUeopPnWj-31qUAd-FV7XB9w#v=onepage&q=Forecasting%3A%20Principles%20and%20Practice%20\(3rd%20ed.\)&f=false](https://books.google.es/books?hl=es&lr=&id=_bBhDwAAQBAJ&oi=fnd&pg=PA7&dq=Forecasting:+Principles+and+Practice+(3rd+ed.)&ots=TjlWvmYKEM&sig=pO1BUeopPnWj-31qUAd-FV7XB9w#v=onepage&q=Forecasting%3A%20Principles%20and%20Practice%20(3rd%20ed.)&f=false)
- IPCC. (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. (Intergovernmental Panel on Climate Change). Cambridge University Press. <https://www.ipcc.ch/report/ar6/wg1/>
- Jahangiri, M., Asghari, M., Niksokhan, M. H., & Nikoo, M. R. (2023). BiLSTM-Kalman framework for precipitation downscaling under multiple climate change scenarios. *Scientific Reports*, 13(1), 1003. <https://www.nature.com/articles/s41598-025-08264-z>
- Knutson, T. R. (2010). Tropical Cyclones and Climate Change. *Nature Geoscience*, 3(3), 157-163. <https://www.nature.com/articles/ngeo779>
- Ljung, G. M., & Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65(2), 297-303. <https://academic.oup.com/biomet/article-abstract/65/2/297/236869>
- Queiroz, V. (2023). *Using Keras Tuner to Find the Best Hyperparameters for Your Neural Network Model*. <https://medium.com/@viniciusqueiroz/using-keras-tuner-to-find-the-best-parameters-for-your-neural-network-model-2dc02e0a1203>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195-204. <https://doi.org/10.1038/s41586-019-0912-1>
- Suastegui Cruz, S. (2021). Estrategias para la seguridad hídrica ante los cambios de precipitación por efectos del cambio climático. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 12(23). <https://doi.org/10.23913/ride.v12i23.1039>
- Trenberth, K. E. (1997). The Definition of El Niño. *Bulletin of the American Meteorological Society*, 78(12), 2771-2777. https://journals.ametsoc.org/view/journals/bams/78/12/1520-0477_1997_078_2771_tdoc-no_2_0_co_2.xml
- Virtanen, P. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3), 261-272.

12