

Data driven and expert knowledge approach for climate zoning

Enfoque basado en datos y conocimiento experto para la zonificación climática

 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

*Corresponding author: Ariamna Crespo Lara, e-mail: ariamnacrespo@gmail.com

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.

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.

INTRODUCTION

The proliferation of hydrometeorological data, driven by advances in sensor and telemetry technologies, has established time series analysis as an indispensable tool in water resource management and climate studies (Bonaccorso y Peres, 2022; Loukas et al., 2021). Among these, precipitation time series represent a critical component of the hydrological cycle, whose accurate

characterization is essential for understanding extreme events, agricultural planning, and assessing the impacts of climate change (F. Zhang et al., 2021). In this context, time series clustering emerges as a powerful data mining technique for discovering spatiotemporal patterns, identifying homogeneous regions, and understanding the underlying structure of large volumes of precipitation data without the need for prior labels, thereby facilitating robust exploratory analysis (Duan et al., 2024; Singh et al., 2025).

Received: 19/11/2025

Accepted: 02/04/2026

Conflicts of interest: The authors declare that there are no conflicts of interest.

Author contributions: **Conceptualization:** Ariamna Crespo Lara, Yankier González Crespo. **Data curation:** Ariamna Crespo Lara, Yankier González Crespo, Julio C. Quintana Zaez. **Investigation:** Ariamna Crespo Lara, Yankier González Crespo. **Methodology:** Ariamna Crespo Lara, Yankier González Crespo, Oscar Brown Manrique. **Supervision:** Julio C. Quintana Zaez, Oscar Brown Manrique. **Validation:** Julio C. Quintana Zaez, Oscar Brown Manrique. **Writing - original draft:** Ariamna Crespo Lara, Yankier González Crespo. **Writing - review & editing:** Ariamna Crespo Lara, Yankier González Crespo, Julio C. Quintana Zaez, Oscar Brown Manrique.

The mention of trademarks of specific equipment, instruments or materials is for identification purposes, there being no promotional commitment in relation to them, neither by the authors nor by the publisher.



This is an open access article under terms of the license Creative Commons Attribution-NonCommercial (BY-NC 4.0).
<https://creativecommons.org/licenses/by-nc/4.0/>



The international scientific community has developed multiple approaches for time series clustering, which can be categorized into three main groups: methods based on temporal proximity that use measures like Dynamic Time Warping (DTW) to capture similarities in the shape of the series (Fransiska et al., 2024; Lee et al., 2025); model-based methods that employ parameters from autoregressive or Hidden Markov Models (Hern ndez et al., 2022; Ghassempour et al., 2014); and feature-based methods that transform the series into vectors of precomputed features (Arulraj et al., 2023; Tiano et al., 2021). More recently, advanced techniques have emerged that incorporate deep learning architectures, such as autoencoders and recurrent neural networks, which enable learning non-linear representations of time series (Paparrizos et al., 2024). In parallel, specialized computational tools like tslearn have been developed to facilitate the implementation and comparison of these algorithms (Tavenard et al., 2020).

In the specific domain of precipitation, clustering has demonstrated its utility in diverse applications. Studies have successfully employed techniques such as k-Shape for spatiotemporal mobility analysis (Paparrizos & Gravano, 2017), hierarchical clustering with correlation measures for regional frequency analysis (Zaifoglu et al., 2018), and feature-based methods for identifying rainfall pattern hietographs (Zhang et al., 2024). In particular, research in Indonesia has used ensemble distance for rainfall data clustering, while in India, combined approaches of hierarchical clustering and wavelet transform have been applied to analyze seasonal precipitation trends (Hendrawati et al., 2024). Those international research highlight the potential of these techniques for regional hydroclimatic analysis.

However, in the national context, specifically in Cuba, the application of these modern clustering methodologies to precipitation time series represents a largely unexplored research niche. Existing studies have predominantly focused on traditional statistical analyses or conventional climatic approaches, with limited implementation of advanced data mining and machine learning techniques for the spatiotemporal characterization of rainfall regimes (Garcia Moya et al., 2024). This methodological gap represents a significant opportunity to develop research that adapts and optimizes these algorithms to the particularities of the Cuban climate, characterized by its temporal and spatial variability and the influence of complex meteorological systems.

To bridge this methodological gap and leverage the identified opportunities, this article proposes a clustering framework for precipitation time series based on the extraction of precomputed features and the application of the Hierarchical Clustering Algorithm (HCA). This methodology combines the computational efficiency of HCA with the informational richness of a set of features designed to capture key aspects of the rainfall regime, such as its intensity, variability, seasonality, and extreme behavior. The main objective is to develop an accessible and scalable methodology that enables the identification of climatically homogeneous regions in Cuba with high

precision, thereby contributing to the improvement of prediction models, drought and flood risk management, and water resource planning in a context of climate change. The proposal is validated using long-term precipitation data, demonstrating its effectiveness in the coherent grouping of stations and its utility as a tool for regional hydroclimatic analysis.

MATERIALS AND METHODS

This section outlines the methodology employed in the paper, detailing the research design and its practical implementation. The selection of specific methods and techniques is justified to ensure the study can be replicated by other researchers.

Context and dataset

Precipitation data from the experimental stations in Ciego de  vila were obtained from CITMA upon request by the Head of the Water Resources Scientific Group at the University of Ciego de  vila. The dataset covers the period from 1960 to 2013 and was compiled through specialist surveys. The database is provided in an Excel file and includes records from 107 experimental stations distributed across the province. For each station, the data are organized by daily precipitation, month, year, station ID_Code, and municipality. Precipitation values are expressed in millimeters (mm).

As tools the Python Pandas library was used to read and process the data, offering a range of methods that enable efficient data extraction through streamlined code. During the data loading process several experimental stations were discarded due to significant data recording errors, resulting in a final set of 17 stations with valid data for the study. The information for all station staked into account can be requested in the supplementary materials.

Methodology

The methodology applied in this work consisted of five stages that formed the workflow of the proposed process for clustering time series data using HCA (Figure 1). These stages were as follows: data loading and normalization, feature generation, data analysis, algorithm tuning, metric evaluation, and cluster description:

- Load data - normalization: Following rigorous data curation that excluded stations with extensive missing data, metadata flaws, or low reliability scores, the refined dataset underwent MinMaxScaler normalization to ensure equitable feature contribution for clustering.
- Features generation: Feature engineering was conducted through a hybrid approach combining computational and domain expertise. Using tsfresh (Christ et al., 2018), sliding window segmentation was applied to extract temporal patterns, while statistical features (mean, variance, trends) were computed from each window.

These data-driven features were enriched with expert-defined hydrological indicators, creating a multidimensional feature set that captures both algorithmic patterns and physical significance for subsequent clustering analysis.

- **Data analysis:** A correlation analysis identified key hydrological drivers and potential multicollinearity, establishing the foundational interpretability framework for subsequent cluster characterization.
- **Algorithm fitting:** Hierarchical clustering with Ward's linkage was selected to model the station relationships. This algorithm excels at generating a clear dendrogram structure that visually represents the nested hierarchy of stations. Ward's method minimizes within-cluster variance, producing compact, spherical clusters that offer superior explanatory value by directly illustrating the aggregation logic and relative similarities between stations at different hierarchical levels.
- **Metrics evaluation:** Cluster quality was assessed through a triple-metric framework: Silhouette Score (measuring cohesion and separation), Calinski-Harabasz Index (evaluating between-cluster dispersion), and Davies-Bouldin Index (quantifying cluster distinctness). This multi-faceted validation approach provides complementary perspectives on partition optimality, ensuring the selected solution demonstrates both mathematical robustness and hydrological relevance beyond what any single metric could confirm.
- **Clusters description:** The final cluster characterization employs a comprehensive five-pillar analytical framework: inferential statistical testing to identify significant inter-cluster differences, vulnerability assessment to evaluate climate risk exposure, temporal trend analysis to uncover precipitation patterns, regional mapping to contextualize spatial distributions, and expert validation to ground algorithmic findings in domain knowledge. This multidimensional interpretation transforms abstract clusters into actionable hydrological typologies.

Algorithm

This study combined feature engineering with Ward's linkage to form compact and interpretable clusters, mitigating dimensional complexity and revealing distinct

dynamic regimes through a multiresolution visualization of the data:

- **Hierarchical Clustering:** uncovers latent structures within high-dimensional time series data, leveraging its inherent adaptability to this complex domain. The methodological framework strategically combines feature engineering with Ward's linkage to address key challenges. By transforming raw series into a curated set of salient features (using tsfresh and feature engineering), the algorithm mitigates the high-dimensionality problem and enhances computational efficiency.
- **Ward's Linkage Method:** minimizes intra-cluster variance to form compact, spherical clusters that facilitate interpretation (Duan et al., 2024; Radovanovic et al., 2020; Rodrigues et al., 2006). This approach replaces the need for direct similarity measures on raw data, such as DTW, with a robust distance metric in the domain of engineered features.
- **Resulting Dendrogram:** provides a multiresolution visualization of the data hierarchy, revealing both broad patterns and detailed anomalies. Consequently, this workflow facilitates the discovery of meaningful clusters corresponding to distinct dynamic regimes within the temporal data, balancing analytical rigor with the extraction of actionable insights.

Internal validity metrics

The clustering solution was quantitatively evaluated using a triad of internal validation metrics: the Silhouette Coefficient, the Calinski-Harabasz Index, and the Davies-Bouldin Index:

- **Silhouette Coefficient:** evaluates how well each object lies within its cluster by measuring cohesion and separation, with values close to 1 indicating a superior model.
- **Calinski-Harabasz Index:** also known as the Variance Ratio Criterion, evaluates cluster quality by comparing the ratio between inter-cluster dispersion and intra-cluster dispersion, where a higher score signifies better-defined clusters (Shahapure y Nicholas, 2020; Wang y Xu, 2019).
- **Davies-Bouldin Index:** quantifies the average similarity between each cluster and its most similar cluster, where an optimal solution achieves the lowest possible value, indicating maximum distinctness among clusters (Ashari et al., 2022).



Figure 1. Pipeline of process for clustering time series

Together, this multifaceted validation framework provides a robust and non-redundant assessment of partition quality, ensuring that the results are both statistically sound and structurally meaningful.

RESULTS AND DISCUSSION

The application of the proposed methodology yielded significant insights into the precipitation regimes of Ciego de  vila. The initial phase involved a comprehensive analysis of the generated features to understand their relationships and relevance. Subsequently, the Hierarchical Clustering Algorithm was fitted, identifying the optimal partition of the data. Finally, the resulting clusters were rigorously interpreted and discussed by integrating the quantitative findings with expert domain knowledge, translating the algorithmic output into actionable climatic characterization.

Feature correlation analysis

This research developed a feature engineering strategy that integrates three distinct feature groups to capture comprehensive time series characteristics. The features groups are show in the Table 1. The first group comprises automated features generated using the tsfresh Python library, which systematically extracts a wide range of temporal patterns. The second group incorporates domain-specific features derived from expert knowledge, ensuring alignment with theoretical foundations. The third group employs a rolling window approach with 30 days of lags to capture temporal dependencies and evolution patterns.

Collectively, these feature sets compress multidimensional time series information into a structured representation suitable for clustering algorithms. Additionally, a correlation analysis was conducted

to identify key relationships, providing foundational insights for subsequent cluster interpretation and characterization studies.

It is noteworthy that the expert-derived features were specifically designed to capture climate dynamics relevant to the Caribbean region. Variables such as `dry_season_intensity`, `hurricane_season_intensity`, `extreme_event_frequency`, and `interannual_trend` were incorporated based on domain expertise regarding their significance in characterizing rainfall patterns.

A Spearman correlation analysis revealed several key relationships: `dry_season_intensity` and `total_precipitation` exhibited a moderate positive correlation ($\rho = 0.544$), while `rainfall_regime_contrast` and `hurricane_season_intensity` showed a strong positive correlation ($\rho = 0.861$). Conversely, `extreme_event_frequency` and `extreme_event_intensity` demonstrated a strong negative correlation ($\rho = -0.761$). These relationships provide crucial insights for interpreting the resulting clusters and identifying distinct time series regimes.

The Figure 2 illustrates the correlation structure among all features, revealing several distinct groups of interrelated variables. Understanding these relationships is particularly valuable for agricultural applications, as they can inform farmers' decision-making processes and support optimization of farming practices in response to climatic variability.

Clustering analysis

In the Figure 3, the hierarchical clustering analysis identified six distinct groups organized within a well-defined hierarchical structure. This grouping reveals different seasonal regimes that are critical for the agricultural sector, as they enable the characterization of specific climate patterns and their impact on crop cycles.

Table 1. Features groups for time series analyzing

Tfresh	Experts	Rolling window
<code>fft_coefficient_attr__real__coeff_0</code>	<code>dry_season_intensity</code>	<code>window_variability</code>
<code>fft_aggregated_aggtype__centroid</code>	<code>dry_season_std</code>	<code>window_stability</code>
<code>number_peaks_n_3</code>	<code>rainy_season_intensity</code>	<code>window_autocorrelation</code>
<code>change_quantiles_isabs_qh_0.9_ql_0.1</code>	<code>rainfall_regime_contrast</code>	<code>window_consistency</code>
<code>change_quantiles_isabs_qh_0.8_ql_0.2'</code>	<code>hurricane_season_intensity</code>	<code>n_valid_windows</code>
<code>abs_energy'</code>	<code>hurricane_season_variability</code>	
<code>mean_abs_change</code>	<code>extreme_event_frequency</code>	
<code>autocorrelation_lag_1</code>	<code>extreme_event_intensity</code>	
<code>autocorrelation_lag_7</code>	<code>interannual_variability</code>	
<code>autocorrelation_lag_30</code>	<code>interannual_trend</code>	
<code>quantile_q_0.95</code>	<code>total_precipitation</code>	
<code>quantile_q_0.99</code>	<code>mean_precipitation</code>	
<code>c3_lag_1</code>	<code>std_precipitation</code>	
<code>cid_ce_normalize_True</code>	<code>cv_precipitation</code>	
	<code>zero_ratio</code>	

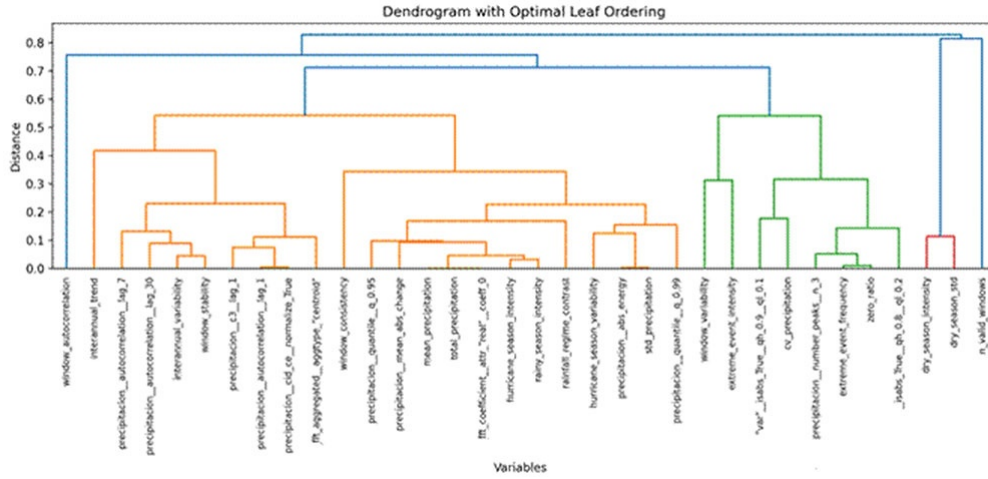


Figure 2. Feature correlation dendrogram for selected variables

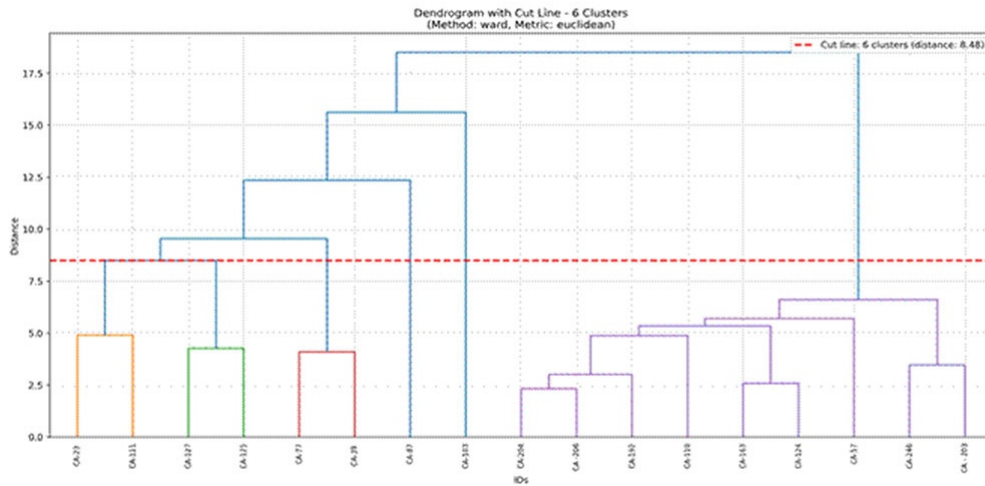


Figure 3. Hierarchical clustering of the stations

The determination of the optimal number of clusters was based on validation metrics such as Silhouette, Calinski-Harabasz, and Davies-Bouldin, which consistently confirmed that six clusters represent the most optimal configuration, as shown in Figure 4. This is particularly relevant because, although other groupings might be visually suggested in the dendrogram, objective metrics validate that this segmentation most effectively captures seasonal variations that directly affect agricultural

planning, crop selection, and climate risk management in the region.

In the Figure 4, the clustering validation metrics yielded moderate performance, with a Silhouette score of 0.34 indicating weak structure separation. A Calinski-Harabasz index of 9.0 suggests limited between-cluster differentiation, while the Davies-Bouldin score of 0.6 reflects reasonable cluster compactness.

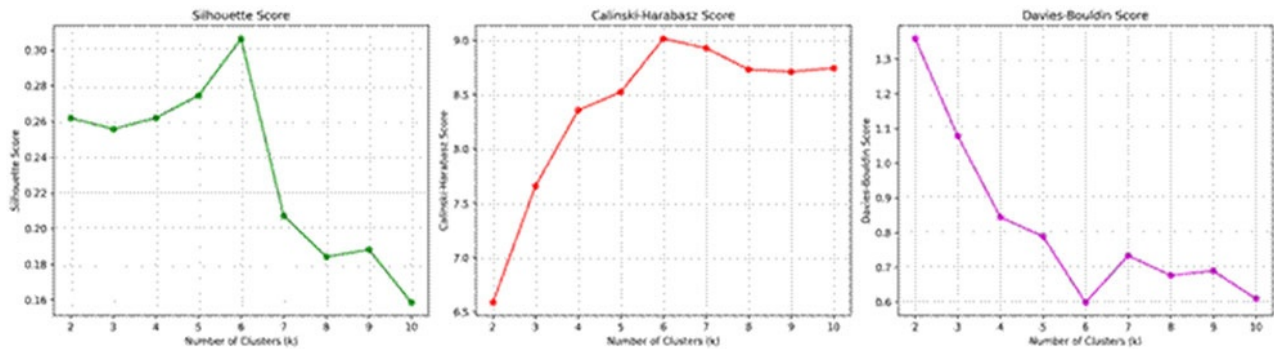


Figure 4. Silhouette Score, Calinski-Harabasz Score and Davies-Bouldin Score results

These collective results imply overlapping clusters with subdued but detectable patterns in the data structure.

Characterization of Clusters

Hierarchical clustering analysis identified six distinct rainfall regimes across the Caribbean region, each characterized by specific seasonal variability patterns. Cluster 5 emerged as the most intense, exhibiting maximum values in precipitation (FFT real component), dry season variability (dry_season_std), and hurricane season intensity, suggesting a high-energy climatic regime. Conversely, Cluster 0 showed minimum values across these variables, indicating a more stable and less intense rainfall pattern. These extremes represent opposite poles of regional climate variability, with direct implications for seasonal agricultural planning and water resource management strategies tailored to each regime's specific characteristics.

Inferential analysis confirmed statistically significant differences ($p < 0.01$) in critical climate variables, including precipitation peak frequency, interannual trends, and hurricane intensity. Post-hoc testing revealed that Cluster 4 exhibits significantly more precipitation peaks compared to Clusters 0, 1, 2, 3, and 5, while Cluster 1 demonstrates a substantially stronger interannual trend than Cluster 5. Vulnerability analysis identified Cluster 1 as the highest-risk group, combining elevated extreme event frequency with pronounced dry season intensity, creating critical challenges for rainfed agriculture systems that depend on predictable seasonal patterns.

The geographical distribution of clusters reveals distinct spatial patterns, with Cluster 0 predominant in 1° de Enero, Cluster 1 in Chambas, and Cluster 5 in Venezuela, suggesting natural climate zonation. This segmentation enables differentiated agricultural strategies: regions like Chambas (Cluster 1) require drought-resistant crops and supplemental irrigation systems, while areas such as Venezuela (Cluster 5) need preparedness protocols for high climate variability and hurricane intensity. This rainfall regime characterization provides a scientific foundation for crop selection, planting calendar optimization, and water management system design adapted to each identified climate pattern, ultimately enhancing agricultural resilience across the Caribbean basin.

CONCLUSIONS

As conclusion, this research successfully demonstrates the efficacy of a hybrid clustering framework, combining data-driven features with domain-driven expertise, to decipher the complexity of rainfall regimes in Ciego de Ávila.

The methodology not only identified six statistically distinct pluviometric regimes-ranging from stable to high-energy systems with marked cyclonic influence-but also effectively translates these findings into agricultural practice.

By characterizing high-vulnerability clusters, it provides an objective tool for decision-making, enabling the prioritization of interventions such as drought-resistant crops or irrigation systems in specific zones.

This contribution transcends academia, laying the groundwork for precise and scalable climate zoning at a national level. This is fundamental for designing resilient agricultural and water policies, optimizing planting calendars, and ultimately strengthening Cuba's food security in the face of climate variability.

REFERENCES

- 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>
- 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/jaic.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.: "Spatio-temporal 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>