

ARTÍCULO DE INVESTIGACIÓN

Predicción de la humedad relativa en bosques de galería de Santa Clara mediante inteligencia artificial/ *Prediction of relative humidity in Santa Clara gallery forests using artificial intelligence*

Julio Enrique Rojas Cantero,

<https://orcid.org/0009-0001-3389-9660>

Centro Meteorológico Provincial de Villa Clara,

jerojas1998@gmail.com

Cynthia Rodríguez-Alfaro,

<https://orcid.org/0000-0003-2307-8567>

Centro Meteorológico Provincial de Villa Clara,

cralfaro@vcl.insmet.cu

Emmanuel Gómez Pérez,

<https://orcid.org/0000-0003-2307-8567>

Centro Meteorológico Provincial de Villa Clara

emmanuel.gomez@vcl.insmet.cu

Recibido: 26/09/2025

Aceptado: 26/10/2025

Resumen

Los bosques de galería sobre serpentinitas al sureste de Santa Clara se caracterizan por un altísimo endemismo. Estos bosques son el hábitat de *Sachsia coronopifolia*, una planta endémica local en Peligro Crítico, considerada entre las 50 plantas más amenazadas de Cuba. La distribución de esta especie se halla restringida a estos parches de bosques, los cuales se encuentran sometidos a la fragmentación y degradación de la cobertura boscosa. El objetivo de este trabajo es modelar algoritmos computacionales capaces de predecir el comportamiento futuro de la humedad relativa del aire en los bosques de galería al sureste de Santa Clara mediante inteligencia artificial. En este sentido, el uso de series temporales se considera el recurso más adecuado. Los resultados del ARIMA modelado mediante técnicas de inteligencia artificial pueden ser útiles para identificar y realizar a tiempo acciones de reforestación que prevengan los efectos negativos del cambio climático sobre estos bosques.

Palabras clave: ARIMA, Conservación, Especie endémica, Series temporales

Abstract

The gallery forests on serpentinite soils southeast of Santa Clara are characterized by a very high level of endemism. These forests are the habitat of **Sachsia coronopifolia**, a local endemic plant

that is critically endangered and considered among the 50 most threatened plants in Cuba. The distribution of this species is restricted to these forest patches, which are subject to fragmentation and degradation of forest cover. The objective of this work is to model computational algorithms capable of predicting the future behavior of relative air humidity in the gallery forests southeast of Santa Clara using artificial intelligence. In this regard, the use of time series is considered the most appropriate resource. The results of the ARIMA modeling using artificial intelligence techniques can be useful for identifying and carrying out timely reforestation actions that prevent the negative effects of climate change on these forests.

Keywords: ARIMA, Conservation, Endemic species, Time series.

Introducción

Los bosques de galería que atraviesan los matorrales xeromorfos espinosos sobre serpentinitas en la Reserva Florística Manejada “Sabanas de Santa Clara”, Villa Clara, Cuba se caracteriza por un altísimo endemismo ((Ruiz Plasencia et al., 2019). Estos bosques de galería sobre serpentinas se localizan sobre un área expuesta a la fragmentación del hábitat por acciones antrópicas (Montalvo et al.; 2012).

Además, la degradación de la cobertura boscosa provocada por la presencia de especies invasoras e incendios forestales es notable en algunos parches de bosques (Ruiz Plasencia et al., 2019). Estos bosques son el hábitat de *Sachsia coronopifolia* (Griseb.) Anderb. (Asteraceae) una planta endémica local en Peligro Crítico (CR) (UICN, 2012; (González Torres et al., 2016), considerada entre las 50 plantas más amenazadas de Cuba (González-Torres et al, 2017). La distribución de esta especie se halla restringida a estos parches de bosques (Rodríguez-Alfaro & Castañeda-Noa, 2022).

De acuerdo con Murray & Kavanagh, (2016) la humedad relativa del aire es un factor crítico para la salud de los bosques de galería, ya que influye en la disponibilidad de agua para las plantas y en la regulación de la temperatura. La alta humedad favorece el crecimiento de especies umbrófilas, que son aquellas que requieren condiciones de sombra y humedad para prosperar como *Sachsia coronopifolia* (Ruiz Plasencia et al., 2019). La disminución de la humedad relativa, a menudo causada por la deforestación y el cambio climático, puede llevar a la degradación de estos ecosistemas y a la pérdida de especies sensibles (Murray & Kavanagh, 2016).

El estudio de las condiciones meteorológicas y su predicción, en específico de la humedad relativa del aire, resulta relevante para establecer prioridades de conservación dirigidas hacia estos bosques de galería. El análisis de esta variable mediante de modelos computacionales puede ser una herramienta útil en la predicción y cálculo de la tendencia de la misma en el futuro inmediato.

El propósito de este trabajo es realizar un estudio de la variabilidad de la humedad mediante técnicas de aprendizaje automático, lo que incluye determinar mediante el pronóstico a partir de modelos computacionales para la valoración desde el punto de vista ecológico de su marcha y tendencia.

Materiales y métodos

Para el desarrollo de la presente investigación se trabajó en tres etapas que se corresponden con la (1) recolección y procesamiento de los datos, (2) análisis de los resultados y (3) el análisis espacial del objeto de investigación.

Los valores de las variables Humedad Relativa (%) y Temperatura fueron tomados con termohigrómetros HOBO Pro 2, ubicados in situ en 5 bosques de galería al sureste de Santa Clara, abarcando las localidades: Corojito, Río Primero, El Playazo, Rebacadero y Sagüita (Figura 1).

Durante el período comprendido entre el 19 de julio al 31 de diciembre del año 2018 se recolectaron los datos. Las variables se tomaron diariamente de forma aleatoria en un rango de cinco a sesenta minutos durante la etapa especificada.

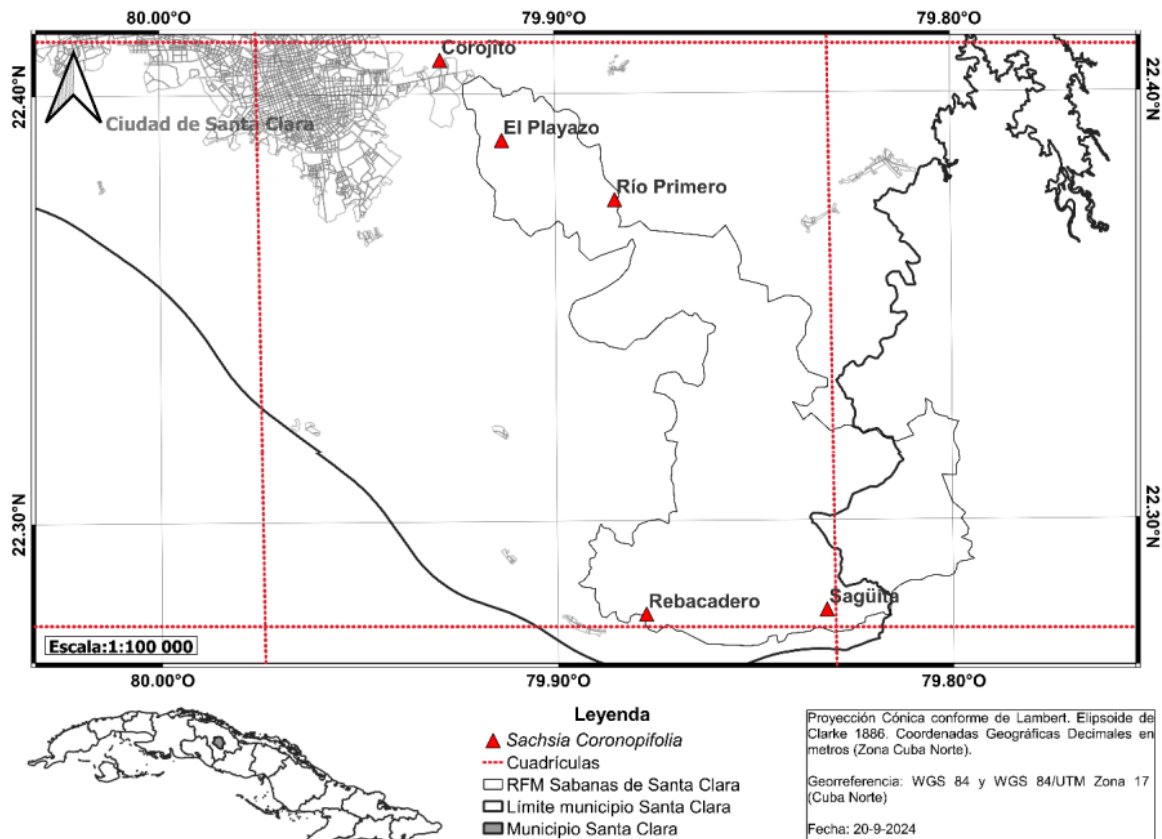


Figura 1. Área de estudios. Fuente: Elaboración grupo de trabajo

A los datos anteriores se les agregó el acumulado de precipitaciones procedentes de la red pluviométrica informativa del Instituto de Recursos Hidráulicos (INRH) y el Centro Meteorológico Provincial de Villa Clara. Se seleccionaron los equipos que por su proximidad resultaban los más representativos para el análisis. En la tabla 1 muestra los datos correspondientes al conjunto final de variables analizadas.

Tabla 1. Variables meteorológicas.

Nombre de la variable	Unidad	Instrumento de medición
Temperatura del aire	oC	Termohigrómetro HOBO
Humedad relativa del aire	%	Termohigrómetro HOBO Pro 2
Acumulado de precipitaciones en 24 horas	mm	Pluviómetro

Las mismas se delimitaron de la siguiente forma: fecha, hora, localidad, humedad relativa, temperatura y precipitaciones en 24 horas.

Mediante el uso de Sistemas de Información Geográfica (SIG) (QGIS versión 3.24 para sistema operativo Windows 10, se determinó que todas las poblaciones de plantas estudiadas se ubican en la misma rejilla, por lo que la variabilidad de los datos no es muy notable entre las localizaciones de las poblaciones dada su relativa cercanía.

Las series temporales son conjuntos de datos ordenados cronológicamente, donde cada observación está asociada a un punto específico en el tiempo. Estos datos pueden ser recopilados a intervalos regulares (como diarios, mensuales o anuales) o irregulares (Peixeiro, 2022) Para el análisis de las variables meteorológicas se siguieron los procedimientos estadísticos propios del análisis de series temporales. De las mismas se analizó:

Análisis de gráficos de la serie temporal

- Análisis de la tendencia de la serie temporal
- Análisis de la estacionalidad de la serie temporal
- Análisis de los residuos temporales
- Análisis de la correlación de la serie temporal
- Test de Dickey – Fuller Aumentado
- Detección de anomalías
- Análisis de la pendiente, para determinar la tendencia de la serie

Como los datos que se van a utilizar cumplen con lo anterior, se selecciona el ARIMA (*del inglés, AutoRegressive Integrated Moving Average*): Modelos que combinan componentes autoregresivas, integradas y de media móvil (Hamilton, 1994).

Seguidamente se listan los componentes de los modelos ARIMA (Atwan, 2022):

- Autorregresivo (AR): Este componente asume que el valor actual de la serie es una función de sus propios valores pasados.
- Integrado (I): Este componente maneja la no estacionariedad en los datos mediante el diferenciado de la serie para hacerla estacionaria.

- Media móvil (MA): Este componente asume que el valor actual es una función de errores pasados en el modelo.

El modelo ARIMA se representa típicamente por la notación ARIMA (p,d,q), donde (Nielsen, 2020):

- p: El orden del componente autorregresivo.
- d: El grado de diferenciación.
- q: El orden del componente de media móvil.

La búsqueda en cuadrícula es una técnica de ajuste de hiperparámetros comúnmente utilizada en el aprendizaje automático, incluido el análisis de series de tiempo con modelos ARIMA. Implica probar sistemáticamente diferentes combinaciones para encontrar el modelo de mejor rendimiento. Para la obtención se utilizó una búsqueda en cuadrícula (del inglés, Grid Search) (Cerqueira & Roque, 2024).

Para la validación del modelo ARIMA se usó la validación de avance secuencial (del inglés, Walk-Forward Validation). La misma es una técnica utilizada para evaluar el rendimiento de un modelo predictivo a lo largo del tiempo. A diferencia de otros métodos de validación, esta se enfoca en cómo el modelo se comporta con datos secuenciales o temporales, lo cual es esencial en aplicaciones donde los datos cambian con el tiempo, como en la predicción de series temporales. Dicha propiedad permite ajustar el modelo continuamente a medida que se dispone de nuevos datos y tener una precisión temporal, ideal para datos que tienen una dependencia de este tipo, y de esta forma el modelo no se “sobreajusta” a datos pasados (Brownlee, 2020).

Resultados y discusión

Verificación de la estacionalidad

Un requisito de los modelos ARIMA es el uso de series temporalmente estacionarias, por lo cual se realizaron varios test para comprobar el cumplimiento de tal condición.

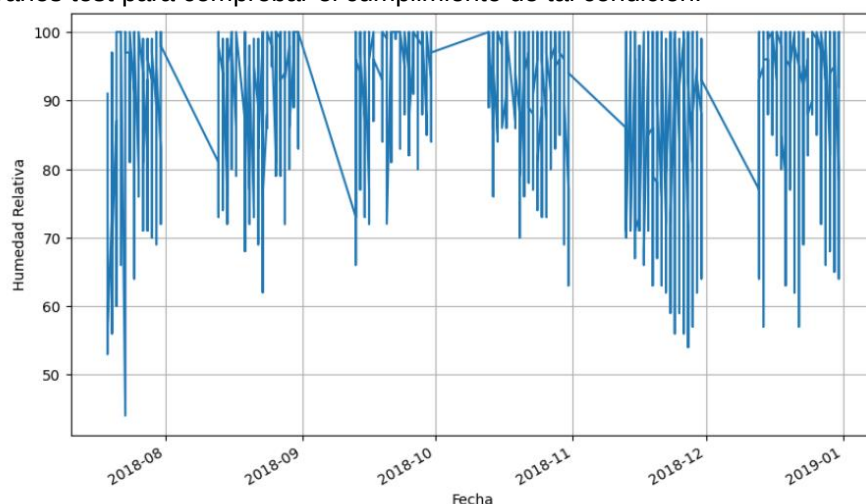


Figura 2. Análisis de gráficos de la serie temporal. Fuente: Elaboración propia

En la figura 2 se muestra la representación a través del tiempo de los valores de la serie de datos original; desde el mes de julio hasta diciembre del año 2018. La gráfica visible es la suma de todas las componentes (tendencia, estacionalidad y residuo) e ilustra la evolución de la variable en el periodo analizado.

Al descomponer la misma, (figura 3) la tendencia muestra una línea suave que captura el crecimiento o decrecimiento de la variable en un período de tiempo prolongado. Al ser estacional, se observan picos de aumento/disminución de la humedad relativa en el punto estudiado.

Se alcanza el máximo de humedad relativa en los meses de septiembre u octubre. Este pico coincide con el periodo lluvioso en Cuba y está influenciado positivamente por el aumento de las precipitaciones de acuerdo con Vichot-Llano y Martinez-Castro (2017). El bimestre septiembre-octubre es el último del periodo lluvioso en Cuba y es el segundo bimestre más lluviosos después del mayo-junio, además septiembre y octubre es el pico de temporada ciclónica donde por lo general octubre es un mes donde predominan bajas presiones en el mar caribe occidental y a la misma vez es un mes de transición entre el verano y el invierno y por lo general llega el primer frente frío en el mes de octubre. Todo esto hace que sea un mes con precipitaciones y por supuesto, esto trae consigo una mayor humedad relativa, mayor lluvia mayor sistemas de bajas presiones.

El mínimo de humedad relativa se obtuvo en diciembre, mes que corresponde al período seco en Cuba. El periodo seco en Cuba comienza en noviembre y se extiende hasta abril, en el periodo este entran cuatro frentes fríos al país por lo general, pero predomina en el país la presencia de las altas presiones continentales migratorias que tienen una gran masa de aire muy seco y estable que inhibe los procesos de precipitaciones y disminuye significativamente la humedad relativa. Anteriormente, durante los meses de noviembre de 2016 a abril de 2017 en la provincia de Villa Clara los mayores déficits de lluvias (extremos y severos) representaron el 40 % en Villa Clara, concentrándose en el municipio de Santa Clara y sus cercanías según (Otero Martín & Barcia Sardiñas, 2018). Este hecho se corresponde con los resultados obtenidos en el estudio sobre los meses de menor humedad relativa dentro del bosque de galería.

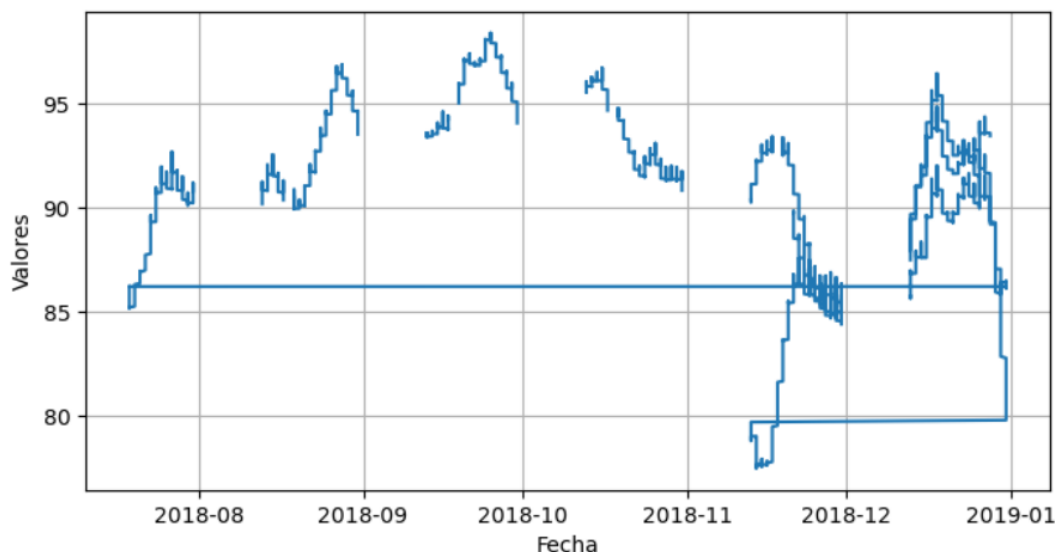
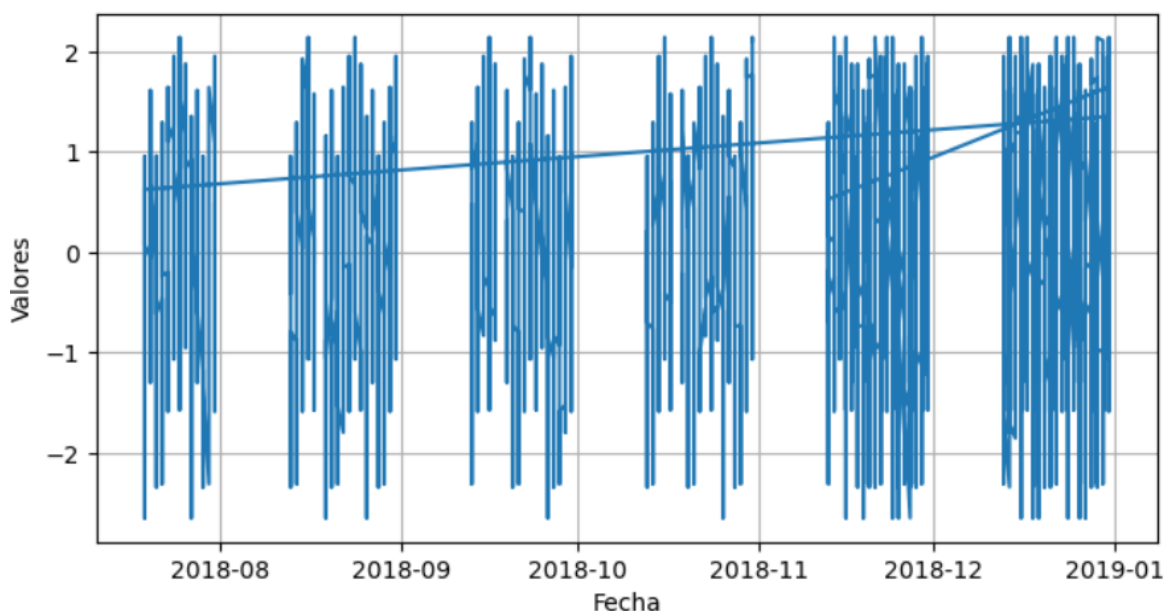
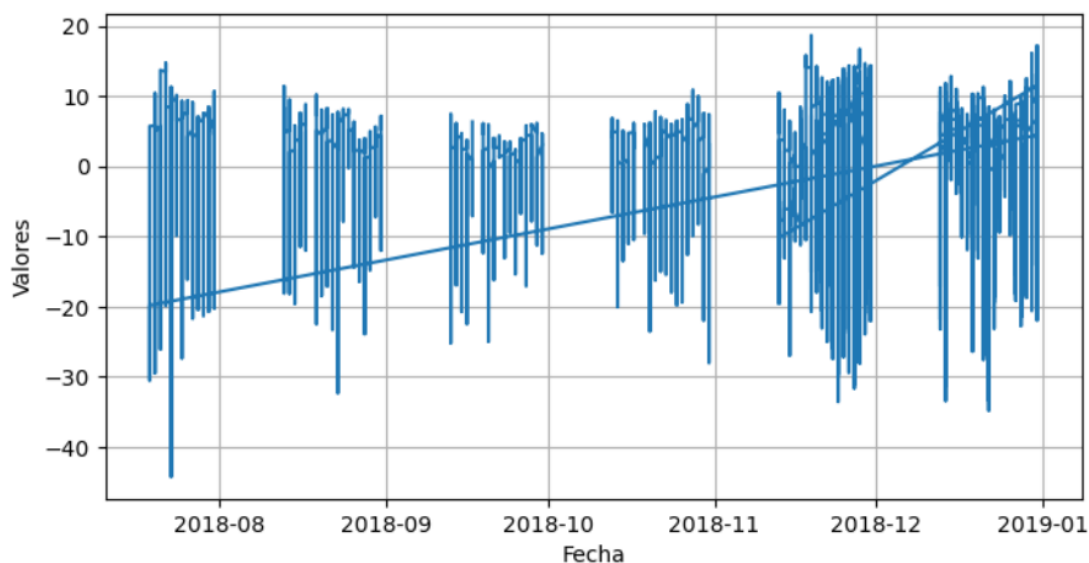


Figura 3. Análisis de la tendencia de la serie temporal. Fuente: Elaboración propia

La estacionalidad capta los patrones repetitivos que se producen en intervalos regulares dentro de la serie temporal. Estos suelen estar relacionados con fenómenos periódicos, como estaciones del año, días de la semana, o ciclos económicos. La figura 4 muestra las fluctuaciones regulares de la serie que ocurren de manera recurrente en un periodo definido (en este caso, seis meses). Si la humedad relativa varía en diferentes épocas del año (como en la temporada de lluvias), esto se reflejará en la estacionalidad.



Como se aprecia en el gráfico, dicho principio se cumple. Se pueden apreciar los patrones recurrentes en la marcha de la variable.



El residuo representa la parte de la serie de datos que no puede explicarse ni por la tendencia ni por la estacionalidad. Se considera el "ruido" o los componentes aleatorios que no siguen un patrón predecible. Si los residuos son pequeños y no muestran patrones visibles, significa que los modelos de tendencia y estacionalidad explican bien la serie temporal. Sin embargo, si hay patrones claros o grandes fluctuaciones en el residuo, podría ser necesario investigar más para entender esas variaciones. En este caso, no es necesario realizar exámenes posteriores, puesto están bien definidos (figura 5). Por otra parte, la autocorrelación mide el grado de similitud entre los valores de una serie temporal en diferentes momentos del tiempo. En otras palabras, es una medida que indica cuánto se relaciona un valor de la serie con valores anteriores o posteriores de la misma serie. La correlación de una variable consigo misma, pero desplazada en el tiempo.

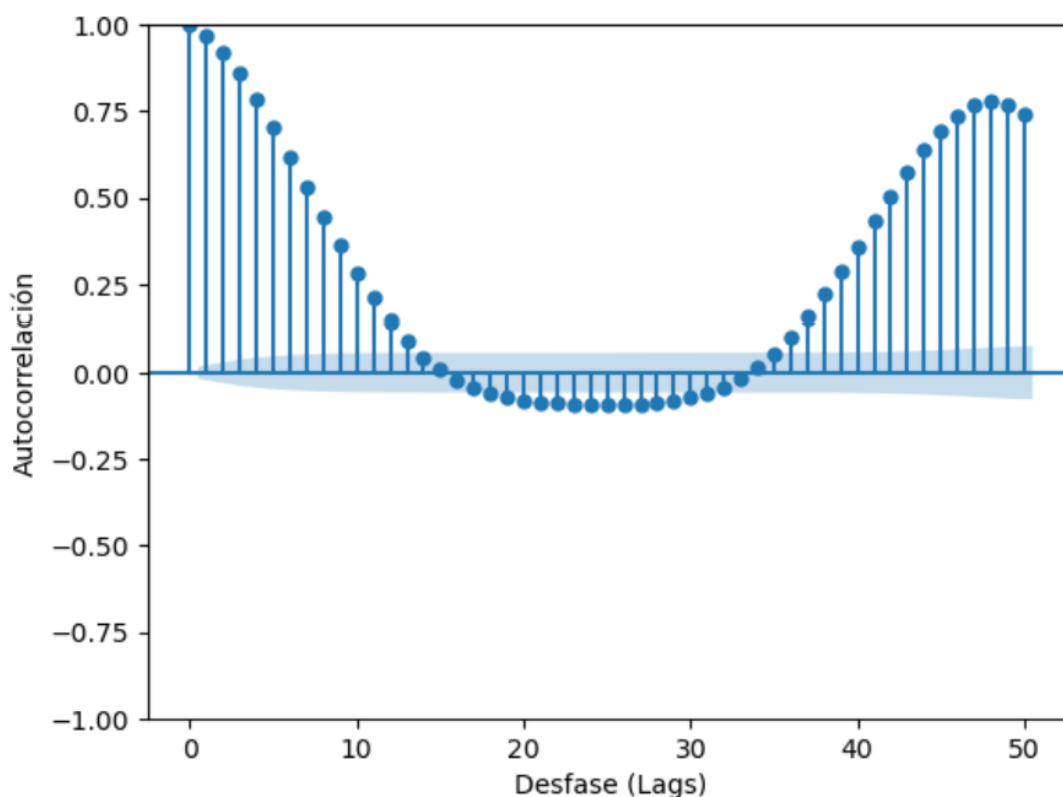


Figura 6. Análisis de la correlación de la serie temporal. Fuente: Elaboración propia

Se observa una alta autocorrelación inicial, ya que un valor cercano a 1.00 en el desfase 0 indica que los datos están perfectamente correlacionados consigo mismos en ese punto. Además, el patrón ondulante o la naturaleza sinusoidal sugieren la presencia de periodicidad en los datos. Esto significa que hay ciclos o patrones repetitivos en la serie temporal. Finalmente, el desfase negativo presente indica una relación inversa, lo que puede ser útil para identificar ciclos de comportamiento opuesto.

Test de Dickey-Fuller Aumentado (ADF)

El Test de Dickey – Fuller Aumentado (ADF) es una prueba estadística utilizada para determinar

si una serie temporal es estacionaria o no. La estacionariedad es una propiedad importante en el análisis de series temporales, ya que muchas técnicas de modelado, como el modelo ARIMA, asumen que los datos son estacionarios.

El Test de Dickey – Fuller Aumentado es una versión extendida del test original de Dickey – Fuller, que incluye términos adicionales para capturar mejor la dinámica temporal de la serie, como la presencia de retardos (lags). Se usa para probar la hipótesis nula (H_0) de que una serie temporal contiene una raíz unitaria, lo que indicaría que no es estacionaria.

Interpretación de los Resultados del Test ADF

- Valor P: Este valor es el indicador más importante para la interpretación.
- Si el valor p es menor que un nivel de significancia (0.05), rechazamos la hipótesis nula. Esto significa que la serie es estacionaria.
- Si el valor p es mayor que 0.05, no podemos rechazar la hipótesis nula, lo que indica que
- la serie no es estacionaria.

Tabla 2. Resultados del test

	ADF	p	Estacionalidad
Humedad relativa	-5.706706050599361	7.45950060847113e-07	SI

Boxplots por periodos

En la figura 6 se puede observar mediante un diagrama de caja, una visualización gráfica que muestra la distribución de una variable a través de cinco estadísticas resumidas:

1. Mediana: El valor central de la distribución (50° percentil).
2. Cuartiles: Dividen los datos en cuatro partes iguales. El primer cuartil (Q1) es el 25° percentil, y el tercer cuartil (Q3) es el 75° percentil.
1. Rango Intercuartil (IQR): La distancia entre Q1 y Q3, que cubre el 50% central de los datos.
2. Bigotes: Extienden desde Q1 y Q3 hasta el valor más lejano dentro de $1.5 * IQR$. Los valores fuera de este rango se consideran outliers (valores atípicos).
3. Outliers: Datos que están significativamente por fuera de los bigotes, indicados como puntos individuales en el gráfico.

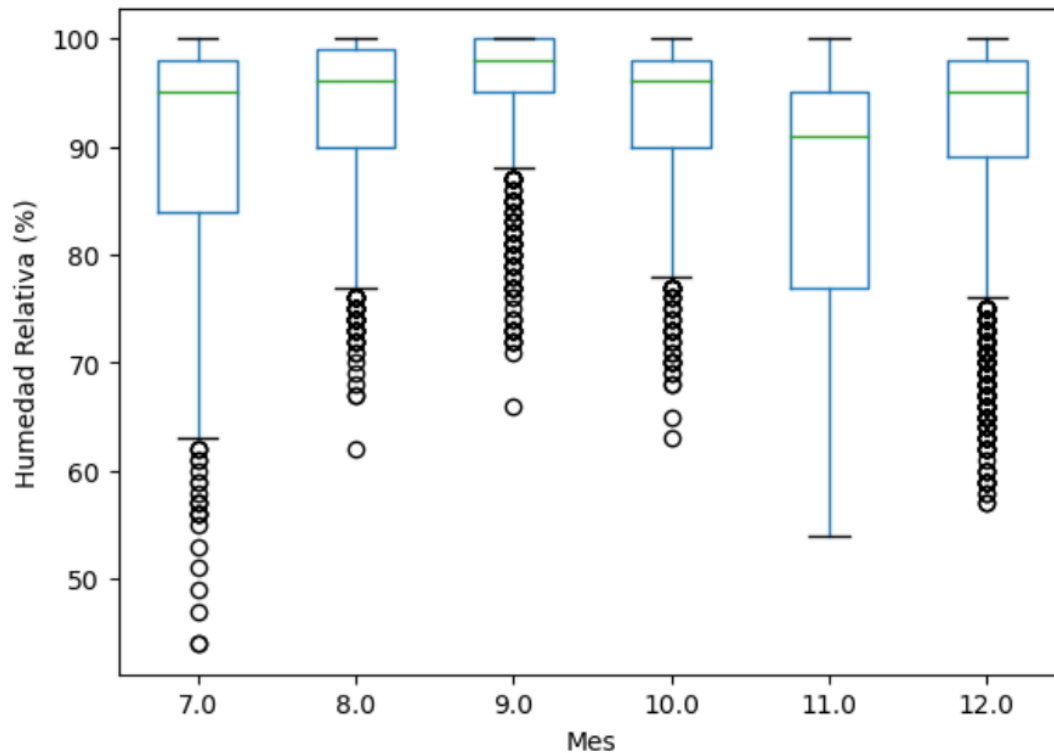


Figura 6. Análisis de las anomalías de la serie temporal. Fuente: Elaboración propia

Como se observa en la Figura 6, podemos afirmar la presencia de valores outliers. Estos valores describen valores de humedad bajos producto a condiciones secas en el ambiente.

Luego de analizar y demostrar la estacionalidad de la serie, se pasa al trabajo con el modelo ARIMA, el cual se realizó por etapas. Primero se buscó el mejor ajuste posible antes de la realización del pronóstico, mediante el uso del Grid Search para los parámetros (p, d, q) y como métrica de comparación se usó la Raíz del Error Medio Cuadrático (RMSE).

Tabla 2. Resultados de la búsqueda de los parámetros óptimos

ARIMA	(p ,d ,q)	RMSE
Rango de los valores (p, d, q)	(0,0,0)	6.202
	(0,0,1)	4.188
	(0,0,2)	3.589
p(0,1,2,4,6,8,10)	(0, 0, 3)	3.355
d(0,1,2,3,4,5)	(0, 1, 0)	2.977
	(0, 1, 2)	2.958
q(0,1,2,3,4,5)	(0, 1, 3)	2.884
	(1, 0, 3)	2.876

Como se observa en la tabla anterior la mejor configuración para nuestra serie de tiempo es (1, 0, 3) con un RMSE de 2.8757210550012258.

Como método de validación del modelo se usó la validación Walk – Forward sigue una estrategia en la que el modelo se entrena y evalúa en una serie de "ventanas" de datos a medida que avanzan en el tiempo. En lugar de usar un único conjunto de entrenamiento y prueba, se mueve la ventana de entrenamiento hacia adelante en el tiempo y se evalúa el modelo en cada paso. Para la comprobación del modelo también se usó la métrica RMSE, en este caso dejó que modelo solo presenta un 2.380. Posteriormente se realizó un pronóstico de 100 pasos adelante para comprobar cómo se comportaría la humedad relativa como se muestra en la figura 7. Conclusiones

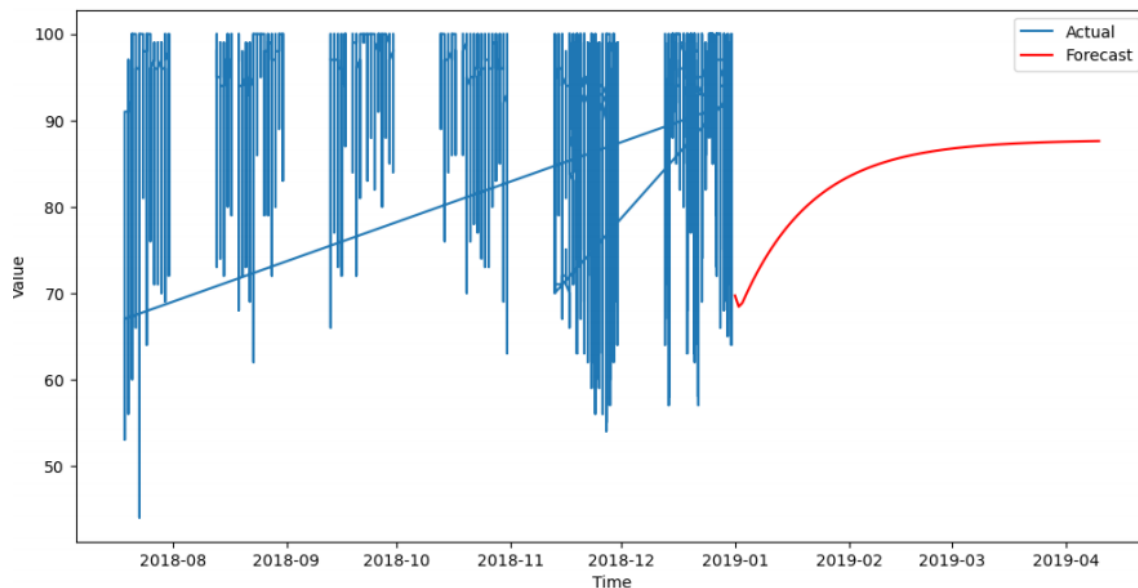


Figura 7. Pronostico 100 pasos adelante. Fuente: Elaboración propia

Para la comparación de los datos se utilizó el cálculo de la pendiente de los dos juegos de datos (reales y pronosticados) (Figura 8).

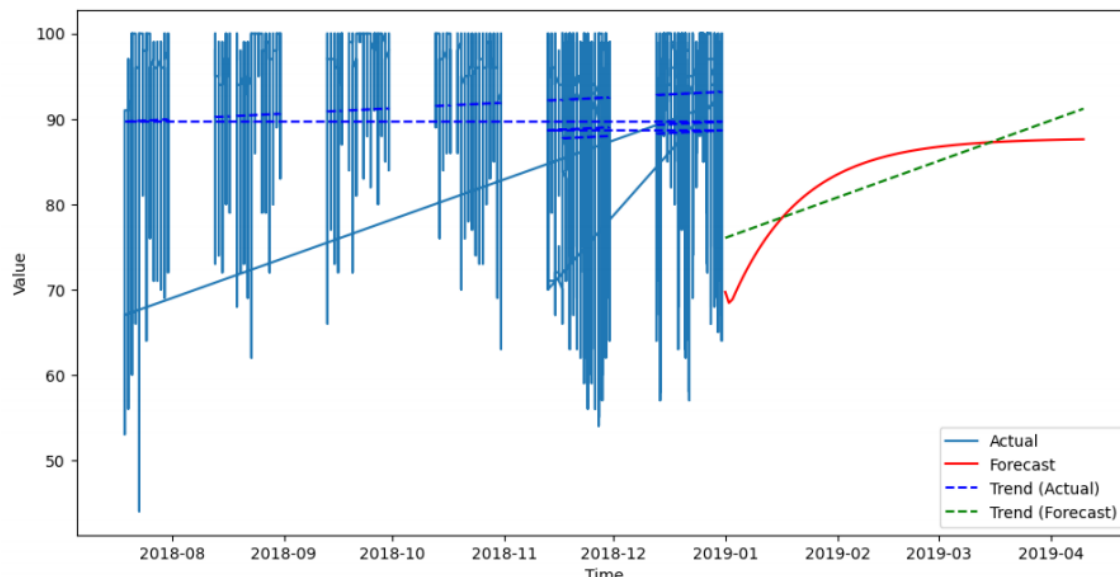


Figura 8. Comparación de las pendientes. Fuente: Elaboración propia

El resultado del cálculo de las pendientes mediante un simple algoritmo un ligero incremento de la humedad en la predicción (tabla 2). Tabla 2. Resultados del caculo de las pendientes

Pendiente	Resultado
Reales	0.00044035019366382006
Pronóstico	0.15303543114659382

La determinación de si el aumento es pronunciado es algo necesario para sostener el principio de la estacionalidad y de la meteorología. Ya se ha demostrado que la humedad en meteorología es estacional y tiende en el mes de octubre a disminuir; para luego en los meses de diciembre y enero con los frentes fríos a aumentar. Como se ve en la figura 6. La pendiente es 0.15303543114659382, el aumento no es pronunciado de manera que puede interpretarse de la siguiente forma que por cada unidad de cambio en el tiempo (eje x), el valor de los datos aumenta en promedio de 0.153 unidades. Aunque es una tendencia ascendente, no es pronunciada ni rápida. Esto puede ser útil para la planificación o la toma de decisiones, ya que indica un crecimiento sostenido, pero no agresivo.

Conclusiones

La marcha de la humedad relativa en la zona analizada sigue un patrón estacional, con máximos durante los meses de septiembre/octubre y mínimos en diciembre. Dicha condición permite su análisis mediante el modelo de aprendizaje ARIMA. Por otra parte, se demostró que la tendencia de esta variable es al aumento sostenido, pero no agresivo llegando aproximadamente a un pico de 90% de humedad relativa coincidente con la temporada invernal donde este variable tiende a aumentar.

Los meses de verano (especialmente agosto) y los meses de diciembre y enero se consideran los meses más secos. La humedad es un factor abiótico esencial para la supervivencia de *Sachsia coronopifolia*, su aumento sostenido a futuro dependerá de las precipitaciones anuales y de la cobertura boscosa, aspecto a considerar para futuras medidas dirigidas a la conservación de los bosques de galería al sureste de Santa Clara. Sin embargo, no se encontraron valores de humedad inferiores al 40%, esto demuestra que, a pesar de las sequías extremas comprendidas durante los años anteriores en los alrededores del área, los cursos de agua y la vegetación del bosque de galería crean un microclima con las condiciones de humedad propicias para el establecimiento de especies umbrófilas como *Sachsia coronopifolia*.

Referencias Bibliográficas

- González-Torres, L. R., Alejandro, P., Bécquer, E. R., Berazaín, R., Barrios, D., & Gómez, J. L. (2013). TOP 50 – Las 50 plantas más amenazadas de Cuba. *Bissea*, 7(NE 1).
- González-Torres, L. R., Palmarola, A., González-Oliva, L., Bécquer, E. R., Testé, E., & Barrios, D. (2016). Lista Roja de la Flora de Cuba. En L. R. González Torres, A. Palmarola, L. González Oliva, E. R.
- Bécquer, E. Testé, & D. Barrios (Eds.), *Bissea* (Vol. 10). *Bissea*.
[http://repositorio.geotech.cu/jspui/Murray, B. R., & Kavanagh, R. \(2016\). "The Role of Humidity in Forest Ecosystems: Implications for Biodiversity and Ecosystem Function." Forest Ecology and Management, 375, 1-10. https://doi.org/10.1016/j.foreco.2016.05.017](http://repositorio.geotech.cu/jspui/Murray, B. R., & Kavanagh, R. (2016). \)
- Otero Martín, M., y Barcia Sardiñas, S. (2018). Manifestaciones de la sequía meteorológica en la provincia de Villa Clara (Cuba) en el período 2000-2017. *Investigaciones Geográficas*. 70: 197-217.
<https://doi.org/10.14198/INGEO2018.70.10>
- Rodríguez-Alfaro, C. & Castañeda-Noa, I. E. (2022). Caracterización del hábitat de *Rhodogeron coronopifolius* Griseb. [Tesis de Titulación, Universidad Central "Marta Abreu" de Las Villas]. Repositorio institucional <https://dspace.uclv.edu.cu>
- UICN. (2012). Categorías y Criterios de la Lista Roja de la UICN: Versión 3.1 .Segunda edición. Gland, Suiza y Cambridge, Reino Unido: UICN. vi + 34pp.
- Originalmente publicado como IUCN Red List Categories and Criteria: Version 3.1 .Second edition. (Gland, Switzerland and Cambridge, UK: IUCN, 2012).
- Vichot-Ilano, A., & Martinez-Castro, D. (2017). Estado actual de las representación de los principales factores del clima del Caribe por modelos climáticos regionales . *Estudios de sensibilidad y validación*, 23(2), 232–261.
- Atwan, T. A. (2022). Time series analysis with Python cookbook: Practical recipes for exploratory data analysis, data preparation, forecasting, and model evaluation. Packt Publishing.
- Brownlee, J. (2020). How to Create an ARIMA Model for Time Series Forecasting in Python.
- Cerqueira, V., & Roque, L. (2024). Deep learning for time series cookbook: Use PyTorch and Python recipes for forecasting, classification, and anomaly detection (1st edition). Packt Publishing Ltd.

- González Torres, L. R., Palmarola, A., González Oliva, L., Bécquer, E. R., Testé, E., & Barrios, D. (2016). Lista Roja de la Flora de Cuba. En L. R. González Torres, A. Palmarola, L. González Oliva, E. R.
- Bécquer, E. Testé, & D. Barrios (Eds.), Bissea (Vol. 10, Número NE1). Bissea.
- Hamilton, J. D. (1994). Time series analysis. Princeton Univ. Press.
- Nielsen, A. (2020). Practical Time Series Analysis.
- Otero Martín, M., & Barcia Sardiñas, S. (2018). Manifestaciones de la sequía meteorológica en la provincia de Villa Clara (Cuba) en el período 2000-2017. Investigaciones Geográficas, 70, 197-217. <https://doi.org/10.14198/INGEO2018.70.10>
- Peixeiro, M. (2022). Time Series Forecasting in Python.
- Ruiz Plasencia, I., Hernández Albornas, J., & Ruiz Rojas, E. (2019). Catálogo de las áreas protegidas de Cuba. Centro Nacional de Áreas Protegidas, 386.