

Enfoque basado en datos y conocimiento experto para la zonificación climática

Data driven and expert knowledge approach for climate zoning

 Ariamna Crespo Lara*,  Yankier González Crespo,
 Julio C. Quintana Zaez and  Oscar Brown Manrique

Universidad de Ciego de Ávila “Máximo Gómez Báez”, Ciego de Ávila, Cuba.
E-mail: Yankier2008@gmail.com, cjquintanazaez@gmail.com, obrown@unica.cu

*Autor para correspondencia: Ariamna Crespo Lara, e-mail: ariamnacrespo@gmail.com

RESUMEN: La creciente variabilidad climática exige metodologías robustas para caracterizar los regímenes de lluvia, un factor crítico para la seguridad alimentaria y la gestión del agua. En Cuba, la aplicación de técnicas avanzadas de minería de datos para este fin representa un área subexplorada. Este estudio propone un marco novedoso para el agrupamiento de series de tiempo de precipitación que combina la extracción de características automatizadas (tsfresh) con indicadores hidroclimáticos definidos por expertos, aplicando posteriormente el algoritmo de Clustering Jerárquico (HCA). La metodología, validada con datos de 17 estaciones en Ciego de Ávila (1960-2013), identificó seis regímenes pluviométricos homogéneos con patrones estacionales e intensidades de eventos extremos significativamente diferentes. Este aporte permite una zonificación climática precisa, superando los enfoques estadísticos tradicionales. La proyección de este trabajo es directa para el contexto agrícola, ya que facilita el desarrollo de estrategias diferenciadas de cultivo, calendarios de siembra optimizados y una mejor gestión del riesgo climático, incrementando así la resiliencia del sector agropecuario cubano.

Palabras clave: agricultura, agrupamiento jerárquico, aprendizaje no supervisado, serie cronológica.

ABSTRACT: Increasing climate variability demands robust methodologies to characterize rainfall regimes, a critical factor for food security and water management. In Cuba, the application of advanced data mining techniques for this purpose remains an underexplored area. This study proposes a novel framework for clustering precipitation time series by combining automated feature extraction (tsfresh) with expert-defined hydroclimatic indicators, followed by the application of the Hierarchical Clustering Algorithm (HCA). The methodology, validated with data from 17 stations in Ciego de Ávila (1960-2013), identified six homogeneous rainfall regimes with significantly different seasonal patterns and extreme event intensities. This contribution enables precise climate zoning, surpassing traditional statistical approaches. The projection of this work is direct for the agricultural context, as it facilitates the development of differentiated cultivation strategies, optimized planting calendars, and improved climate risk management, thereby increasing the resilience of the Cuban agricultural sector.

Keywords: agriculture, hierarchical clustering, time series, unsupervised learning.

INTRODUCCIÓN

La proliferación de datos hidrometeorológicos, impulsada por los avances en las tecnologías de sensores y telemetría, ha consolidado el análisis de series temporales como una herramienta indispensable en la gestión de recursos hídricos y los estudios climáticos (Bonaccorso y Peres, 2022; Loukas et al., 2021). Entre estas, las series temporales de precipitación representan un componente crítico del ciclo hidrológico, cuya caracterización precisa

es esencial para comprender los eventos extremos, la planificación agrícola y la evaluación de los impactos del cambio climático (Zhang et al., 2021). En este contexto, la clusterización de series temporales emerge como una poderosa técnica de minería de datos para descubrir patrones espacio-temporales, identificar regiones homogéneas y comprender la estructura subyacente de grandes volúmenes de datos de precipitación sin necesidad de etiquetas previas, facilitando así un análisis exploratorio robusto (Duan et al., 2024; Singh et al., 2025).

Recibido: 19/11/2025

Aceptado: 02/04/2026

Conflicto de intereses: Los autores declaran que no existen conflictos de interés.

Contribución de los autores: **Conceptualización:** Ariamna Crespo Lara, Yankier González Crespo. **Curación de datos:** Ariamna Crespo Lara, Yankier González Crespo, Julio Cesar Quintana Zaez. **Investigación:** Ariamna Crespo Lara, Yankier González Crespo. **Metodología:** Ariamna Crespo Lara, Yankier González Crespo, Oscar Brown Manrique. **Supervisión:** Julio Cesar Quintana Zaez, Oscar Brown Manrique. **Validación:** Julio Cesar Quintana Zaez, Oscar Brown Manrique. **Redacción - borrador original:** Ariamna Crespo Lara, Yankier González Crespo. **Redacción - revisión y edición:** Ariamna Crespo Lara, Yankier González Crespo, Julio Cesar Quintana Zaez, Oscar Brown Manrique.

La mención de marcas comerciales de equipos, instrumentos o materiales específicos obedece a propósitos de identificación, no existiendo ningún compromiso promocional con relación los mismos, ni por los autores ni por el editor.



Este artículo se encuentra bajo los términos de la licencia Creative Commons Attribution-NonCommercial (CC BY-NC 4.0).
<https://creativecommons.org/licenses/by-nc/4.0/>



La comunidad científica internacional ha desarrollado múltiples enfoques para la clusterización de series temporales, los cuales pueden categorizarse en tres grupos principales: métodos basados en la proximidad temporal que utilizan medidas como Dynamic Time Warping (DTW) para capturar similitudes en la forma de las series [Fransiska et al. \(2024\)](#); [Lee et al. \(2025\)](#); métodos basados en modelos que emplean parámetros de modelos autorregresivos o modelos ocultos de Márkov [Hernández et al. \(2022\)](#); [Ghassempour et al. \(2014\)](#); y métodos basados en características (features) que transforman la serie en vectores de atributos precomputados ([Arulraj et al., 2023](#); [Tiano et al., 2021](#)).

Más recientemente, han surgido técnicas avanzadas que incorporan arquitecturas de aprendizaje profundo, como autoencoders y redes neuronales recurrentes, que permiten aprender representaciones no lineales de las series temporales ([Paparrizos et al., 2024](#)). Paralelamente, se han desarrollado herramientas computacionales especializadas como tslearn para facilitar la implementación y comparación de estos algoritmos ([Tavenard et al., 2020](#)).

En el dominio específico de la precipitación, la clusterización ha demostrado su utilidad en diversas aplicaciones. Estudios previos han empleado con éxito técnicas como k-Shape para el análisis de movilidad espacio-temporal [Paparrizos y Gravano \(2017\)](#), clusterización jerárquica con medidas de correlación para el análisis de frecuencia regional [Zaifoglu et al. \(2018\)](#), y métodos basados en características para identificar hietogramas de patrones de lluvia ([Zhang et al., 2024](#)). En particular, investigaciones en Indonesia han utilizado la distancia de ensamble para la clusterización de datos de lluvia, mientras que en India se han aplicado enfoques combinados de clusterización jerárquica y transformada wavelet para analizar tendencias estacionales de precipitación ([Hendrawati et al., 2024](#)). Estas investigaciones internacionales resaltan el potencial de estas técnicas para el análisis hidroclimático regional.

Sin embargo, en el contexto nacional, la aplicación de estas metodologías modernas de clusterización a series temporales de precipitación representa un nicho de investigación en gran medida inexplorado. Los estudios existentes se han centrado predominantemente en análisis estadísticos tradicionales o enfoques climáticos convencionales, con una implementación limitada de técnicas avanzadas de minería de datos y aprendizaje automático para la caracterización espacio-temporal de los regímenes de lluvia ([García Moya et al., 2024](#)). Esta brecha metodológica representa una oportunidad significativa para desarrollar investigaciones que adapten y optimicen estos algoritmos a las particularidades del clima cubano, caracterizado por su variabilidad temporal y espacial y la influencia de sistemas meteorológicos complejos.

Para cubrir esta brecha metodológica y aprovechar las oportunidades identificadas, este artículo propone un marco de trabajo (framework) para la clusterización de series temporales de precipitación basado en la extracción de características precomputadas y la aplicación del Algoritmo

de Clusterización Jerárquica (Hierarchical Clustering Algorithm, HCA). Esta metodología combina la eficiencia computacional del HCA con la riqueza informativa de un conjunto de características diseñadas para capturar aspectos clave del régimen de lluvias, tales como su intensidad, variabilidad, estacionalidad y comportamiento extremo. El objetivo principal es desarrollar una metodología accesible y escalable que permita identificar regiones climáticamente homogéneas en Cuba con alta precisión, contribuyendo así al perfeccionamiento de los modelos de predicción, la gestión de riesgos de sequías e inundaciones y la planificación de los recursos hídricos en un contexto de cambio climático. La propuesta se valida utilizando datos de precipitación a largo plazo, demostrando su eficacia en la agrupación coherente de estaciones y su utilidad como herramienta para el análisis hidroclimático regional.

MATERIALES Y MÉTODOS

Esta sección describe la metodología empleada en el artículo, detallando el diseño de la investigación y su implementación práctica. La selección de métodos y técnicas específicas se justifica para asegurar que el estudio pueda ser replicado por otros investigadores.

Contexto y conjunto de datos

Los datos de precipitación de las estaciones experimentales de Ciego de Ávila fueron obtenidos del CITMA mediante solicitud realizada por el Jefe del Grupo Científico de Recursos Hídricos de la Universidad de Ciego de Ávila. El conjunto de datos cubre el período de 1960 a 2013 y fue recopilado a través de encuestas a especialistas. La base de datos se proporciona en un archivo de Excel e incluye registros de 107 estaciones experimentales distribuidas en la provincia. Para cada estación, los datos están organizados por precipitación diaria, mes, año, código ID de la estación y municipio. Los valores de precipitación se expresan en milímetros (mm).

Como herramientas, se utilizó la librería Pandas de Python para leer y procesar los datos, ofreciendo una gama de métodos que permiten una extracción eficiente de datos mediante un código simplificado. Durante el proceso de carga de datos, varias estaciones experimentales fueron descartadas debido a errores significativos en el registro de datos, resultando en un conjunto final de 17 estaciones con datos válidos para el estudio. La información de todas las estaciones consideradas puede solicitarse en los materiales suplementarios.

Metodología

La metodología aplicada en este trabajo constó de cinco etapas que conformaron el flujo de trabajo del proceso propuesto para la clusterización de datos de series temporales mediante HCA ([Figura 1](#)). Dichas etapas fueron las siguientes: carga de datos - normalización, generación de características (*features*), análisis de datos, ajuste del algoritmo, evaluación de métricas y descripción de los clústeres:

- Carga de datos - normalización: Tras una rigurosa depuración de datos que excluyó estaciones con datos faltantes extensos, fallos de metadatos o bajos puntajes de confiabilidad, el conjunto de datos depurado se sometió a una normalización MinMaxScaler para garantizar una contribución equitativa de las características en la clusterización.
- Generación de características (*Features*): La ingeniería de características se realizó mediante un enfoque híbrido que combina la computación con la experiencia en el dominio. Utilizando tsfresh [Christ et al. \(2018\)](#), se aplicó segmentación por ventanas deslizantes para extraer patrones temporales, mientras que las características estadísticas (media, varianza, tendencias) se calcularon a partir de cada ventana. Estas características basadas en datos se enriquecieron con indicadores hidrológicos definidos por expertos, creando un conjunto de características multidimensional que captura tanto patrones algorítmicos como significado físico para el posterior análisis de clusterización.
- Análisis de datos: Un análisis de correlación identificó los impulsores hidrológicos clave y la posible multicolinealidad, estableciendo el marco de interpretabilidad fundamental para la caracterización posterior de los clústeres.
- Ajuste del algoritmo: Se seleccionó la clusterización jerárquica con el enlace de Ward para modelar las relaciones entre estaciones. Este algoritmo destaca por generar una estructura de dendrograma clara que representa visualmente la jerarquía anidada de las estaciones. El método de Ward minimiza la varianza intra-clúster, produciendo clústeres compactos y esféricos que ofrecen un valor explicativo superior al ilustrar directamente la lógica de agregación y las similitudes relativas entre estaciones en diferentes niveles jerárquicos.
- Evaluación de métricas: La calidad del clúster se evaluó mediante un marco de triple métrica: el Coeficiente de Silueta (Silhouette Score, que mide la cohesión y separación), el Índice de Calinski-Harabasz (que evalúa la dispersión entre clústeres) y el Índice de Davies-Bouldin (que cuantifica la distinción entre clústeres). Este enfoque de validación multifacético proporciona

perspectivas complementarias sobre la optimalidad de la partición, asegurando que la solución seleccionada demuestre tanto robustez matemática como relevancia hidrológica más allá de lo que cualquier métrica única podría confirmar.

- Descripción de los clústeres: La caracterización final de los clústeres emplea un marco analítico integral de cinco pilares: pruebas estadísticas inferenciales para identificar diferencias significativas entre clústeres, evaluación de vulnerabilidad para examinar la exposición al riesgo climático, análisis de tendencias temporales para descubrir patrones de precipitación, mapeo regional para contextualizar las distribuciones espaciales y validación por expertos para fundamentar los hallazgos algorítmicos en el conocimiento del dominio. Esta interpretación multidimensional transforma los clústeres abstractos en tipologías hidrológicas procesables.

Algoritmo

Este estudio combinó la ingeniería de características con el enlace de Ward para formar grupos compactos e interpretables, mitigando la complejidad dimensional y revelando regímenes dinámicos distintos mediante una visualización multiresolución de los datos:

- Clusterización Jerárquica: descubre estructuras latentes dentro de datos de series temporales de alta dimensionalidad, aprovechando su adaptabilidad inherente a este dominio complejo. El marco metodológico combina estratégicamente la ingeniería de características con el enlace de Ward para abordar desafíos clave. Al transformar las series brutas en un conjunto curado de características sobresalientes (mediante tsfresh e ingeniería de características), el algoritmo mitiga el problema de la alta dimensionalidad y mejora la eficiencia computacional.
- Método de enlace de Ward: minimiza la varianza intra-clúster para formar grupos compactos y esféricos que facilitan la interpretación ([Duan et al., 2024](#); [Radovanovic et al., 2020](#); [Rodrigues et al., 2006](#)). Este enfoque sustituye la necesidad de medidas de similitud directas sobre datos brutos, como DTW, con una métrica de distancia robusta en el dominio de las características diseñadas.



Figura 1. Flujo de trabajo para la clusterización de series de tiempo

- Dendrograma resultante: proporciona una visualización multiresolución de la jerarquía de datos, revelando tanto patrones generales como anomalías detalladas. En consecuencia, este flujo de trabajo facilita el descubrimiento de clústeres significativos que corresponden a regímenes dinámicos distintos dentro de los datos temporales, equilibrando el rigor analítico con la obtención de información procesable.

Métricas de validación interna

La solución de clusterización se evaluó cuantitativamente utilizando una tríada de métricas de validación interna; el Coeficiente de Silueta, el Índice de Calinski-Harabasz y el Índice de Davies-Bouldin:

- Coeficiente de Silueta: evalúa qué tan bien se sitúa cada objeto dentro de su clúster midiendo la cohesión y la separación, y valores cercanos a 1 indican un modelo superior.
- Índice de Calinski-Harabasz: conocido también como Criterio de Razón de Varianza, evalúa la calidad de los clústeres comparando la relación entre la dispersión entre clústeres y la dispersión dentro del clúster, donde una puntuación más alta significa clústeres mejor definidos (Shahapure y Nicholas, 2020; Wang y Xu, 2019).
- Índice de Davies-Bouldin: cuantifica la similitud promedio entre cada clúster y su clúster más similar, donde una solución óptima alcanza el valor más bajo posible, indicando la máxima distinción entre clústeres (Ashari et al., 2022).

En conjunto, este marco de validación multifacético proporciona una evaluación robusta y no redundante de la calidad de la partición, asegurando que los resultados sean tanto estadísticamente sólidos como estructuralmente significativos.

RESULTADOS Y DISCUSIÓN

La aplicación de la metodología propuesta generó información significativa sobre los regímenes de precipitación de Ciego de Ávila. La fase inicial implicó un análisis exhaustivo de las características generadas para comprender sus relaciones y relevancia. Posteriormente, se ajustó el Algoritmo de Clusterización Jerárquica, identificando la partición óptima de los datos. Finalmente, los clústeres resultantes se interpretaron y discutieron rigurosamente, integrando los hallazgos cuantitativos con el conocimiento experto del dominio, traduciendo el resultado algorítmico en una caracterización climática procesable.

Análisis de la correlación de características

Esta investigación desarrolló una estrategia de ingeniería de características que integra tres grupos distintos de características para capturar las características integrales de las series temporales. La [Tabla 1](#) muestra los grupos de características. El primer grupo comprende características automatizadas generadas mediante la librería tsfresh de Python, que extrae sistemáticamente una amplia gama de patrones temporales. El segundo grupo incorpora características específicas del dominio derivadas del conocimiento experto, asegurando la alineación con los fundamentos teóricos. El tercer grupo emplea un enfoque de ventana deslizante con 30 días de rezago para capturar dependencias temporales y patrones de evolución.

Colectivamente, estos conjuntos de características comprimen la información multidimensional de las series temporales en una representación estructurada adecuada para algoritmos de clusterización. Adicionalmente, se realizó un análisis de correlación para identificar relaciones clave, proporcionando información fundamental para la posterior interpretación de los clústeres y estudios de caracterización.

Tabla 1. Grupos de características para el análisis de series temporales

Tfresh	Expertos	Rolling window
fft_coefficient_attr_ "real"__coeff_0	dry_season_intensity	window_variability
fft_aggregated_aggtype_ "centroid"	dry_season_std	window_stability
number_peaks__n_3	rainy_season_intensity	window_autocorrelation
change_quantiles_isabs_qh_0.9_ql_0.1	rainfall_regime_contrast	window_consistency
change_quantiles_isabs_qh_0.8_ql_0.2'	hurricane_season_intensity	n_valid_windows
abs_energy'	hurricane_season_variability	
mean_abs_change	extreme_event_frequency	
autocorrelation_lag_1	extreme_event_intensity	
autocorrelation_lag_7	interannual_variability	
autocorrelation_lag_30	interannual_trend	
quantile__q_0.95	total_precipitation	
quantile__q_0.99	mean_precipitation	
c3_lag_1	std_precipitation	
cid_ce_normalize_True	cv_precipitation	
	zero_ratio	

Es notable que las características derivadas por expertos fueron diseñadas específicamente para capturar la dinámica climática relevante para la región del Caribe. Variables como *dry_season_intensity*, *hurricane_season_intensity*, *extreme_event_frequency* e *interannual_trend* se incorporaron basándose en la experiencia del dominio sobre su importancia en la caracterización de los patrones de lluvia.

Un análisis de correlación de Spearman reveló varias relaciones clave: *dry_season_intensity* y *total_precipitation* exhibieron una correlación positiva moderada ($\rho = 0,544$), mientras que *rainfall_regime_contrast* y *hurricane_season_intensity* mostraron una fuerte correlación positiva ($\rho = 0,861$). Por el contrario, *extreme_event_frequency* y *extreme_event_intensity* demostraron una fuerte correlación negativa ($\rho = -0,761$). Estas relaciones proporcionan información crucial para interpretar los clústeres resultantes e identificar regímenes distintivos de series temporales.

La Figura 2 ilustra la estructura de correlación entre todas las características, revelando varios grupos distintos de variables interrelacionadas. Comprender estas relaciones es particularmente valioso para aplicaciones agrícolas, ya

que puede informar los procesos de toma de decisiones de los agricultores y apoyar la optimización de las prácticas agrícolas en respuesta a la variabilidad climática.

Análisis de la clusterización

En la Figura 3, el análisis de la clusterización jerárquica identificó seis grupos distintos organizados dentro de una estructura jerárquica bien definida. Esta agrupación revela diferentes regímenes estacionales que son críticos para el sector agrícola, ya que permiten la caracterización de patrones climáticos específicos y su impacto en los ciclos de los cultivos.

La determinación del número óptimo de clústeres se basó en métricas de validación como Silueta, Calinski-Harabasz y Davies-Bouldin, que confirmaron consistentemente que seis clústeres representan la configuración más óptima, como se muestra en la Figura 4. Esto es relevante porque, aunque otras agrupaciones podrían sugerirse visualmente en el dendrograma, las métricas objetivas validan que esta segmentación captura de manera más efectiva las variaciones estacionales que afectan directamente la planificación agrícola, la selección de cultivos y la gestión del riesgo climático en la región.

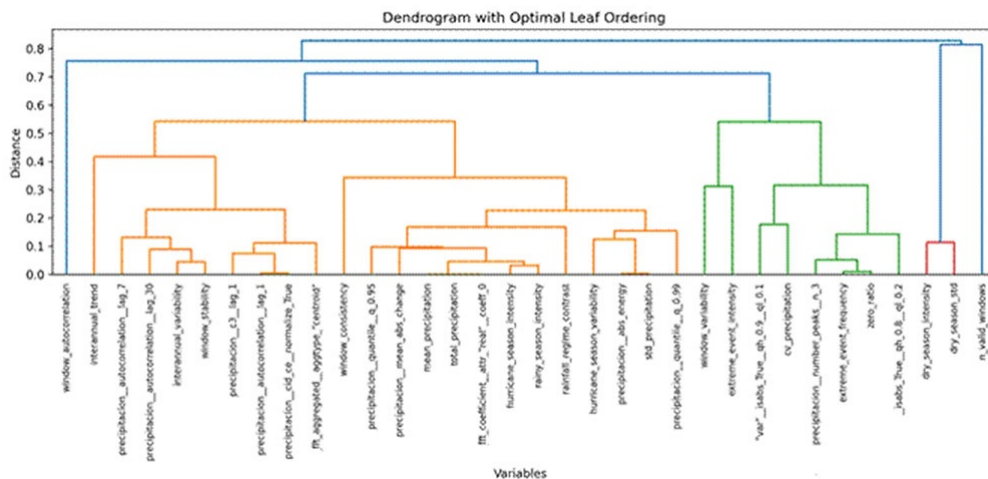


Figura 2. Dendrograma de correlación de características para variables seleccionadas.

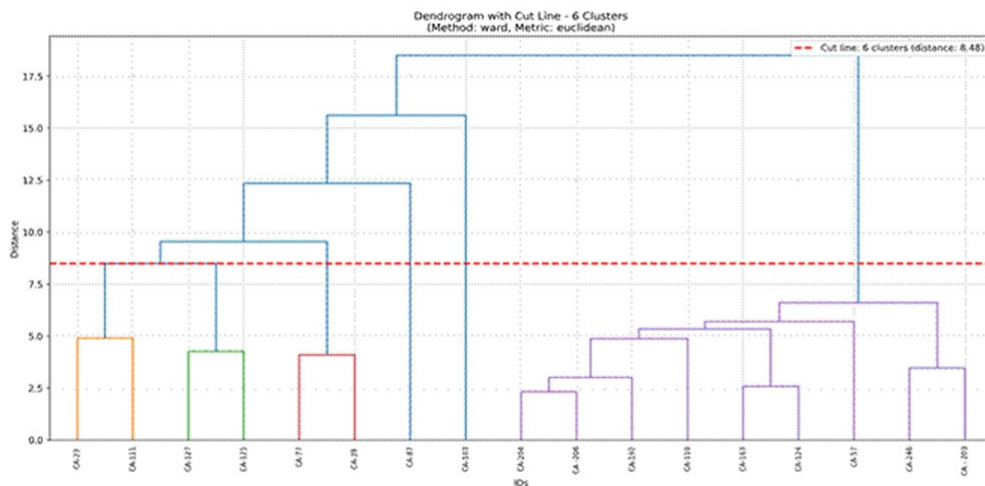


Figura 3. Clusterización jerárquica de las estaciones.

En la propia Figura 4, las métricas de validación de la clusterización arrojaron un rendimiento moderado, con una puntuación de Silueta de 0,34, lo que indica una separación de la estructura débil. Un índice de Calinski-Harabasz de 9,0 sugiere una diferenciación limitada entre clústeres, mientras que la puntuación de Davies-Bouldin de 0,6 refleja una compacidad razonable de los clústeres. Estos resultados colectivos implican clústeres superpuestos con patrones atenuados pero detectables en la estructura de los datos.

Caracterización de los Clústeres

El análisis de clusterización jerárquica identificó seis regímenes de lluvia distintos en la región del Caribe, cada uno caracterizado por patrones de variabilidad estacional específicos. El Clúster 5 emergió como el más intenso, exhibiendo valores máximos en precipitación (componente real FFT), variabilidad de la estación seca e intensidad de la temporada de huracanes, lo que sugiere un régimen climático de alta energía. Por el contrario, el Clúster 0 mostró valores mínimos en estas variables, indicando un patrón de lluvia más estable y menos intenso. Estos extremos representan polos opuestos de la variabilidad climática regional, con implicaciones directas para la planificación agrícola estacional y las estrategias de gestión de recursos hídricos adaptadas a las características específicas de cada régimen.

El análisis inferencial confirmó diferencias estadísticamente significativas ($p < 0.01$) en variables climáticas críticas, incluyendo la frecuencia de picos de precipitación, las tendencias interanuales y la intensidad de huracanes. Las pruebas post-hoc revelaron que el Clúster 4 exhibe significativamente más picos de precipitación en comparación con los Clústeres 0, 1, 2, 3 y 5, mientras que el Clúster 1 demuestra una tendencia interanual sustancialmente más fuerte que el Clúster 5. El análisis de vulnerabilidad identificó al Clúster 1 como el grupo de mayor riesgo, combinando una elevada frecuencia de eventos extremos con una pronunciada intensidad en la estación seca, creando desafíos críticos para los sistemas agrícolas de secano que dependen de patrones estacionales predecibles.

La distribución geográfica de los clústeres revela patrones espaciales distintos, con predominio del Clúster 0 en 1º de Enero, el Clúster 1 en Chambas y el Clúster

5 en Venezuela, lo que sugiere una zonificación climática natural. Esta segmentación permite estrategias agrícolas diferenciadas: las regiones como Chambas (Clúster 1) requieren cultivos resistentes a la sequía y sistemas de riego suplementario, mientras que áreas como Venezuela (Clúster 5) necesitan protocolos de preparación para alta variabilidad climática e intensidad de huracanes. Esta caracterización de los regímenes de lluvia proporciona una base científica para la selección de cultivos, la optimización de los calendarios de siembra y el diseño de sistemas de gestión del agua adaptados a cada patrón climático identificado, mejorando en última instancia la resiliencia agrícola en la cuenca del Caribe.

CONCLUSIONES

Como conclusión, esta investigación demuestra la eficacia de un marco de clusterización híbrido, que combina características basadas en datos con conocimiento experto del dominio, para descifrar la complejidad de los regímenes pluviométricos en Ciego de Ávila.

La metodología no solo identificó seis regímenes pluviométricos estadísticamente distintos, que van desde sistemas estables hasta sistemas de alta energía con marcada influencia ciclónica, sino que también traduce estos hallazgos a la práctica agrícola.

La caracterización de los clústeres de alta vulnerabilidad, proporciona una herramienta objetiva para la toma de decisiones, permitiendo la priorización de intervenciones como cultivos resistentes a la sequía o sistemas de riego en zonas específicas.

Esta contribución trasciende el ámbito académico, sentando las bases para una zonificación climática precisa y escalable a nivel nacional. Esto es fundamental para diseñar políticas agrícolas e hídricas resilientes, optimizar los calendarios de siembra y además, fortalecer la seguridad alimentaria frente a la variabilidad climática.

REFERENCIAS BIBLIOGRÁFICAS

ARULRAJ, M.; PETKOVIC, V.; FERRARO, R.R.; MENG, H.: "Precipitation Vertical Structure Characterization: A Feature-Based Approach", *Journal of Hydrometeorology*, 24(12): 2281-2297, 2023, ISSN: 1573-1650, DOI: <https://doi.org/10.1175/JHM-D-23-0034.1>

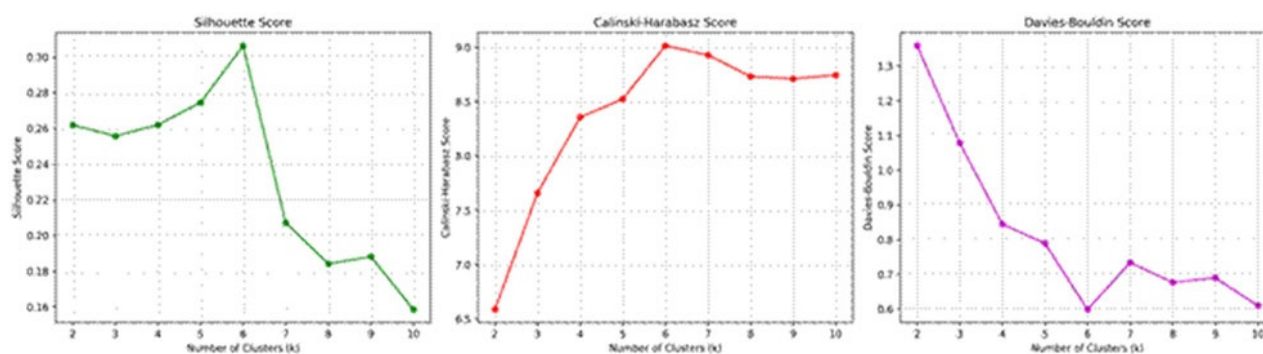


Figura 4. Resultados del Índice de Silueta, Índice de Calinski-Harabasz e Índice de Davies-Bouldin.

- ASHARI, I.F.; BANJARNAHOR, R.; FARIDA, D.R.; AISYAH, S.P.; DEWI, A.P.; HUMAYA, N.: "Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies", *Journal of Applied Informatics and Computing*, 6(1): 07-15, 2022, ISSN: 2548-6861, DOI: <https://doi.org/10.30871/ja-ic.v6i1.3485>
- BONACCORSO, B.; PERES, D.J.: "Analysis of Extreme Hydrometeorological Events", *Resources*, 11(6): 55, 2022, ISSN: 2079-9276, DOI: <https://doi.org/10.3390/resources11060055>
- CHRIST, M.; BRAUN, N.; NEUFFER, J.; KEMPA-LIEHR, A.W.: "Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh - A Python package)", *Neurocomputing*, 307: 72-77, 2018, ISSN: 1872-8286, DOI: <https://doi.org/10.1016/j.neucom.2018.03.067>
- DUAN, H.; LI, Q.; HE, L.; ZHANG, J.; AN, H.; ALI, R.; VAZIFEDOUST, M.: "Climate Classification for Major Cities in China Using Cluster Analysis", *Atmosphere*, 15(7): 741, 2024, ISSN: 2073-4433, DOI: <https://doi.org/10.3390/atmos15070741>
- FRANSISKA, H.; AGUSTINA, D.; SETYORINI, D.; SUMARTAJAYA, I.M.; KURNIA, A.: "Time Series Clustering Analysis Using Dynamic Time Warping Technique of Daily Rainfall in Bengkulu Province", *IOP Conference Series: Earth and Environmental Science*, 1359(1): 012026, 2024, ISSN: 1755-1315, DOI: <https://doi.org/10.1088/1755-1315/1359/1/012026>
- GARCIA-MOYA, A.; ALONSO-HERNÁNDEZ, C.M.; SÁNCHEZ-MURILLO, R.; MORERA-GÓMEZ, Y.; SÁNCHEZ-LLULL, M.; DÍAZ RIZO, O.; CUESTA SANTOS, O.; LÓPEZ LEE, R.; BRÍGIDO FLORES, O.; RAMOS VILTRE, E.O.; ORTEGA, L.: "Spatiotemporal characterization of the isotopic composition of meteoric waters in Cuba", *Scientific Data*, 11(1): 1398, 2024, ISSN: 2542-2987, DOI: <https://doi.org/10.1038/s41597-024-04178-z>
- GHASSEMPOUR, S.; GIROSI, F.; MAEDER, A.: "Clustering Multivariate Time Series Using Hidden Markov Models", *International Journal of Environmental Research and Public Health*, 11(3): 2741-2763, 2014, ISSN: 1660-4601, DOI: <https://doi.org/10.3390/ijerph110302741>
- HERNÁNDEZ, C.A.; MARTÍNEZ, F.H.; GIRAL, D.A.: "Evaluación del desempeño de modelos de decisión espectral predictivos", *Información tecnológica*, 33(3): 149-158, 2022, ISSN: 0718-0764, DOI: <http://dx.doi.org/10.4067/S0718-07642022000300149>
- HENDRAWATI, T.; WIGENA, A.H.; SUMERTAJAYA, I.M.; SARTONO, B.; PRAVITASARI, A.A.; ASNAWI, M.H.: "The ensemble distance on model-based clustering for regions clustering based on rainfall: The case of rainfall in West Java Indonesia", *International Journal of Data and Network Science*, 8(2): 1187-1196, 2024, ISSN: 2561-8148, DOI: <https://doi.org/10.5267/j.ijdns.2023.11.015>
- LEE, S.; DANANDEH MEHR, A.; MORIASI, D.; MIRCHI, A.: "Large-Scale Drought Forecasting in the U.S. Southern Plains Through a Hybrid Cluster-Based Wavelet-Machine Learning Approach", *Water Resources Research*, 61(11): e2024WR039744, 2025, ISSN: 0043-1397, DOI: <https://doi.org/10.1029/2024WR039744>
- LOUKAS, A.; GARROTE, L.; VASILIADES, L.: "Hydrological and Hydro-Meteorological Extremes and Related Risk and Uncertainty", *Water*, 13(3): 377, 2021, ISSN: 0043-1397, DOI: <https://doi.org/10.3390/w13030377>
- PAPARRIZOS, J.; GRAVANO, L.: "Fast and Accurate Time-Series Clustering", *ACM Transactions on Database Systems*, 42(2): 1-49, 2017, ISSN: 0362-5915, DOI: <https://doi.org/10.1145/3044711>
- PAPARRIZOS, J.; YANG, F.; LI, H.: "Bridging the Gap: A Decade Review of Time-Series Clustering Methods", *arXiv:2412.20582*, 2024, ISSN: 2331-8422, DOI: <https://doi.org/10.48550/arXiv.2412.20582>
- RADOVANOVIC, A.; LI, J.; MILANOVIC, J.V.; MILOSAVLJEVIC, N.; STORCHI, R.: "Application of Agglomerative Hierarchical Clustering for Clustering of Time Series Data", 2020, IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe): 640-644, 2020, ISSN: 2165-4824, DOI: <https://doi.org/10.1109/ISGT-Europe47291.2020.9248759>
- RODRIGUES, P.P.; GAMA, J.; PEDROSO, J.P.: "ODAC: Hierarchical Clustering of Time Series Data Streams", *Proceedings of the 2006 SIAM International Conference on Data Mining*: 499-503, 2006, ISBN: 978-0-89871-611-5, DOI: <https://doi.org/10.1137/1.9781611972764.48>
- SHAHAPURE, K.R.; NICHOLAS, C.: "Cluster Quality Analysis Using Silhouette Score", 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA): 747-748, 2020, ISSN: 2472-1573, DOI: <https://doi.org/10.1109/DSAA49011.2020.00096>
- SINGH, J.; SINGH, V.; OJHA, C.S.P.: "A Cluster-Based Data Assimilation Approach to Generate New Daily Gridded Time Series Precipitation Data in the Himalayan River Basins", *Water Resources Research*, 61(1): e2024WR037324, 2025, ISSN: 2165-4824, DOI: <https://doi.org/10.1029/2024WR037324>
- TAVENARD, R.; FAOUZI, J.; VANDEWIELE, G.; DIVO, F.; ANDROZ, G.; HOLTZ, C.; PAYNE, M.; YURCHAK, R.; RUßWURM, M.; KOLAR, K.; WOODS, E.: "Tslern, A Machine Learning Toolkit for Time Series Data", *Journal of Machine Learning Research*, 21: 1-6, 2020, ISSN: 1533-7928
- TIANO, D.; BONIFATI, A.; NG, R.: "Feature-driven Time Series Clustering", *OpenProceedings.org*, 2021, ISSN: 2367-2005, DOI: <https://doi.org/10.5441/002/EDBT.2021.33>

- WANG, X.; XU, Y.: "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index", *IOP Conference Series: Materials Science and Engineering*, 569(5): 052024, 2019, ISSN: 1873-0191, DOI: <https://doi.org/10.1088/1757-899X/569/5/052024>
- ZAIFOGLU, H.; AKINTUG, B.; YANMAZ, A.M.: "Regional Frequency Analysis of Precipitation Using Time Series Clustering Approaches", *Journal of Hydrologic Engineering*, 23(6): 05018007, 2018, ISSN: 1084-0699, DOI: [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001659](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001659)
- ZHANG, F.; BIEDERMAN, J.A.; DANNENBERG, M.P.; YAN, D.; REED, S.C.; SMITH, W.K.: "Five Decades of Observed Daily Precipitation Reveal Longer and More Variable Drought Events Across Much of the Western United States", *Geophysical Research Letters*, 48(7): e2020GL092293, 2021, ISSN: 1944-8007, DOI: <https://doi.org/10.1029/2020GL092293>
- ZHANG, Z.; LI, D.; ZHANG, Z.; DUFFIELD, N.: "Mining Spatiotemporal Mobility Patterns Using Improved Deep Time Series Clustering", *ISPRS International Journal of Geo-Information*, 13(11): 374, 2024, ISSN: 2220-9964, DOI: <https://doi.org/10.3390/ijgi13110374>