

Article

Microclimate Zoning Based on Double Clustering Method for Humid Climates with Altitudinal Gradient Variations: A Case Study in Colombia

Cristian Mejía-Parada ¹, Viviana Mora-Ruiz ¹, Jonathan Soto-Paz ¹, Brayan A. Parra-Orobio ²
and Shady Attia ^{3,*}

¹ Research Group on Threats, Vulnerability and Risks to Natural Phenomena, Faculty of Engineering, Research and Development University (UDI), Bucaramanga 680001, Colombia; cmejia5@udi.edu.co (C.M.-P.); jmora13@udi.edu.co (V.M.-R.); jsoto3@udi.edu.co (J.S.-P.)

² GE&TES Research Group, Department of Environmental and Health Sciences, Popular University of Cesar, Aguachica 205010, Colombia; baparra@unicesar.edu.co

³ Sustainable Building Design Lab, Department UEE, Faculty of Applied Sciences, University of Liège, 4000 Liège, Belgium

* Correspondence: shady.attia@uliege.be

Abstract: Climatic classification is essential for evaluating climate parameters that allow sustainable urban planning and resource management in countries with difficult access to meteorological information. Clustering methods are on trend to identify climate zoning; however, for microclimate, it is necessary to apply a double clustering technique to reduce the variability from former clusters. This research raised a climate classification of an emerging country, Colombia, using climatological models based on freely available satellite image data. A double clustering approach was applied, including climatological, geographic, and topographic patterns. The research was divided into four stages, covering the collection and selection of climatic and geographic data, and multivariate statistical analysis including principal components analysis (PCA) and agglomerative hierarchical clustering (HAC). The meteorological data were from reliable sources from the Center for Hydrometeorology and Remote Sensing (CHRS) and the National Renewable Energy Laboratory (NREL). The results showed that a total of 17 microclimates distributed across the country were identified, each characterized by a different threshold of the climatic and geographic factors evaluated. This subdivision provided a detailed understanding of local climatic conditions, especially in the mountain chains of the Andes.

Keywords: climate zoning; clustering; microclimate; principal component analysis; K-means



Citation: Mejía-Parada, C.; Mora-Ruiz, V.; Soto-Paz, J.; Parra-Orobio, B.A.; Attia, S. Microclimate Zoning Based on Double Clustering Method for Humid Climates with Altitudinal Gradient Variations: A Case Study in Colombia. *Atmosphere* **2024**, *15*, 709. <https://doi.org/10.3390/atmos15060709>

Academic Editor: Ferdinando Salata

Received: 15 May 2024

Revised: 7 June 2024

Accepted: 10 June 2024

Published: 14 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Conceptualization

A detailed understanding of the spatial distribution of climate is essential for effective land-use planning and management [1]. This accurate climate classification provides a framework for anticipating and mitigating climate impacts on various human activities and the natural environment. Also, it serves as a fundamental basis for developing climate change adaptation policies and strategies [2]. In this context, climate data from satellite imagery emerge as an invaluable tool in contrast to meteorological terrain data. The wide spatial distribution between weather stations and the lack of continuous data usually generates more significant uncertainty, translating into reduced accuracy in describing more local environments [3]. On the other hand, satellite-based technologies offer complex, validated statistical models that provide a broader and more detailed perspective of climate patterns, allowing a better understanding of microclimates and geographic variations at regional and local scales.

The integration of satellite data into the climate characterization of a region provides crucial information for the identification of vulnerable areas and the implementation of specific adaptation measures. In the field of urban design and sustainable building construction, the ability to analyze climate patterns at the local and regional levels allows for more precise planning of urban infrastructure, the orientation of buildings to maximize solar energy capture, or the implementation of passive heating and cooling strategies [4]. The integration of satellite data into the climate characterization of a region contributes to improving the assessment of thermal comfort of people and the potential energy demand and greenhouse gas emissions associated with heating and cooling systems [5]. The lack of an accurate climate classification can lead to inefficient or suboptimal policies, as decision-making is based on incomplete or inaccurate data. This can result in inappropriate use of resources, increased costs, and reduced adaptive capacity to climate change impacts [6]. Therefore, the availability of accurate and up-to-date climate data is crucial to inform policy and strategic decisions that promote sustainability and resilience in resource management and energy consumption.

In most emerging countries, there are significant deficiencies in the collection of climate information. This situation is closely linked to problems related to continuity and a lack of data at obsolete weather stations. This problem is compounded by the fact that climate observation points in the field often lack adequate distribution and coverage at the local or regional levels. In addition, it is common for these countries to lack climate models based on their satellite data, which makes it even more difficult to assess the climate of a region under these conditions. From this perspective, the implementation of open-access tools, such as the typical meteorological year (TMY) model produced by the National Renewable Energy Laboratory (NREL) [7], and precipitation models, such as PERSIANN Dynamic Infrared-Rain Rate (PDIR-Now) from the Center for Hydrometeorology and Remote Sensing (CHRS) [8], emerge as reliable and up-to-date satellite-based tools for the detailed climate classification of a study region.

1.2. Climate Zoning Background

A climate classification is a system that organizes and categorizes different types of climates based on common climatic patterns. Traditionally, this zoning has been carried out using qualitative and subjective approaches based on observations and experience [9]. One of the most widely recognized climate classification systems is the Köppen climate classification system, developed by the climatologist Wladimir Köppen [10]. This classification is based on the observation of vegetation types associated with different climates, as well as temperature and precipitation regimes. The zoning is divided into five major groups: A (tropical climates), B (dry climates), C (temperate climates), D (cold climates), and E (ice climates). Each of these groups is further subdivided into specific categories, allowing for a detailed description of the climatic characteristics of a given region. The Köppen climate classification is constantly updated [11]. It is widely used in various disciplines, from agriculture to architecture, as it provides valuable information for understanding the climate of a region. However, other authors [6] have indicated that it is not as suitable for bioclimatic designs since it does not consider aspects such as solar radiation, wind, or relative humidity, which are fundamental parameters for designing efficient and sustainable built environments [9].

Another commonly used method in climatic classification is the degree day classification method. This methodology focuses on the accumulation of temperatures above or below certain thresholds, which allows the assessment of crop growth potential, energy consumption, or heating and cooling demand [12]. This method is based on calculating the cumulative sum of temperatures above certain specified thresholds over a given period. It provides detailed information on local thermal conditions, allowing for more accurate planning of building orientation, material selection, and implementing passive cooling and heating strategies [13]. However, one limitation is its exclusive focus on the accumulation of temperatures based on a benchmark evaluated by the model, which may not fully capture

the complexity of climate systems. This methodology tends to simplify climate variability by focusing only on temperature, leaving out other important climatic factors, such as precipitation, relative humidity, wind, or solar radiation. In addition, degree day may not be accurate in areas with extreme or changing climates. This is because the accumulation of temperatures may not accurately reflect actual climatic conditions. Therefore, it is necessary to complement this tool with broader, multidimensional approaches to fully understand a region's climate variability.

Within climate classification methodologies, some countries have adopted climate analysis using the Givoni psychrometric diagram as a climate characterization tool applied to bioclimatic zoning classification [14]. This technique presents a correlation of fundamental environmental variables, such as dry bulb temperature and relative humidity, facilitating the understanding of the climatic conditions of a specific region. This classification method is based on a series of zones that suggest design strategies to be implemented according to climatic ranges [15]. This makes it possible to visually identify the most appropriate passive strategies to achieve indoor thermal comfort without resorting to mechanical heating or cooling systems. Among the advantages of this approach are its ability to provide quick and general recommendations on bioclimatic strategies and its accessibility and ease of use. However, it has limitations, such as the simplicity of the climatic data used, the subjectivity of thermal comfort, and the lack of consideration of specific local factors.

On the other hand, local climate zoning (LCZ) is a standardized methodology for classifying land cover and its thermal, climatic, and surface emission properties in urban and rural areas [16]. This tool is used in the context of climate change, such as sustainable urban planning, urban heat island mitigation, air quality modeling, and environmental risk assessment [17]. The development of LCZs involves a systematic approach combining geospatial data analysis, field observations, and remote sensing techniques. Land-use patterns and urban morphology are identified and mapped, considering features such as building height, building density, soil permeability, vegetation cover, and thermal properties of surface materials [16]. These characteristics are grouped into standardized zones, each with specific thermal, climatic, and surface emission properties. This rigorous approach allows for a detailed characterization of the urban environment and its interaction with the local climate. It facilitates the development of climate-smart urban planning strategies and the integration of LCZs into numerical climate and air quality simulation models [18]. In addition, this tool provides a consistent and comparative framework for studying urban climate effects in different cities and regions [18]. Although it provides a local characterization, the complexity of this methodology and its scalability is very low, making it difficult to evaluate large extension areas

In recent decades, clustering techniques have gained popularity due to their quantitative and objective approach [4,19,20]. This methodology for climate zoning uses unsupervised clustering algorithms to identify patterns and similarities in large sets of climate data [21]. These approaches group regions with similar climatic characteristics into clusters or distinctive climatic zones. One of the main advantages of this methodology is its ability to process large amounts of multivariate data objectively and reproducibly. This enables the evaluation of multiple climatic and geographic parameters, allowing a better climatic understanding of the evaluated region. In addition, it allows for continuous updating of climate zones as new data become available. This makes it a valuable tool for natural resource analysis and management, urban planning, agriculture, and biodiversity conservation. Nadarajah et al. [19] implemented a bioclimatic classification for building energy efficiency using clustering techniques for urban planning design in Sri Lanka, exhibiting three distinct bioclimatic zones and bringing forth specific information regarding each bioclimatic zone. Likewise, Praene et al. [20] performed a hierarchical clustering on PCA in Madagascar, where results showed three climate zones reflecting the territory's reality. Recently, Davinson et al. [3] proposed a new microclimate method, including a double clustering technique with K-means partitioning that allowed the identification of local zones in Reunion Island. Their findings aim to contribute to the understanding of climate at

a local level and can be used as a guide to the adaptation of climate-responsive strategies at a smaller scale. However, there are not many studies with the last-mentioned methodology and few more that include the evaluation of altitudinal gradient as a climatological variable.

Two of the most widely used clustering techniques in climate zoning are the K-means method and hierarchical clustering, especially Ward's approach with Euclidean distance [19,22,23]. The K-means method is an unsupervised partitioning algorithm that divides the data into a predetermined number of clusters (k) [24]. It randomly assigns each data point (region) to one of the k clusters. It then iteratively calculates the centroids of each cluster and reassigns the data points to the cluster with the closest centroid until the groups are stabilized [25]. The main advantage of K-means is its computational efficiency, which makes it suitable for processing large climate datasets. However, it requires specifying the number of clusters beforehand, which can be a limitation if the underlying structure of the data is not known [26].

On the other hand, hierarchical clustering constructs a hierarchy of nested clusters, either in an agglomerative (bottom-up) or divisive (top-down) manner. One of the most used approaches is the Ward method, an agglomerative method that minimizes the sum of squared distances within clusters, using Euclidean distance as the similarity measure [22]. These algorithms have the advantage of not requiring the number of clusters beforehand, as they produce a complete hierarchy of clusters. This makes it particularly useful when the underlying structure of the climate data is not known [19]. In addition, using Euclidean distance as a similarity measure makes it intuitive and easy to interpret. However, this method can be computationally expensive for large climate datasets, and the choice of the cutoff level in the cluster hierarchy can introduce some subjectivity. In addition, it can be sensitive to outliers in the data.

1.3. Aim and Contribution of This Study

Few studies have performed a double clustering climatic classification, especially in emerging countries. Therefore, this research contributes to the definition of microclimate zoning based on open-access satellite data. This classification process facilitates the implementation of policies for energy efficiency and adapting the built environment to climate change, seeking the design of sustainable and resilient communities [9]. Therefore, it is necessary to use local-level zonings to carry out this type of analysis. In addition, the use of databases without uncertainty is required [27]. Although weather stations offer greater accuracy, their wider spatial distribution and problems with data continuity limit their scalability at both local and regional levels. Also, these regions uncommonly had access to their climatology models from satellite imagery. These limitations in the climatic evaluation of these regions can be overcome with the use of open-access, high-resolution statistics models based on satellite images. For the present research, TMY and PDIR-Now climatological models obtained localized information every 4.4 km per pixel. A cluster analysis and local climatic characterization, validated by various statistical methodologies, such as silhouette coefficient and WCSS, was performed in an emerging country geographically located in a tropical region influenced by the altitudinal gradient.

The implemented methodology benefits designers, urban planners, and resource managers by allowing the spatial identification and average characteristics of various climatic typologies in the evaluated territory. This study provides a deeper understanding of climatic behavior, which is crucial in assessing projects across multiple study areas. Also, this methodology can be replicated in different climatic contexts using open data from the NREL and CHRS, which is especially relevant for emerging countries without accurate meteorological characteristics or access to such data. The zoning provides an updated characterization based on data from the last 20 years, representing an updated classification from other climate zoning in the region. This research work reflects its contribution in the following points:

- I. Introduce, using updated TMY and PDIR-Now, open-access climate files based on high-resolution satellite information (4.4 km pixel).

- II. Perform a double clustering for global and local climate zoning for an emerging country located in a tropical region influenced by altitudinal gradient based on seven climate parameters and three geographical parameters.
- III. A new climate classification based on multivariate analysis has been established for Colombia.

The research is divided into four sections. The first section conceptualizes the problem and highlights the need for an updated local climate characterization in developing countries for territorial management. The second section describes the methodology developed, which includes a global and local clustering process for a microanalysis of the regional climate. This process involves dimensionality reduction using principal component analysis (PCA) to relate climate and geographical parameters, followed by a multivariate cluster-based analysis. The third section presents the results of the global and local analysis, identifying the climatic and geographic characteristics of each cluster. This classification is compared with other existing classifications, highlighting similarities, differences, and the possibility of complementing or clearly differentiating them. Finally, a discussion is presented, highlighting the importance of the proposed new climate zoning in the current resource management context for transforming urban environments into more sustainable communities.

2. Existing Climate Zoning of Colombia

2.1. The Study Area

Colombia, situated in the northeast of South America, spans an area of 1,141,748 km² and boasts access to Pacific and Atlantic coasts. Geographically, it ranges from approximately 12° N to 4° S latitude and 68° W to 78° W longitude, positioning it near the equatorial line typically associated with tropical and humid climates. Nevertheless, the region is profoundly influenced by the Andes Mountains' altitudinal gradient, which traverses the country from south to center, resulting in a diverse topography ranging from sea level to 5710 m above sea level. These geographic and topographic characteristics profoundly impact the climate, fostering various ecosystems and climatic conditions [28]. Moreover, Colombia's population, estimated at around 52.22 million, is predominantly concentrated in the northern regions and within the Andes Mountain range, as depicted in Figure 1. This distribution underscores the establishment of urban areas across varying climatic conditions dictated by geographic positioning and climate parameters. Consequently, this presents a significant challenge in formulating guidelines for sustainable urban development, considering the distinct requirements associated with each climate type and topography prevalent in the country.

2.2. Climate Zoning of the Region

Colombia does not follow the classic seasonal pattern of spring, summer, autumn, and winter. Instead, it experiences climatic seasons marked by periods of drought and rainfall. This climatic behavior varies significantly from one region to another due to topographic diversity and proximity to the equator. Depending on these factors, each area may experience one or two rainy seasons throughout the year. In addition, the amount of accumulated precipitation varies considerably across the country. In the northern regions, annual rainfall is low, with values not exceeding 400 mm in some areas. However, areas such as the West Coast experience extremely high precipitation levels, reaching up to 13,000 mm annually [28]. This disparity in climatic conditions significantly impacts demographic and environmental aspects in each region of the country.

Over time, various climate classifications have been developed to understand and define Colombia's climatic behavior. Table 1 presents the maps produced by different authors and governmental entities, each based on different climatic parameters and analysis approaches. These classifications provide essential tools for understanding the country's climate variability and are fundamental for natural resource planning, management, and sustainable development decision-making.

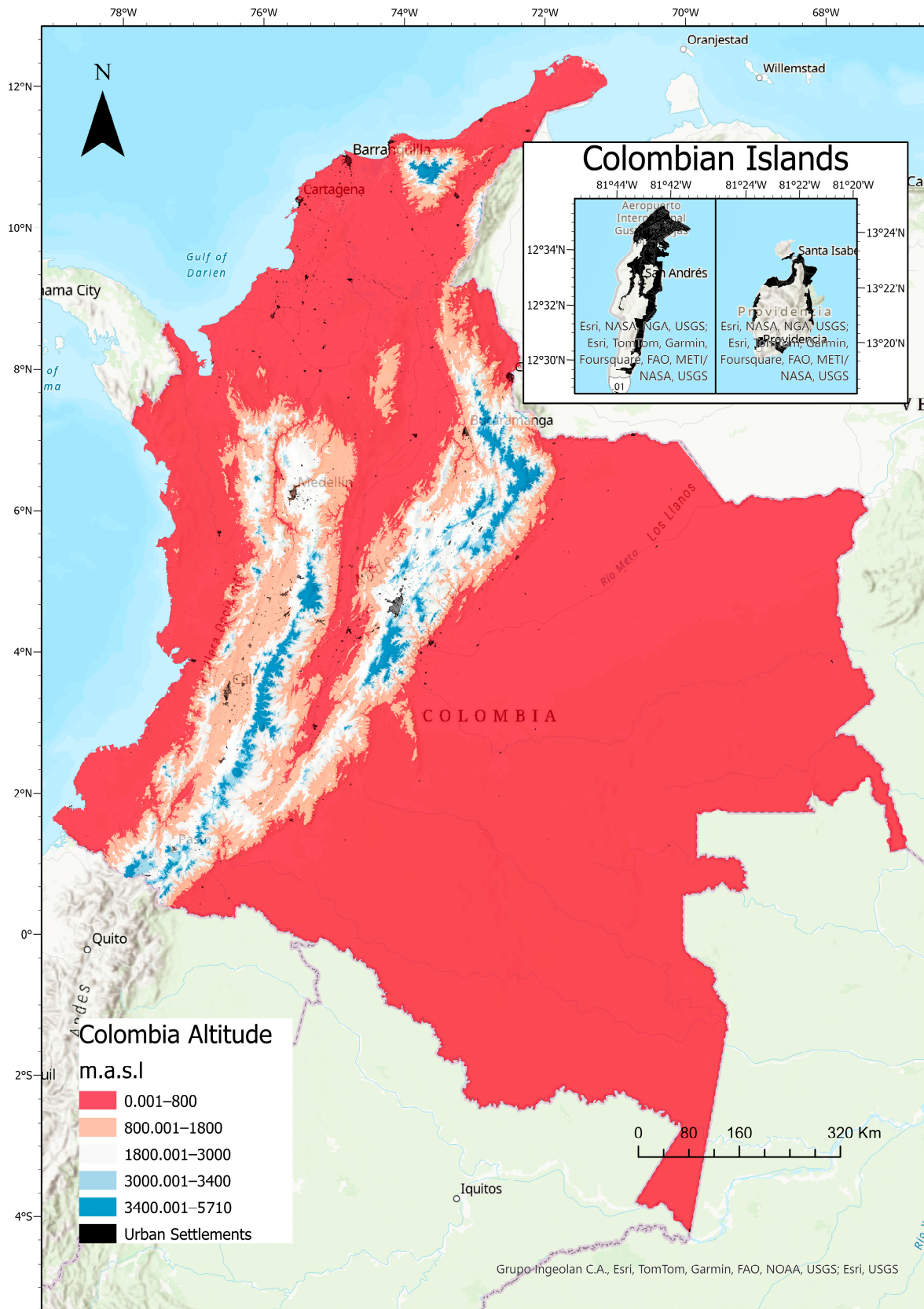
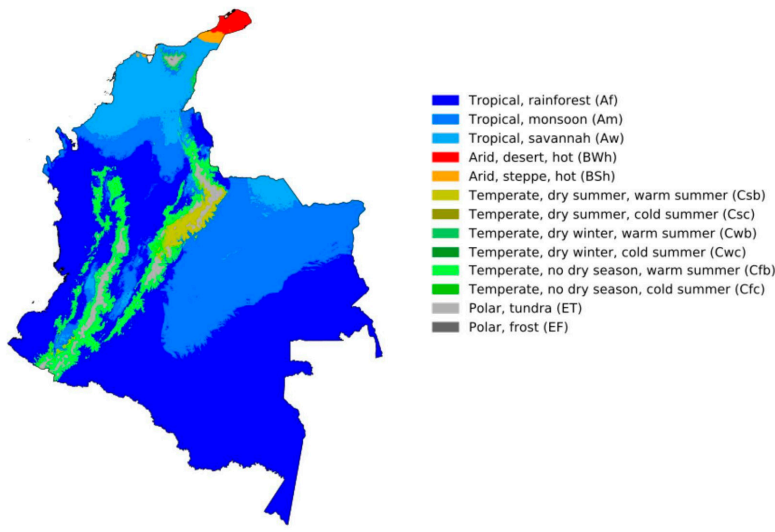


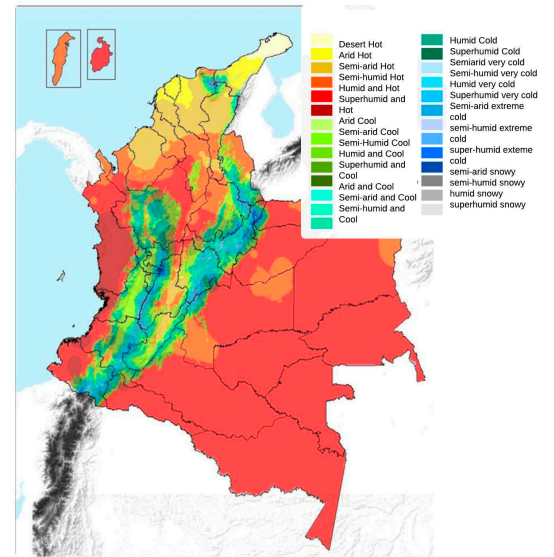
Figure 1. DEM (digital elevation model) of Colombia with the spatial distribution of urban settlements.

Table 1. Climate classification for Colombia.



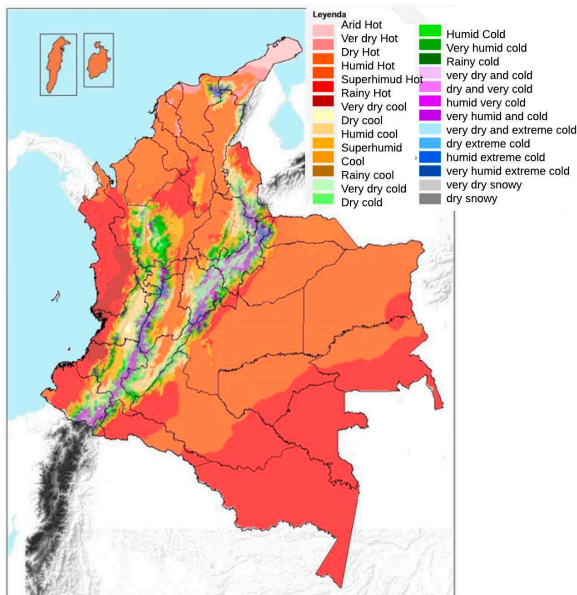
Köppen–Geiger Classification [11]

Use: climatology, agriculture, geography, hydrology.



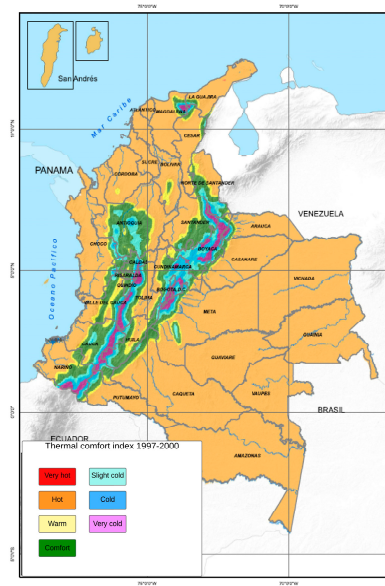
Caldas–Lang Classification [28]

Use: agriculture, climatology, biodiversity.



Holdridge Classification [28]

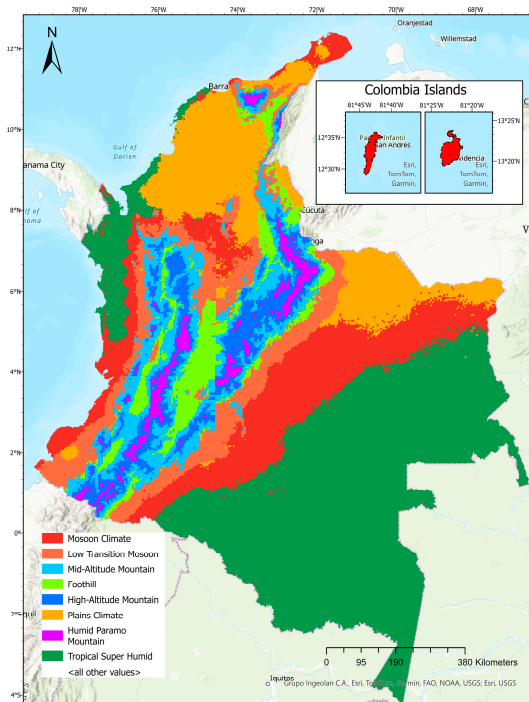
Use: agriculture, ecology, biology, land-use planning.



Thermal Comfort Classification [29]

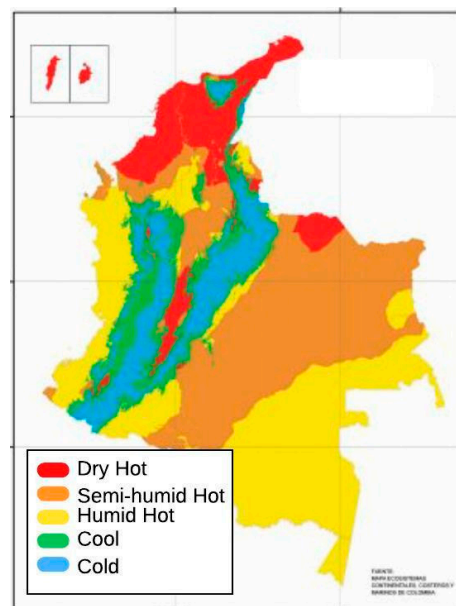
Use: thermal comfort evaluation.

Table 1. Cont.



Bioclimatic design zoning [4]

Use: architecture and building design.



Climate Zoning [30]

Use: architecture and building recommendations.

3. Materials and Methods

The present research is structured in four phases and designed to obtain a detailed climatic classification of a study region, as shown in Figure 2. Each phase addresses a critical stage of the data analysis process, contributing to achieving the research objectives.

In the first phase, weather and geographical data were collected and selected. Subsequently, in the second stage, a multivariate statistical analysis consisting of three fundamental steps was carried out. First, a principal component analysis (PCA) focused on dimensionality reduction was implemented, followed by extracting significant factors using an elbow testing technique. Then, hierarchical agglomerative clustering (HAC) determined the optimal number of clusters. Finally, a cluster analysis was performed using the K-means method to consolidate the identified clusters.

Phase three was based on the information generated in the first cluster analysis. This stage aimed to identify subgroups to obtain more detailed microclimatic details specific to the study region. The statistics processing performed in the second stage was executed again to develop this stage. Finally, the fourth phase focused on the spatial distribution of the results obtained in the previous phases. Geographic Information Systems (GIS) tools were used to map the identified clusters and subgroups. This process allowed a visual identification of the different climatic regions and provided a cartographic representation of the classifications obtained.

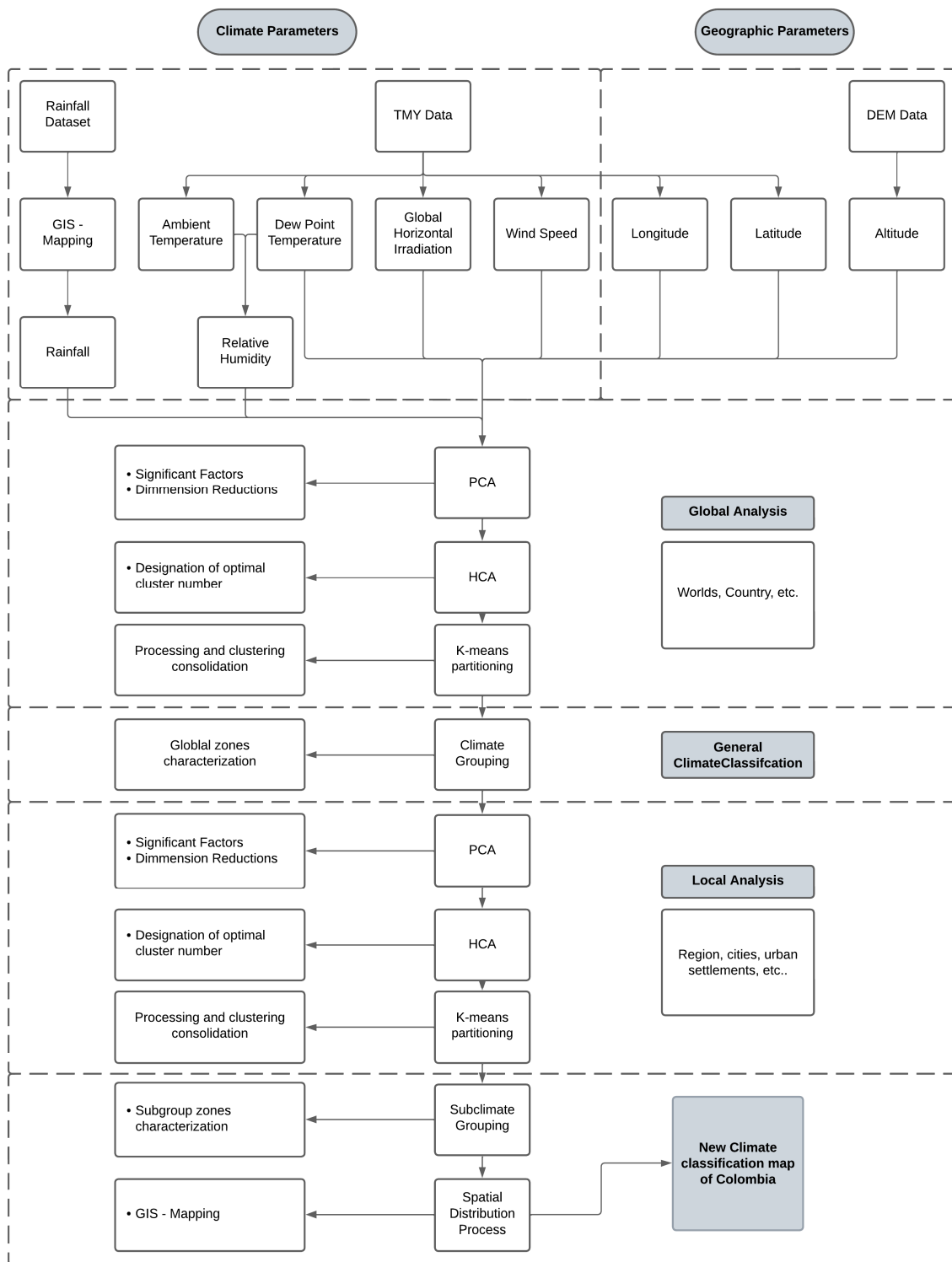


Figure 2. Process diagram for the current study.

3.1. Boundary Conditions

Due to the lack of hourly data in the records of the “Instituto de Hidrología, Meteorología y Estudios Ambientales” (IDEAM), we resorted to the use of typical meteorological year (TMY) data, which will be explained soon. However, since this kind of database generally does not include rain data, Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks—Dynamic Infrared Rain Rate near real-time (PDIR-Now) from the Center for Hydrometeorology and Remote Sensing (CHRS) at the University

of California, Irvine, was used [8]. Since it is a real-time global, high-resolution model (4 km per pixel), it adjusts to the TMY selected for the presented study. Also, these rain data contain raster with daily cumulative rainfall data from 1 March 2000 to 31 December 2021. Once the full range of days (7976) was downloaded, the mean for the cumulative rainfall data was calculated using Python's Rasterio Package [31] due to the largest raster to process.

The TMY data provide information on dry-bulb temperature and dew point temperature. From these values, relative humidity was calculated. The thermofeel library [32], a Python package that offers methods that allow estimating different parameters implicated in thermal comfort, including relative humidity, was used.

3.2. Data Processing

The National Renewable Energy Laboratory (NREL) has accumulated solar radiation and meteorological data for over two decades, creating an extensive worldwide database [7]. Specific to the study region, this database is enriched with information on TMY, a tool widely used in the analysis of the climatic behavior of an area [33]. These data are available at a cell resolution of 4.4 km per pixel and contain detailed hourly data, including dry-bulb temperature, dew point temperature, global horizontal irradiance (GHI), wind speed, and other parameters. In addition, the CHRS provides global rainfall estimations, providing relevant information on rain patterns throughout the study region. Based on the previous statement and methodology designed by Davidson et al. [3], for this study, the following climate factors were used:

- Daily cumulative GHI (W/m^2)
- Daily average wind speed (m/s)
- Daily maximum and minimum relative humidity (%)
- Daily maximum and minimum dry-bulb temperature ($^{\circ}C$)
- Daily cumulative rainfall (mm)

Considering the significant presence of the Andes Mountain range in Colombia's topography, a digital elevation model (DEM) was acquired for the study region. This model was obtained from the ALOS PALSAR satellite with a resolution of 12.5 m per pixel. This elevation layer was accurately extracted for each climatic point in the study area using Geographic Information Systems (GIS) tools. In the analysis, three geographical variables were included (altitude, latitude, and longitude). The original NREL database consists of a matrix with 69,784 points distributed throughout the study area, each with its own TMY data. However, during the data pre-processing phase, the points were reduced to 56,938 points due to the lack of information between the TMY cloud points and the CHRS rain data. This is because the CHRS models are in raster format, as well as the DEM models. During the extraction of data from the CHRS model, certain points were outside the CHRS accumulative rainfall map, so there was no data available, which caused those points to be discarded for the later stage of the analysis.

A data pre-processing process was carried out using Python as a statistics tool with the tools of the Sklearn Package. The maximum and minimum average hourly values were determined based on TMY data for the temperature and relative humidity variables. In the case of global horizontal irradiance (GHI), the accumulated daily average value was calculated and, finally, for wind speed, the annual daily average value was obtained. At the end of the process, the geographic variables and the rainfall data were included in the points database, as previously mentioned.

3.3. Clustering Analysis

Cluster analysis is a multivariate statistical technique that groups similar elements according to their characteristics or similarities. This tool offers several advantages in identifying patterns and hidden structures in large datasets. In climate classification, it is a tool that has proven to be very useful in the grouping of geographic areas with similar climatic patterns [4,19,20,34].

3.3.1. Principal Component Analysis (PCA)

Working with huge data series with many variables (dimensions) makes it difficult to visualize, analyze, and model. Principal component analysis (PCA) is a dimensionality reduction technique used worldwide in statistics and data analysis. The main objective of this technique is to simplify the data structure while maintaining as much information as possible from the original dataset [35]. PCA transforms the original data into a new set of variables (principal components, or PCs), considering a linear combination of the original dimensions. The scree test is a technique commonly used in this process that identifies the number of principal components that group the highest variability in the data. This scree plot is used to achieve an adequate reduction of the dimensionality of the dataset. This technique calculates each PC's eigenvalue, representing the variability in the data this component captures. The generated graph shows a descending curve of the eigenvalues. The point where the slope drops drastically and tends to 1 will indicate the optimal number of principal components to retain [36].

3.3.2. Hierarchical Clustering Analysis

A hierarchical agglomerative classification (HAC) was performed together with a validation by an unsupervised K-means classification. Different studies have shown HAC as an ideal method in the data exploration stage [20,37], where an iterative process of agglomerative clustering of the contained data is sought. The cloud of 56,938 points was grouped following Ward's criterion in the conformation of a large cluster. This grouping was then partitioned into two clusters. It is possible to determine the distance in points that allows the grouping of these in new clusters using the Euclidean distance method. Its basic principle is to minimize the variance between the new groups formed [22].

Secondly, K-means agglomerative classification was carried out. This statistical algorithm is a widely used clustering technique due to its processing efficiency. However, it can have difficulties initiating the centroids of its clusters because they are executed randomly in their initial phase [3].

Nevertheless, the silhouette coefficient is a correct evaluation metric to quantify the quality of clusters generated by clustering algorithms. It measures how well defined the clusters are and how far apart they are from each other [38]. For the present investigation, the silhouette coefficient was calculated for both the HAC method and the K-means algorithm to measure and self-validate the optimal number of clusters. This coefficient varies in the range of -1 to 1 , where higher values indicate better clustering quality, as indicated below:

$$S = \frac{b - a}{\max(a, b)} \quad (1)$$

where,

- a represents the mean distance of a point and other points that belong to the same clusters,
- b represents the mean distance of a point and points that belong to the nearest cluster.

We used the WCSS (within-cluster sum of squares) technique as a final rectification parameter. This metric measures the dispersion of the data points within each cluster. It is calculated by adding the squared distances between each data point and the cluster's centroid to which it is assigned [39]. This metric confirms the optimal number of clusters for the PCs selected in each research phase. This technique is mathematically explained below:

$$WCSS = \sum d(p, c)^2 \quad (2)$$

where,

- $d(p, c)$ is the distance between a point (p) and the cluster centroid (c) to which it belongs.

All the above methodologies were used for the climate data to define the optimal number of clusters in each process step. The validation by different tools helped to increase the quality and reliability of the data.

The K-means partitioning method is a clustering technique that assigns observations to K sets, where K represents the predetermined number of centroids. At each iteration, the observed point is assigned to the nearest centroid, and the centroid is recalculated as the mean of all assigned observations. This iterative process is repeated until the convergence in the assignment of samples to clusters is reached. The suitability of this kind of clustering analysis for handling large databases lies in its computational efficiency [23]. Segmenting the dataset into clusters significantly reduces the complexity of the analysis by focusing on more manageable groups [21].

The abovementioned procedure, including the data, was initially carried out for the global analysis. Subsequently, the same procedure was repeated for each cluster obtained by the first K-means partition analysis to carry out a local study.

3.4. Spatial Distribution Mapping

A GIS tool was necessary to represent the spatial distribution of the obtained clusters. The groups were projected onto a map of Colombia to obtain the spatial localization of each cluster. For the creation of the new classification distribution, the inverse distance-weighted tool was the most appropriate interpolation method, as it seeks to maintain the original values of the observations that are spatially georeferenced [20,40].

4. Results

4.1. Global Analysis and Clustering Process

Considering the procedure outlined in Figure 2, a principal component analysis (PCA) was carried out to perform dimensionality reduction during the global analysis. In line with the Kaiser criterion, only those principal components (PC) whose eigenvalues exceeded the one mark were selected. In this context, the threshold was set at three PCs, as illustrated in Figure 3A, which explains approximately 79% of the variance, as seen in Figure 3B. This proportion encapsulated significant variability in the data, thus ensuring an adequate representation of the factors evaluated.

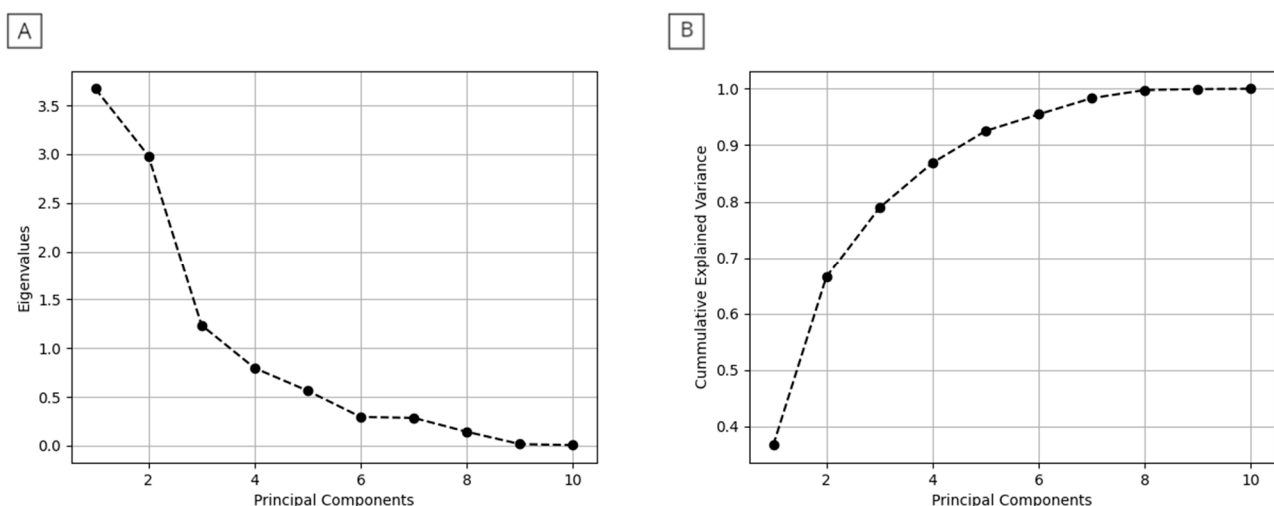


Figure 3. Scree test for selection of PCA components (A). Cumulative explained variance by each principal component (B).

The above analysis was complemented by the projection map of the principal components and climatic variables, as presented in Figure 4. This heat map provides a visualization of the influence of all variables on each of the principal components. Within this context, the contributions of temperature, elevation, and relative humidity in comparable

ranges were highlighted in the first component. On the other hand, in the second dimension, daily accumulated rainfall, latitude, wind speed, and elevation played relevant roles. Finally, variability was mainly determined by longitude, global horizontal irradiance, and precipitation in the third component. Figure 3 shows that both climatic and geographic variables contributed to the variability of the information within these components.

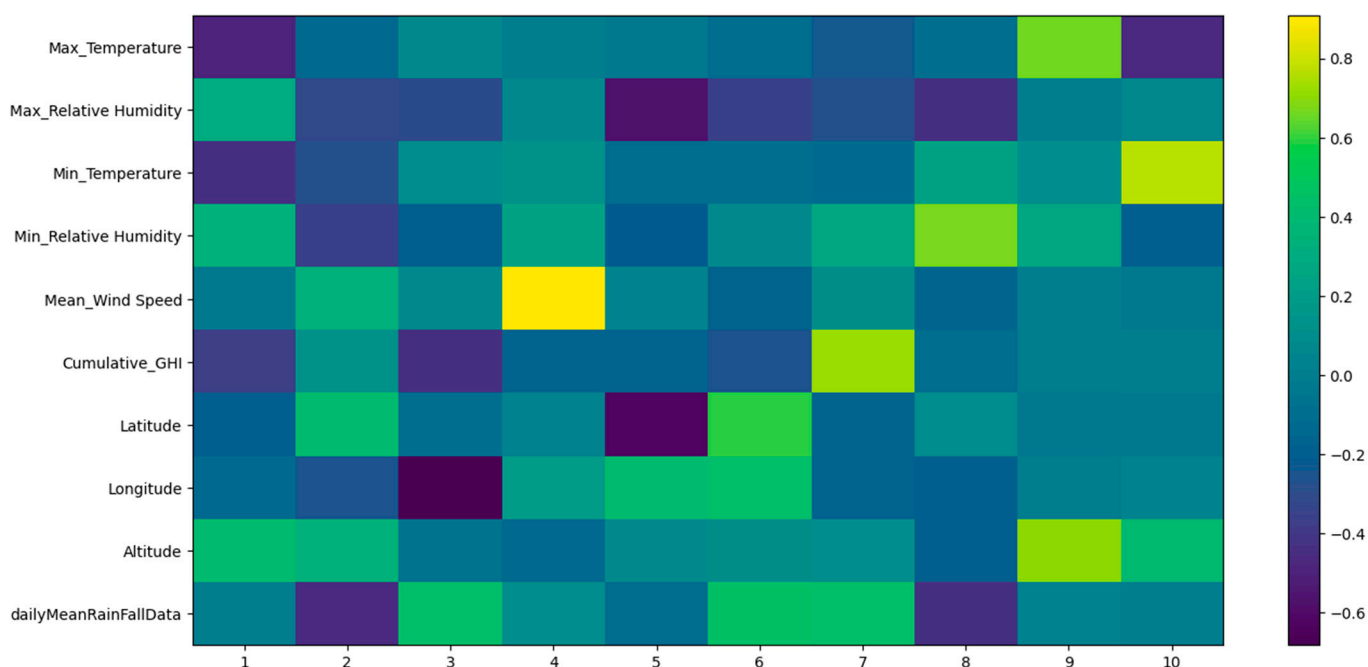


Figure 4. Heat map between initial variables and principal components.

It was also observed that from the fourth component onward, the information's variability was directed toward a single variable, which was inadequate to describe the sample evaluated. The figure confirms that three components are sufficient to explain the variance of the original dataset.

Then, the hierarchical agglomerative clustering (HAC) was carried out in the second stage of the global analysis. Within this procedure, a silhouette coefficient analysis was performed, the results of which were inconclusive since it indicated four, five, and six as the optimal number of clusters (Figure 5). On the other hand, the dendrogram analysis exhibited a Euclidean distance more aligned with the formation of four clusters. Consequently, these groups were finally selected as the optimal numbers for the overall analysis, as shown in Figure 6, where each color represents a different cluster.

Figure 7 presents the spatial distribution of the global climate analysis. Cluster 0 identified the extensive plains of Colombia with the highest average global horizontal irradiance among all the defined clusters. This characteristic generated a climatic condition marked by high temperatures and moderate relative humidity. In correlation with its spatial distribution, this cluster was in three zones of the country: north, east, and center, where the last one was between the three main mountain ranges that make up Colombia's mountain system. These regions are associated with a low-altitude topography, with some plateaus contrasting with a savanna climate. Cluster 1 was in the south of the country, with a small area located in the center of the region, in a geographic position closer to the equator, which resulted in a predominantly tropical climate with high humidity and temperature levels; additionally, due to its geographical condition and abundant vegetation, characteristic of a tropical forest where air currents are minimal, a lower wind speed range was observed in this area.

Cluster 2 described a mountain climate with elevations between 1000 and 5000 m.a.s.l, which inversely affected the temperature, decreasing consistently with increasing altitude.

This characteristic is due to its geographical position, located toward the country’s west, center, and east, along each mountain range of the Andes and the Sierra Nevada de Santa Marta in northern Colombia. This climatic group’s spatial distribution (latitude and longitude) entailed an amplitude between the thresholds of various climatic variables evaluated, such as rainfall, relative humidity, wind speed, and solar radiation. Furthermore, Cluster 4, due to its geographic location and primary climatic characterization, particularly in terms of temperature and relative humidity, shared similarities with Cluster 2. However, the western region of Colombia, which borders the Pacific Ocean, exhibited considerably higher cumulative rainfall than the other clusters. This area presented varied topography due to its location between the Pacific Ocean and the western Andes Mountain range, where it also had relatively less solar radiation than other clusters.

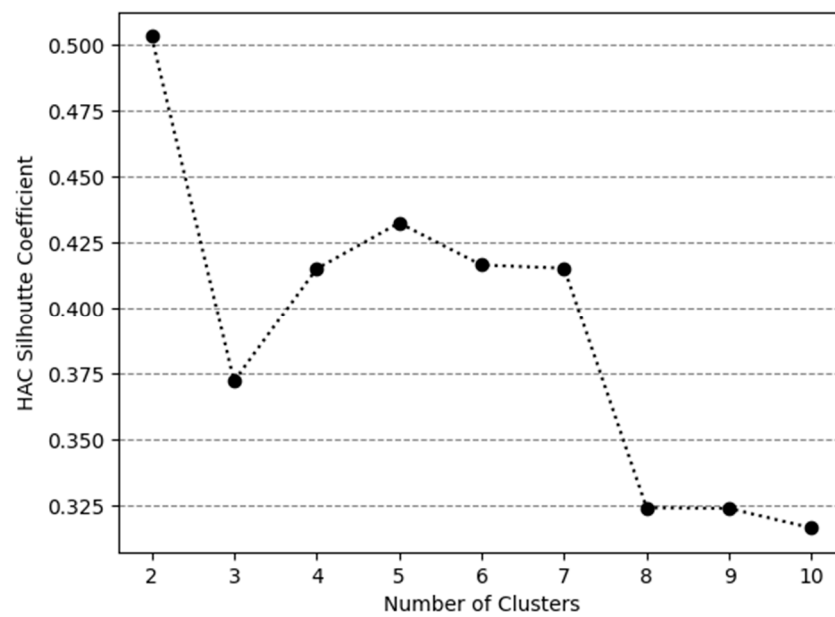


Figure 5. HAC silhouette coefficient for global analysis.

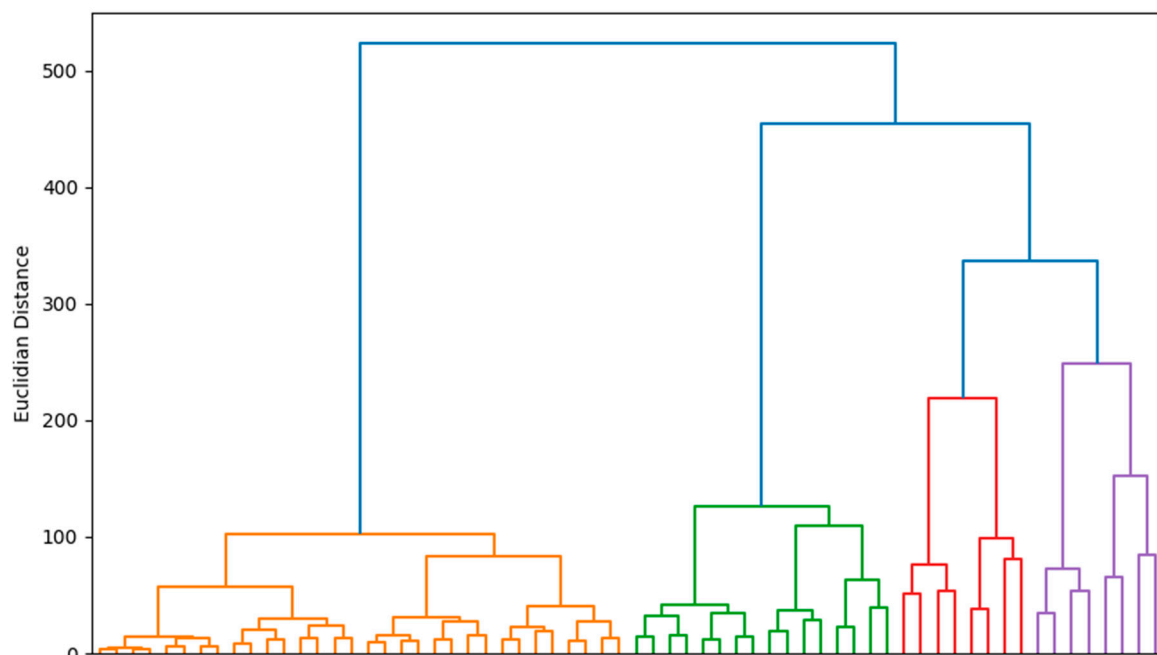


Figure 6. Global analysis dendrogram. Note: Each color of the dendrogram indicates a different former cluster.

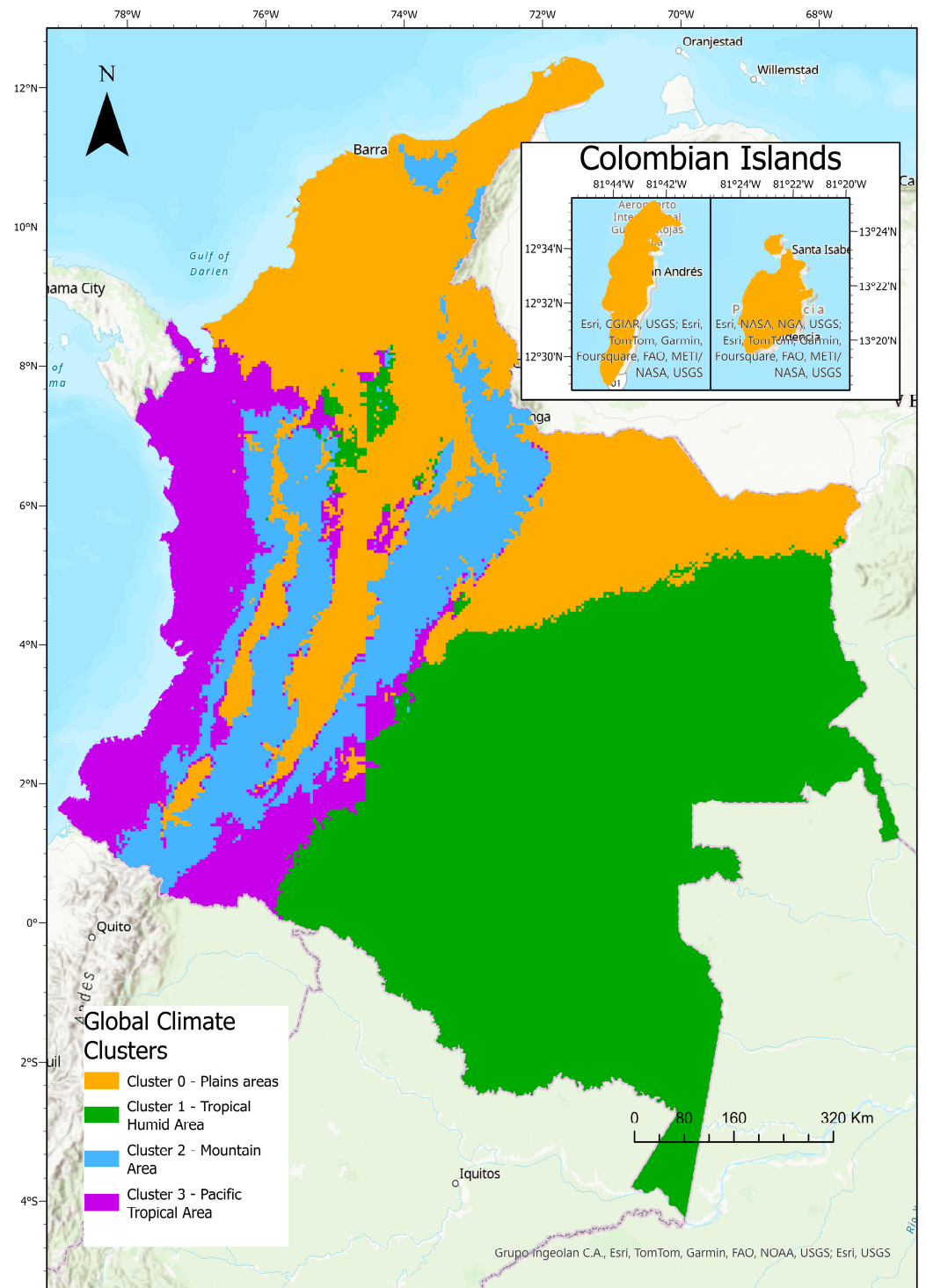


Figure 7. Global cluster spatial distribution.

The above clusters exhibited significant variability in the results obtained for each variable, as shown in Table 2. This variability was expected since the clusters shared similar climatic conditions but may also present marked differences among other geographic or climatic parameters. Considering this premise, a second analysis was indispensable to reduce the variability between regions, generating a more accurate climatic localization, as shown below.

Table 2. Statistics of global cluster for Colombia.

Global Cluster		Max. Temperature (°C)	Max. Relative Humidity (%)	Min. Temperature (°C)	Min. Relative Humidity (%)	Mean Wind Speed (m/s)	Daily Cumulative GHI (w/m ²)	Cumulative Rainfall Mean (mm)	Elevation (m)
Cluster 0	Average	31.13	90.91	22.91	54.90	1.11	5385.29	5.77	330.07
	Min.	20.82	65.15	12.22	35.66	0.02	3614.93	0.83	0.00
	Max.	36.85	100.00	28.24	81.34	7.61	6531.18	12.55	2600
	Std	2.41	6.45	2.29	8.02	1.14	348.03	2.08	416.24
Cluster 1	Average	28.29	99.76	22.51	76.59	0.26	4851.96	9.02	209.29
	Min.	23.04	94.14	15.03	58.02	0.05	4183.88	5.52	40
	Max.	30.89	100.00	24.24	89.27	1.68	5585.80	11.32	1555
	Std	0.68	0.58	0.58	4.23	0.33	144.51	1.06	116.61
Cluster 2	Average	18.76	99.36	10.74	74.12	1.11	4410.98	4.77	2442.84
	Min.	3.40	81.23	−4.76	50.3	0.03	2688.17	1.75	773
	Max.	26.73	100.00	19.44	100.00	3.69	6227.63	11.89	5081
	Std	3.91	2.00	4.05	11.35	0.49	654.43	1.67	693.47
Cluster 3	Average	27.92	95.32	21.92	68.21	1.00	4304.76	11.31	440.47
	Min.	20.93	73.61	13.99	45.12	0.03	2894.31	3.45	3.00
	Max.	31.69	100.00	27.41	97.9	3.94	5555.64	24.4	2118
	Std	1.72	4.63	2.42	9.24	0.85	421.83	3.77	443.85

Source: Authors.

4.2. Microclimate Configuration

The process used in the global analysis was replicated for the delimitation of microclimates through local analysis. For each point located in the main groups, a dimensionality reduction analysis was carried out, maintaining three principal components in all cases, obtaining an explained variance between 70% and 85% for all four former clusters. On the other hand, in the hierarchical cluster analysis, the silhouette coefficient, the dendrogram, and the WCSS plots were evaluated as criteria for selecting the optimal number of groups. The result revealed 17 microclimates distributed within the first 4 main groups. These microclimate results are presented in Figure 8.

The subdivision of the main Cluster 0 generated four subgroups designated with the letters A, B, C, and D. Table 3 shows the local statistics for each climatic variable in the conglomerate analyzed. It is relevant to note that, within these subgroups and through the spatial distribution, the region was subdivided mainly according to the following geographic and climatic parameters: terrain elevation, temperature, and rainfall. In the north of the country and the upper part of the northwest coast, there is a region with unique characteristics compared to the others, presenting an arid climate with minimal precipitation but with a considerably higher wind speed. This is due to its geographic position, where the highest trade winds in the country are concentrated. On the other hand, conglomerate 0-C represents the mountainous slopes, serving as a transition zone toward colder climates present in the main Cluster 2. Finally, group 0-D showed the highest temperature levels, consistent with a lower altitude and wind speed magnitude compared with the other groups of the main cluster.

The results derived from the sub-clustering of cluster 1 revealed the existence of three distinct subgroups with clear difference statistics, as shown in Table 4. Subgroup 2-B stood out mainly for having steeper elevations than the other clusters. On the other hand, subgroup 1-A exhibited the highest mean rainfall and considerably high humidity relative to the different clusters, consistent with its proximity to the equator. Concerning subgroup 1-C, lower mean precipitation and higher temperatures were observed, suggesting a transition climate toward drier zones, considering that it is located between the Colombian Amazon (1-A) and the country’s eastern plains (0-B).

Cluster 2 presented the most significant climatic variability identified within the region under study. In the context of the definition of clusters, six microclimates were identified that exhibited both similarities and important differences (Table 5). A paradigmatic example is the area represented by Cluster 2-D, characterized by a higher mean elevation than the other clusters. Consistent with the altitudinal gradient, this cluster recorded a decrease

in mean temperature (both maximum and minimum) of approximately 5 degrees Celsius compared to the closest cluster in terms of mean altitude (2-E). Despite having the lowest mean cumulative precipitation of all clusters, Cluster 2-D exhibited remarkably high relative humidity.

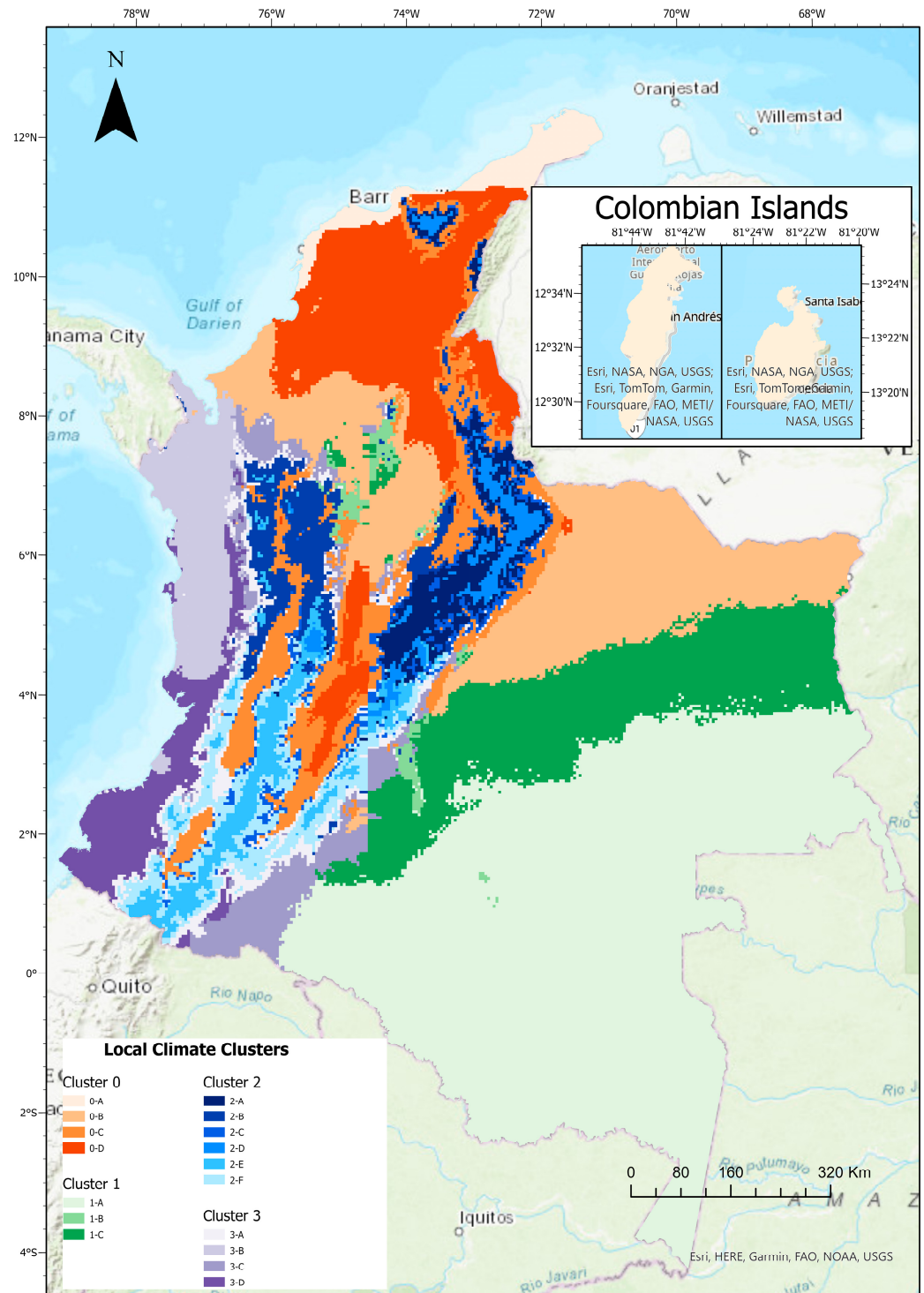


Figure 8. Microclimate spatial distribution for Colombia.

Table 3. Local statistics for Global Cluster 0.

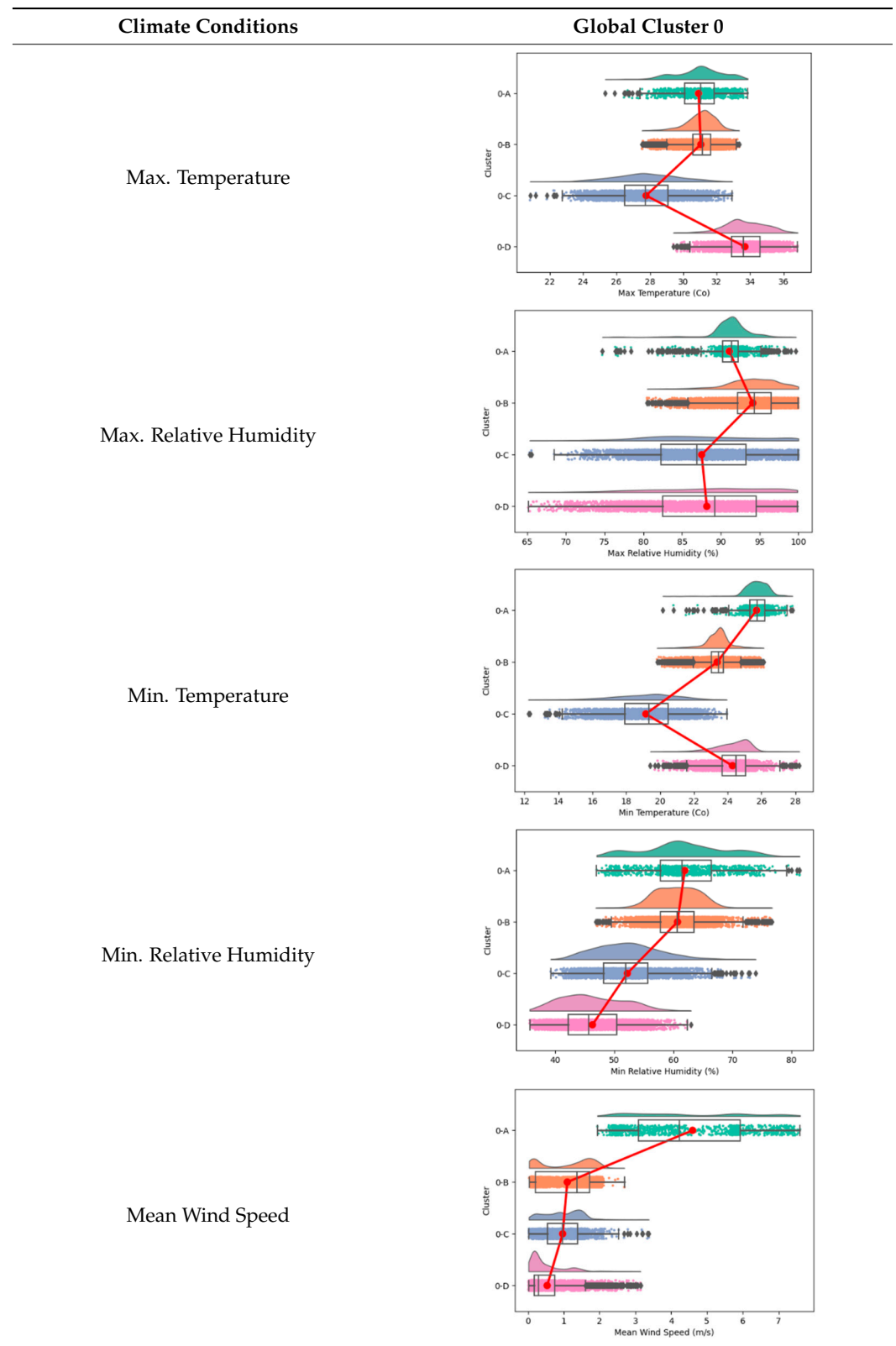


Table 3. Cont.

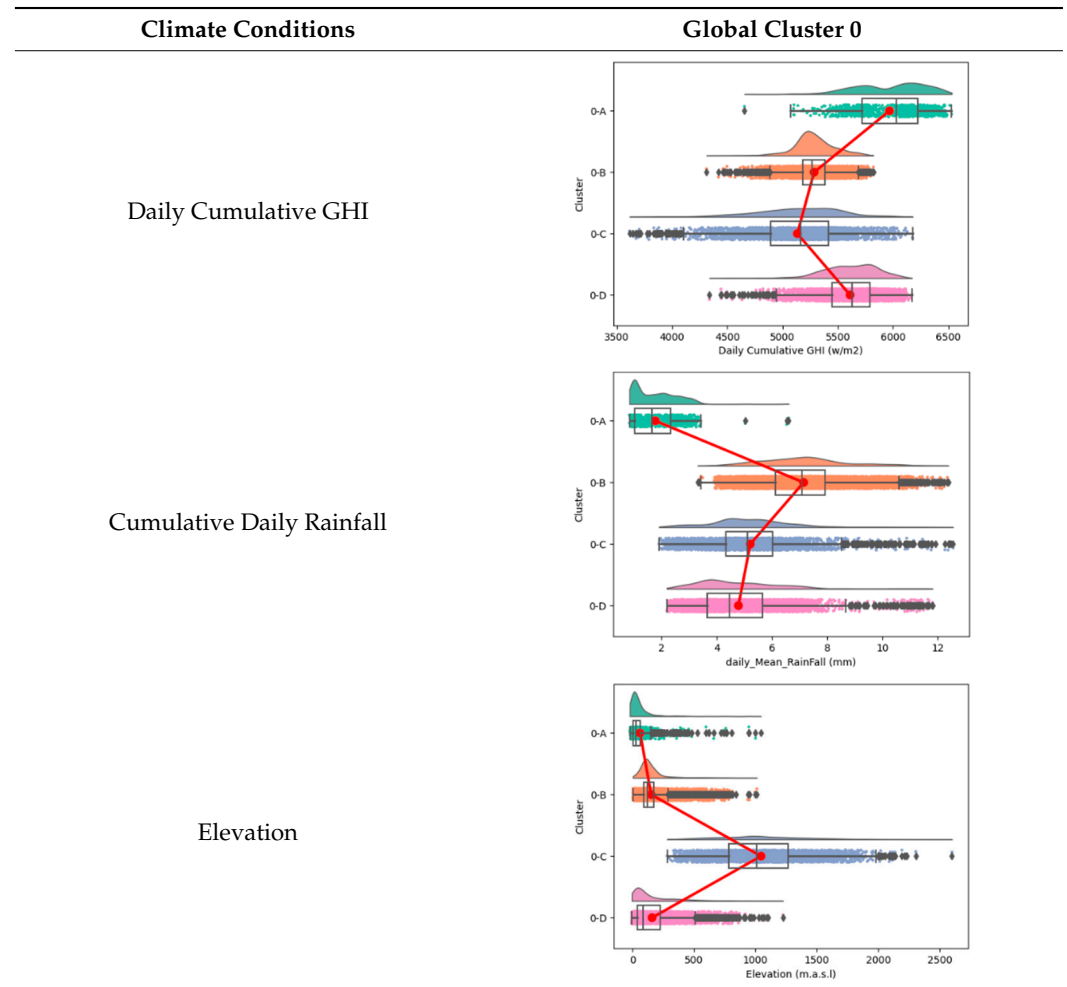


Table 4. Local statistics for Global Cluster 1.

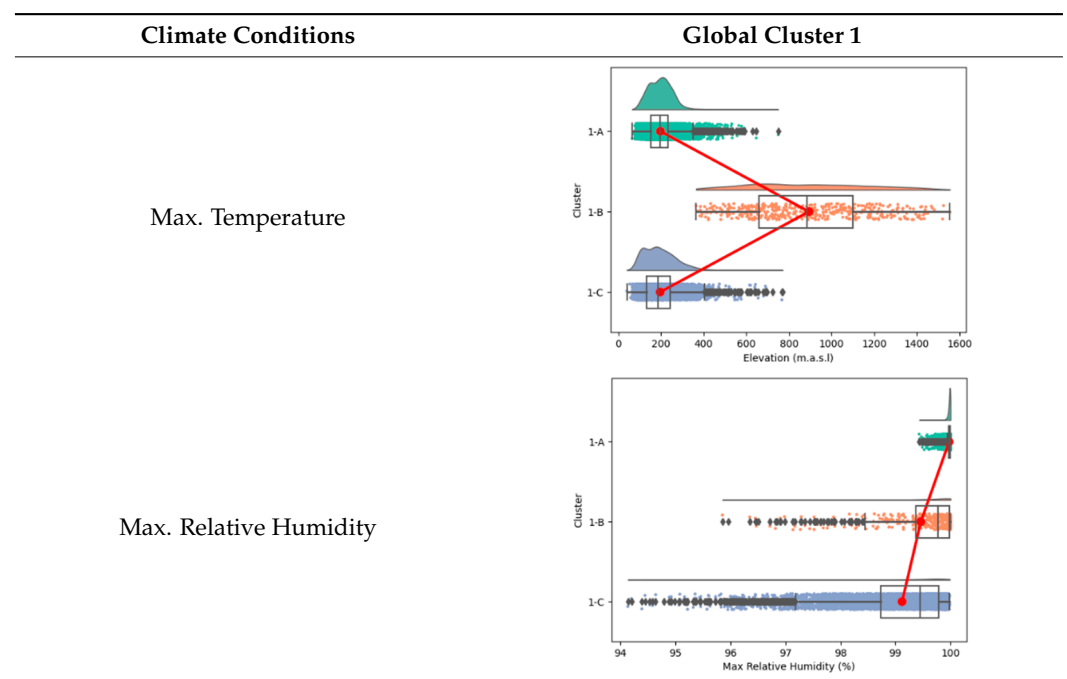


Table 4. Cont.

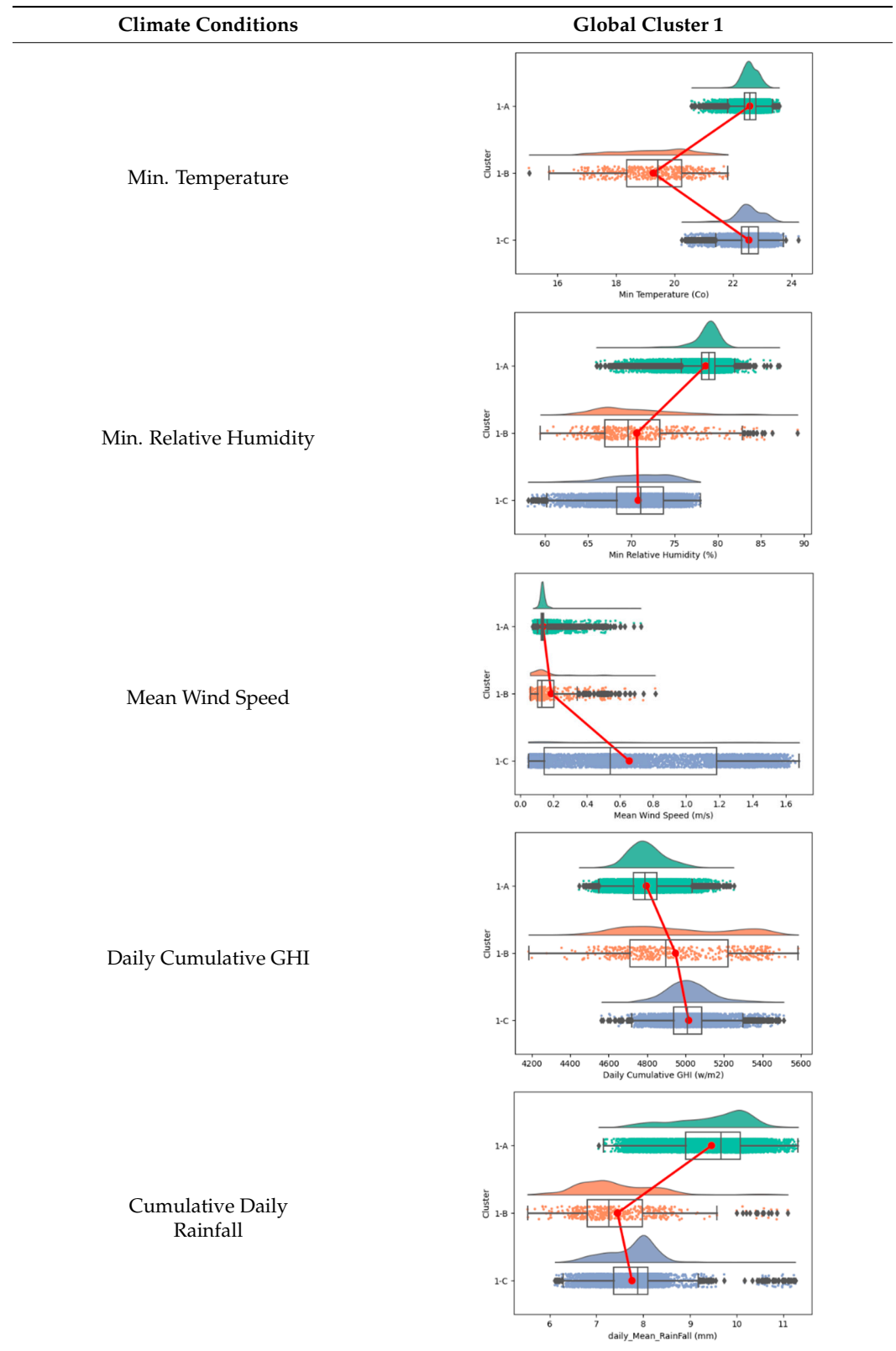


Table 4. Cont.

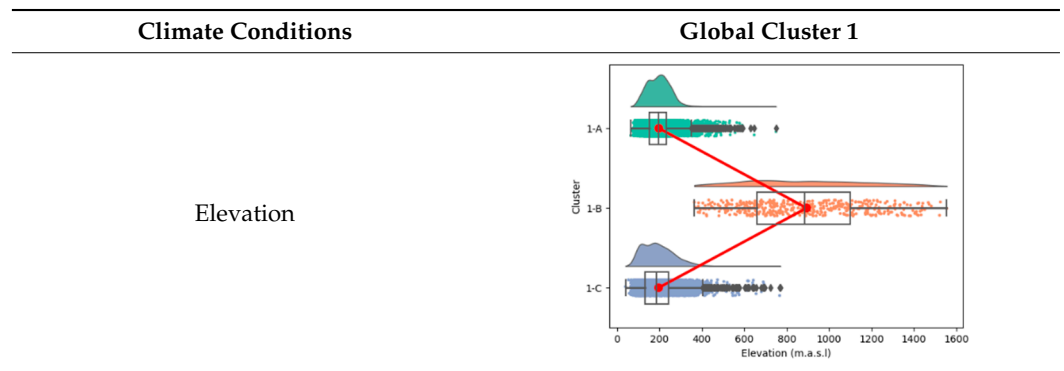


Table 5. Local statistics for Global Cluster 2.

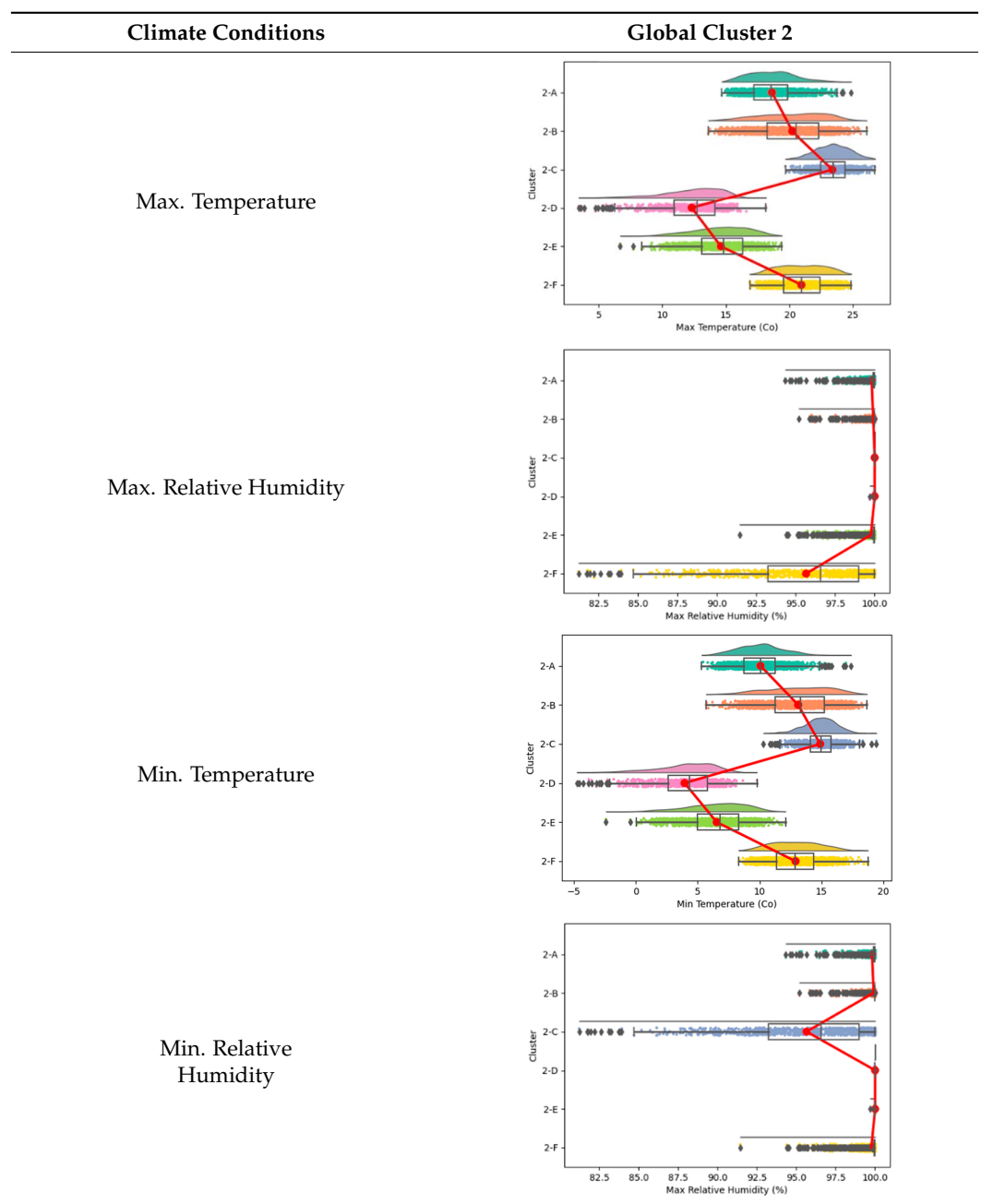
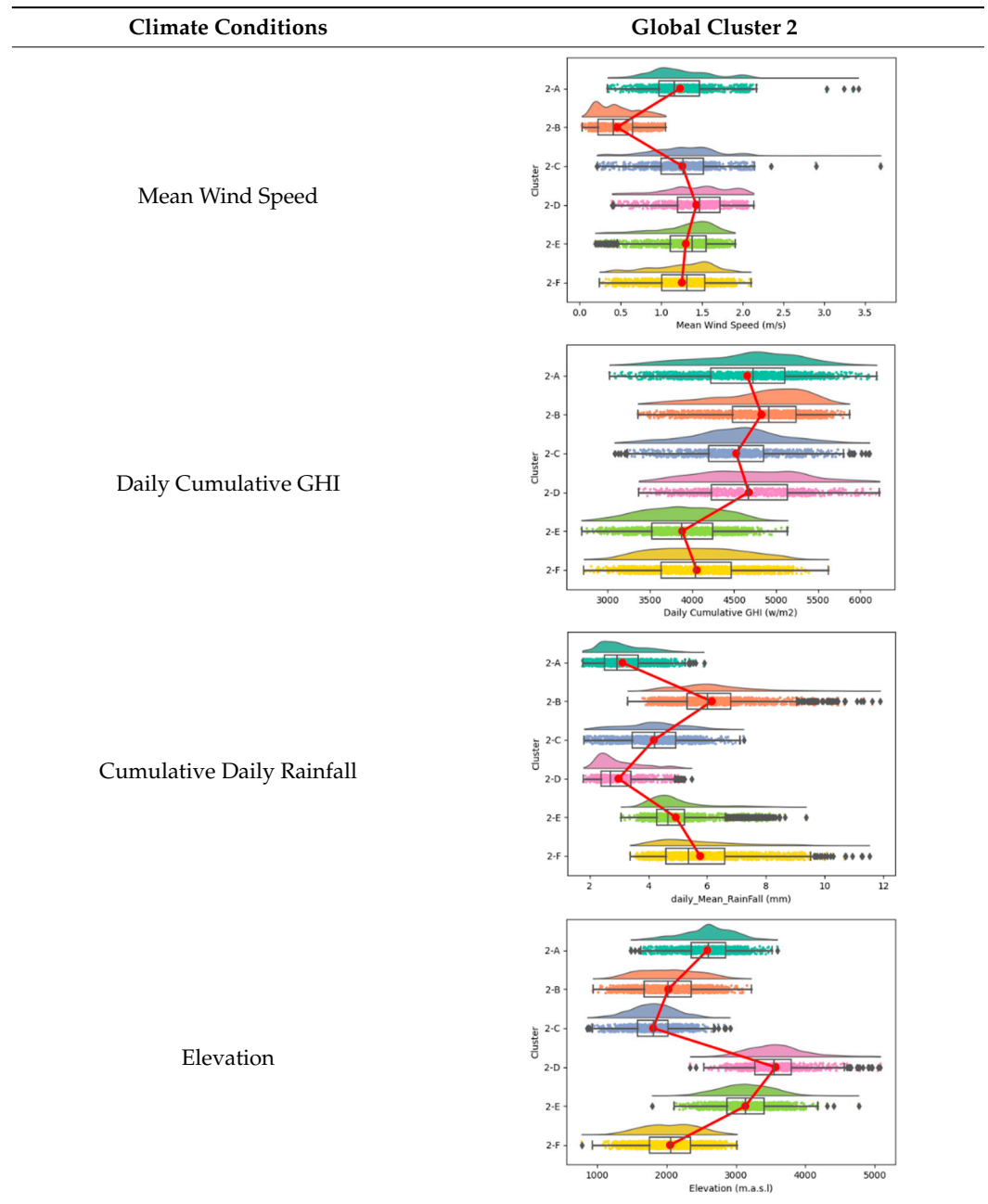


Table 5. Cont.



On the other hand, the Western Mountain was predominantly influenced by Clusters 2-B and 2-E, which, although sharing very similar mean characteristics, showed differences in the radiation present in the area that affected the mean minimum humidity of the clusters. In addition, latitude influenced this cluster, generating a division between the north and south of the Western Mountain. Microclimate 2-C was distributed along the three mountain ranges located on their foothill. This indicated a clear transition between warmer sub-climates and those that were cooler, characteristic of the influence of the altitudinal gradient. It is essential to highlight the high humidity in all these regions, demonstrating the geographic influence due to its proximity to the equator.

For Cluster 3, the subdivision resulted in a total of four microclimates that extend along the Pacific coast and some areas around the Western Mountain. Specifically, subgroup 3-A was characterized by a mean elevation twice that of the other subgroups, distinguishing it markedly from the others, as shown in Table 6. This disparity may be explained by its

proximity to the mountain range, where it is widely distributed. It suggests its role as a transition region between tropical zones, higher elevation areas, and lower temperatures. On the other hand, although subgroups 3-B and 3-D shared close precipitation levels, the latter recorded the highest amount of rainfall and a higher average wind speed. Subgroup 3-B presented higher temperatures related to its geographic location closer to the equator than subgroup 3-D. Finally, subgroup 3-C was the only one not located directly on the Pacific coast; instead, it was in the lower and upper parts of the Western Mountain range, opposite the Pacific. Its location and low wind speed generated a clear distinction concerning the other subgroups.

Table 6. Local statistics for Global Cluster 3.

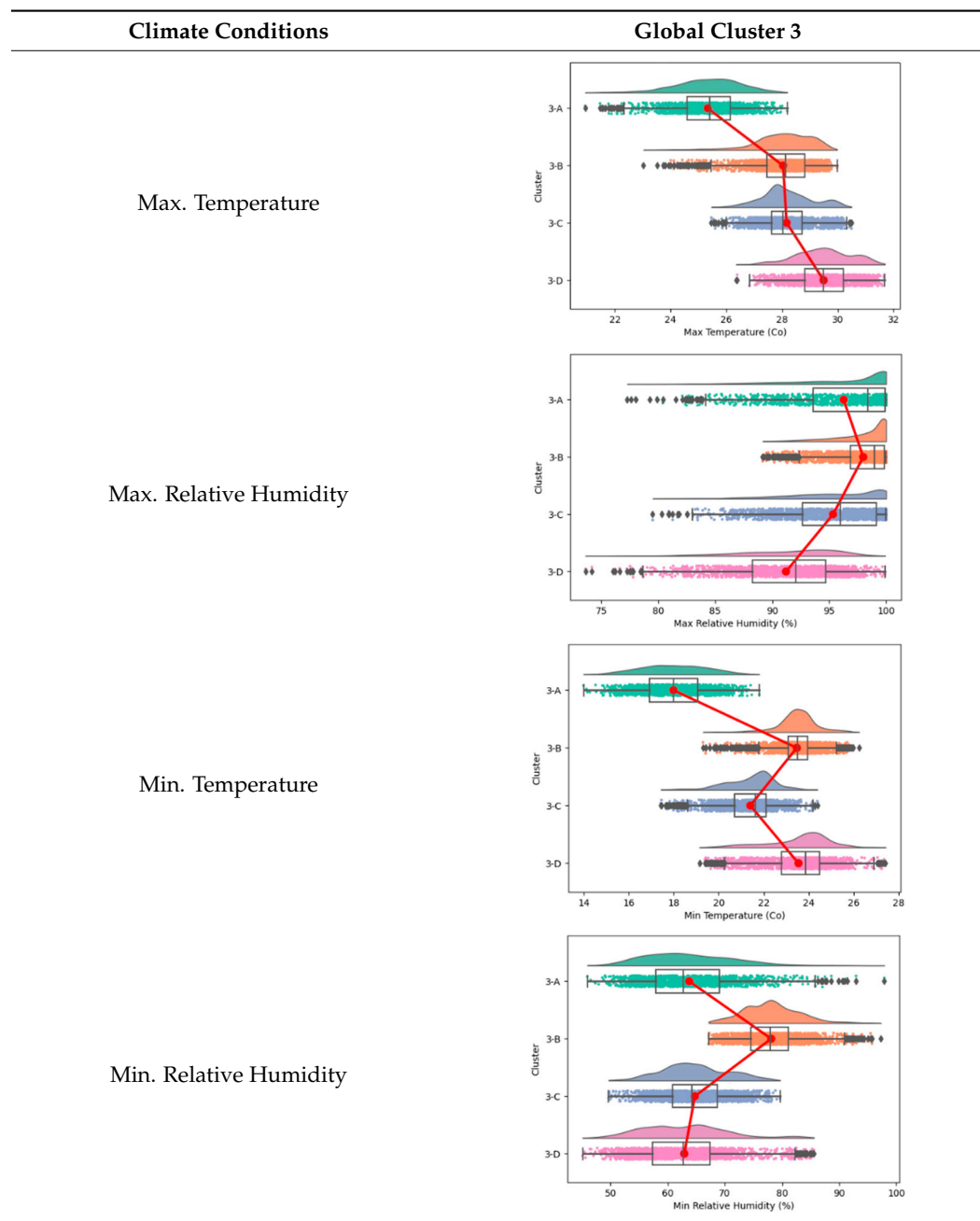
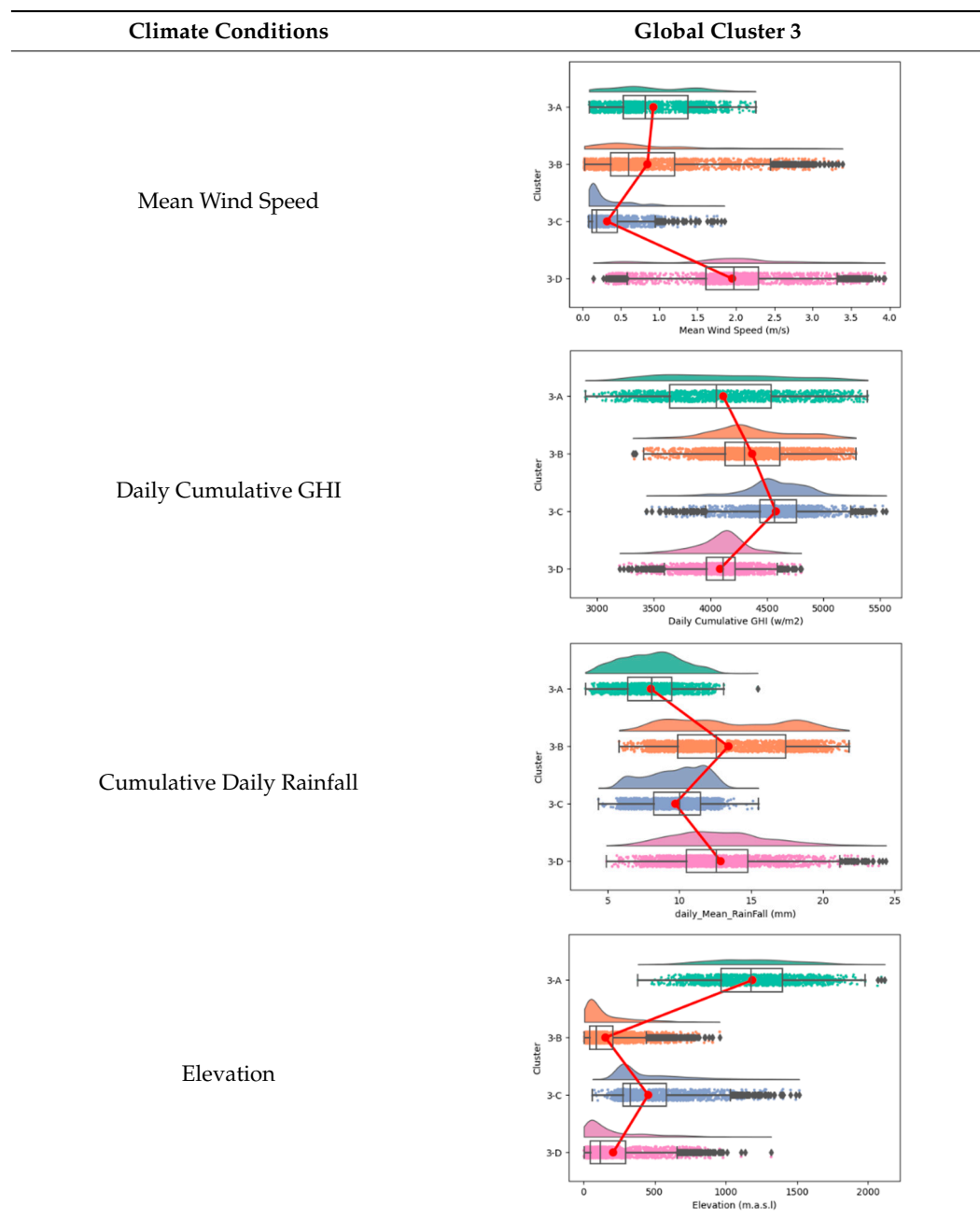


Table 6. Cont.



5. Discussion

5.1. Key Research Findings

Several climate classification methods have been developed and applied in different regions of the world, with the Köppen classification being one of the most recognized and used. This classification is based on parameters such as temperature, rainfall, and vegetation for climatic distribution. Although its use has spread globally, the authors of [41] indicated that it has lower accuracy in assessing the energy performance of buildings compared with other methodologies. Therefore, other climate classifications, such as the degree day method, have been developed to describe heating and cooling needs in specific climates. This method is one of the most extensive approaches for building-oriented climate zoning.

Nevertheless, it only provided a limited view of the influence of climate conditions [13]. On the other hand, to describe local climates, the LCZ method has been implemented in

different regions worldwide due to its potential in the characterization of the local environment, considering various climatic, geographic, and built environments [42,43]. Although all these classifications present particularities and defined uses, they share the characteristic of the need for supervised analyses, which is a relevant aspect of their development.

Conversely, the use of clustering algorithms in climate classification is one of the most popular methods in recent times, especially for its contribution to urban planning and resource management sectors [41]. These methods are typically based on unsupervised analysis, providing a holistic assessment when these labels are unavailable [21]. An area's climatic, topographic, and geomorphological parameters directly impact the regional climate, generating complex patterns and interrelated variables that clustering algorithms can identify more precisely. In addition, unsupervised analysis can reveal conditions that are not obvious to the naked eye, facilitating understanding of climate dynamics and identifying key features that influence climate patterns [44]. In addition, unsupervised analyses offer greater scalability, making them more efficient in processing large volumes of data and simplifying the development of a climate classification [45].

On the other hand, one of the recurring challenges when analyzing the climatic behavior of a development region lies in the limited spatial distribution of weather stations. However, technological advances in the implementation and use of satellite imagery have led to the development of more advanced meteorological models [7]. These models typically offer higher spatial and temporal resolution compared to traditional weather station data [27]. This improvement implies the ability to capture local climate variations in greater detail and frequency, which is especially beneficial for geographic areas with complex topography or local climate variations. Nevertheless, emerging countries usually do not have satellites, making data recollection difficult. However, technological advances have made it easier for a wide variety of users to access and use satellite data and associated weather models [8]. This makes it possible for researchers, urban planners, and government agencies to use this information to identify weather patterns in a region. This tool is especially valuable in areas where the availability of updated data from weather stations is limited, which is common in developing countries [46]. In this context, the use of meteorological models becomes essential for an adequate updated climate classification of any region that lacks the resources or data necessary to carry it out. This allows for an accurate climate assessment in support of the sustainable development of the region.

Developing a microclimatic classification through a multivariate analysis of diverse climatic and geographic characteristics from satellite imagery for Colombia represents a significant advance in understanding and representing an updated local climatic pattern in the region. The implementation of dimensionality reduction models not only provides an advantage in data processing but allows linking variables within the same dimension. This allows a closer relationship between climatic parameters compared to other traditional climate classifications. Therefore, comparing these classifications is ideal for assessing the accuracy of the spatial distribution of climates in the region.

The global analysis divides Colombia into four main zones, reflecting the distinctive climatic characteristics of each region: warm plains, tropical climate of the Amazon Jungle, super-humid tropical climate in the Pacific coast, and mountain climate region in the Andes Mountains range. This clustering provides a geographical division not visualized in other earliest climatic approaches performed in the region. For example, the division between the Pacific region from the Amazon jungle in the south of the country is not presented in the Köppen classification. The Western Mountains near the Pacific coast generate an orographic effect that produces condensation of humid air from the ocean, which results in a considerable increment of precipitation that does not occur in the south of the country.

Within its local analysis, the developed classification generated a clear division in the tropical climate cluster in the south of the country, creating a transition zone between a more humid climate and a drier and less humid zone, such as the savanna climate that borders the country's southeast. This is a similar characteristic to the one defined by Köppen as a transition zone between the tropical rainforest climate and a tropical savanna

zone. However, the microclimatic classification allowed us to note a smoother and more detailed transition zone than the one presented by Köppen.

On the other hand, the Caldas–Lang classification [28] and the new classification map present similarities in the location of warm climates in their different humidity levels, indicating that variables such as temperature, rainfall, altitude, and relative humidity significantly influence these climatic regions more than other geographic or climatic characteristics.

In mountain climates, the Caldas–Lang and Holdridge classifications [28] reflect a climatic diversity similar to Cluster 2, confirming that the altitudinal gradient near the equator causes notable changes in the climate of a region. However, classifications such as Caldas–Lang are based on thermal floors by elevation, with defined ranges for transitions between regions. Although this allows the identification of temperature changes between regions, it may limit the understanding of climate transitions influenced by other climatic or geographic factors. From this perspective, the multivariate model considers all variables rather than a single parameter, thus avoiding assigning variability exclusively to a specific factor.

Other classifications focusing on the effect of climate on humans have been performed in the region. For example, the thermal comfort classification [29] shows the level of thermal comfort throughout the region. This classification indicates how the altitudinal gradient affects human thermal stress, considering mountain slopes as more comfortable climates. However, its information is limited since its data have an outdated time series that only contemplates temperature and relative humidity, leaving out those other factors, such as solar radiation, that clearly impact the level of thermal stress [47–49].

Moreover, Colombian authorities in 2012 suggested a climate classification for architectural and building design. This model includes different climate and geographical variables [30]. Their classification has some similitudes with the global analysis of this research. However, the Andes Mountain was described as having a cold, humid climate in its entirety, oversimplifying the climatic complexity of this area. In contrast, the local classification of this research is a more detailed and precise approach that will allow the development of more effective bioclimatic strategies. This shows that the local climate approach would allow urban planners to improve the planning and design of sustainable architectural projects.

Mejía-Parada et al. [4] generated a clustering analysis in the study region to sectorize and propose passive bioclimatic design strategies based on the psychrometric diagram, a tool widely used in preliminary design stages of climate-resilient buildings [14,15]. However, this tool is limited to temperature and relative humidity interactions. As mentioned before, other variables, such as solar radiation and wind speed, are necessary complements for understanding and taking advantage of local climatic conditions through passive strategies. From this perspective, climatic micro-localization serves as a complementary tool for the understanding of the bioclimatic behavior of the different regions of Colombia.

5.2. Contributions toward Practice

It is essential to recognize that climate is dynamic and complex, subject to short- and long-term changes due to various factors, including natural climatic variations and human activities. Therefore, the new climate map should be considered as a snapshot representation of current climate conditions, considering that the implemented data were updated to 2021. Although challenges remain, the new approach offers a valuable tool for future research in climatology, meteorology, and related disciplines and for decision-making in areas such as agriculture, natural resource management, and urban development planning in emerging countries.

This novel classification of the region provides a deeper and more detailed understanding of global and local climate conditions for countries with difficulties in climatology data access, which is essential for planning and designing urban environments that are both ecological and resilient in the face of current and future climate challenges. This methodology should be applied in countries with social–economic and climatological contexts

similar to Colombia's. For example, regions located in central Africa and southeast Asia share tropical climate similarities and are also emerging countries that can benefit from this approach. Moreover, the use of updated data is essential because climate patterns are undergoing significant changes due to global warming and other environmental factors. Outdated climate zoning would not accurately reflect current conditions, which could lead to inappropriate decisions in areas such as urban planning, agriculture, resource management, and others.

In addition, accurate climate zoning is essential for optimizing bioclimatic designs in the framework of building resilient and friendly urban environments. For agriculture, updated information allows for better planning of planting cycles, irrigation, and crop selection, which in turn can increase yields and efficiency in the use of resources, such as water and fertilizers [50,51]. Also, this type of climate characterization is crucial for the efficient management of water and forestry resources. By better understanding precipitation patterns, temperature, and other climatic factors, resource managers can make more informed decisions about water allocation, forest fire prevention, and natural habitat conservation.

Including renewable energy in sustainable design is becoming increasingly common worldwide [52,53]. For example, evaluating methodologies, such as near-zero energy consumption buildings, have become popular in recent decades [54]. From this perspective, the new classification not only complements traditional passive strategies but also shows a panorama of the possible use of solar radiation as a source of electrical energy and heating, especially considering that Colombia's government had proposed, in its National Development Plan, an energy transition to reduce the use of fossil fuels due to Colombia's high potential of renewable resources for energy generation [55].

This approach was represented as an innovative and scalable climate zoning classification for developing countries. For the study region, the map divided the country into global and local scales, each with different climate thresholds that allow better urban planning and a more complete understanding of climate impacts in diverse areas, from agriculture to natural resource management. Its usefulness extends to identifying climate transitions and the potential use of renewable energies, highlighting its relevance in climate research and the design of sustainable environments.

6. Conclusions

Climate has a considerable impact on human activities, especially in emerging countries with hot and humid climates. Defining an appropriate climate zoning can be a valuable tool for urban planning, resource management, and agricultural fields. For their part, using TMY and PDIR-Now products from open-access satellite databases can be a helpful alternative, especially in regions with insufficient data or outdated information that does not reflect climate change's current challenges.

This study carried out a detailed analysis of Colombia's climate using PCA, HCA, and K-means techniques for 56,938 climate records spaced approximately every 4.4 km. Through the global analysis, four main climatic zones were identified in the country: warm plains, tropical climate of the Amazon rainforest, super humid tropical climate on the Pacific coast, and mountain climate region in the Andes Mountains. These divisions provided a detailed geographic overview of the distinctive climatic characteristics of each region, surpassing previous classifications, such as the Köppen classification. Moreover, the local analysis allowed to precisely delineate the microclimates within each of these major zones. A total of 17 microclimates distributed across the country were identified, each characterized by a different threshold of the climatic and geographic evaluated factors. This subdivision provided a detailed understanding of local climatic conditions, especially in the mountain chains of the Andes, which cross the country. The altitudinal gradient demonstrated an important influence on the climate of tropical regions, which included irregular topographies.

When compared to previous climate classifications, significant similarities and differences were observed in the representation of climate patterns. However, the multivariate

approach of the research showed greater accuracy and scalability by considering all relevant climate variables, avoiding oversimplifications, and providing a more accurate representation of Colombia's climatic complexity. On the other hand, the principal climatic and geographic parameters that showed influence in the region were temperature, rainfall, relative humidity, altitude, latitude, and wind speed. As a result, the climate zoning presented a new division for Colombia, dividing the tropical Pacific coast from the tropical Amazon jungle, where other climate classifications had not generated this separation between zones.

The practical contributions of this study were substantial, providing a valuable tool for future research in climatology, meteorology, and related disciplines in emerging countries. Detailed information on local climatic conditions is crucial for optimizing bioclimatic designs, building energy efficiency, and agriculture resources, among others in the region. In addition, this study will serve as a basis for further research on the impact of climate change in Colombia and other developing countries, including the development of adaptation and mitigation strategies, providing a step toward the development of more sustainable communities.

Author Contributions: All authors contributed to the study's conception and design, material preparation, data collection, and analysis were performed. The first draft of the manuscript was written by C.M.-P., and all authors commented on previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: The authors thank Research and Development University (UDI) for its support in developing this research. We want to acknowledge the Sustainable Building Design (SBD) lab at the Faculty of Applied Sciences at the University of Liege for valuable support during the data collection and data analysis.

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

References

1. Perera, N.G.R.; Emmanuel, R. A "Local Climate Zone" Based Approach to Urban Planning in Colombo, Sri Lanka. *Urban. Clim.* **2018**, *23*, 188–203. [[CrossRef](#)]
2. Reckien, D.; Salvia, M.; Heidrich, O.; Church, J.M.; Pietrapertosa, F.; De Gregorio-Hurtado, S.; D'Alonzo, V.; Foley, A.; Simoes, S.G.; Krkoška Lorencová, E.; et al. How Are Cities Planning to Respond to Climate Change? Assessment of Local Climate Plans from 885 Cities in the EU-28. *J. Clean. Prod.* **2018**, *191*, 207–219. [[CrossRef](#)]
3. Davidson, A.S.; Malet-Damour, B.; Praene, J.P. A New Microclimate Zoning Method Based on Multivariate Statistics: The Case of Reunion Island. *Urban. Clim.* **2023**, *52*, 101687. [[CrossRef](#)]
4. Mejía-Parada, C.; Mora-Ruiz, V.; Attia, S. Bioclimatic Design Recommendations for Novel Cluster Analysis-Based Mapping for Humid Climates with Altitudinal Gradient Variations. *J. Build. Eng.* **2024**, *82*, 108262. [[CrossRef](#)]
5. Ascencio-Vásquez, J.; Brecl, K.; Topič, M. Methodology of Köppen-Geiger-Photovoltaic Climate Classification and Implications to Worldwide Mapping of PV System Performance. *Sol. Energy* **2019**, *191*, 672–685. [[CrossRef](#)]
6. Attia, S.; Lacombe, T. Architect-Friendly Climate Analysis Tool for Bioclimatic Design in Hot Humid Climates. In Proceedings of the Building Simulation 2019: 16th Conference of International Building Performance Simulation Association, Rome, Italy, 2–4 September 2019; Volume 7, pp. 4785–4792.
7. Sengupta, M.; Xie, Y.; Lopez, A.; Habte, A.; Maclaurin, G.; Shelby, J. The National Solar Radiation Data Base (NSRDB). *Renew. Sustain. Energy Rev.* **2018**, *89*, 51–60. [[CrossRef](#)]
8. Nguyen, P.; Ombadi, M.; Goroooh, V.A.; Shearer, E.J.; Sadeghi, M.; Sorooshian, S.; Hsu, K.; Bolvin, D.; Ralph, M.F. Persiann Dynamic Infrared-Rain Rate (PDIR-Now): A near-Real-Time, Quasi-Global Satellite Precipitation Dataset. *J. Hydrometeorol.* **2020**, *21*, 2893–2906. [[CrossRef](#)] [[PubMed](#)]
9. Walsh, A.; Cóstola, D.; Labaki, L.C. Review of Methods for Climatic Zoning for Building Energy Efficiency Programs. *Build. Environ.* **2017**, *112*, 337–350. [[CrossRef](#)]

10. Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated World Map of the Köppen-Geiger Climate Classification. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1633–1644. [[CrossRef](#)]
11. Beck, H.E.; Zimmermann, N.E.; McVicar, T.R.; Vergopolan, N.; Berg, A.; Wood, E.F. Present and Future Köppen-Geiger Climate Classification Maps at 1-Km Resolution. *Sci. Data* **2018**, *5*, 180214. [[CrossRef](#)]
12. Martinopoulos, G.; Alexandru, A.; Papakostas, K.T. Mapping Temperature Variation and Degree-Days in Metropolitan Areas with Publicly Available Sensors. *Urban. Clim.* **2019**, *28*, 100464. [[CrossRef](#)]
13. Omarov, B.; Memon, S.A.; Kim, J. A Novel Approach to Develop Climate Classification Based on Degree Days and Building Energy Performance. *Energy* **2023**, *267*, 126514. [[CrossRef](#)]
14. Roshan, G.; Farrokhzad, M.; Attia, S. Climatic Clustering Analysis for Novel Atlas Mapping and Bioclimatic Design Recommendations. *Indoor Built Environ.* **2021**, *30*, 313–333. [[CrossRef](#)]
15. Manzano-Agugliaro, F.; Montoya, F.G.; Sabio-Ortega, A.; García-Cruz, A. Review of Bioclimatic Architecture Strategies for Achieving Thermal Comfort. *Renew. Sustain. Energy Rev.* **2015**, *49*, 736–755. [[CrossRef](#)]
16. Liu, S.; Shi, Q. Local Climate Zone Mapping as Remote Sensing Scene Classification Using Deep Learning: A Case Study of Metropolitan China. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 229–242. [[CrossRef](#)]
17. Kotharkar, R.; Bagade, A. Local Climate Zone Classification for Indian Cities: A Case Study of Nagpur. *Urban. Clim.* **2018**, *24*, 369–392. [[CrossRef](#)]
18. Wicki, A.; Parlow, E. Attribution of Local Climate Zones Using a Multitemporal Land Use/Land Cover Classification Scheme. *J. Appl. Remote Sens.* **2017**, *11*, 026001. [[CrossRef](#)]
19. Nadarajah, P.D.; Singh, M.K.; Mahapatra, S.; Pajek, L.; Košir, M. Bioclimatic Classification for Building Energy Efficiency Using Hierarchical Clustering: A Case Study for Sri Lanka. *J. Build. Eng.* **2023**, *83*, 108388. [[CrossRef](#)]
20. Praene, J.-P.; Malet-Damour, B.; Harimisa Radanielina, M.; Fontaine, L.; Riviere, G.; Philippe Praene, J.; Rivière, G. GIS-Based Approach to Identify Climatic Zoning: A Hierarchical Clustering on Principal Component Analysis GIS-Based Approach to Define Climatic Zoning: A Hierarchical Clustering on Principal Component Analysis. *Build. Environ.* **2019**, *164*, 106330. [[CrossRef](#)]
21. Zscheischler, J.; Mahecha, M.D.; Harmeling, S. Climate Classifications: The Value of Unsupervised Clustering. *Procedia Comput. Sci.* **2012**, *9*, 897–906. [[CrossRef](#)]
22. Li, T.; Rezaeipanah, A.; Tag El Din, E.S.M. An Ensemble Agglomerative Hierarchical Clustering Algorithm Based on Clusters Clustering Technique and the Novel Similarity Measurement. *J. King Saud. Univ. Comput. Inf. Sci.* **2022**, *34*, 3828–3842. [[CrossRef](#)]
23. Bienvenido-Huertas, D.; Marín-García, D.; Carretero-Ayuso, M.J.; Rodríguez-Jiménez, C.E. Climate Classification for New and Restored Buildings in Andalusia: Analysing the Current Regulation and a New Approach Based on k-Means. *J. Build. Eng.* **2021**, *43*, 102829. [[CrossRef](#)]
24. Sinaga, K.P.; Yang, M.S. Unsupervised K-Means Clustering Algorithm. *IEEE Access* **2020**, *8*, 80716–80727. [[CrossRef](#)]
25. Bradley, P.S.; Fayyad, U.M. Refining Initial Points for K-Means Clustering. *ICML* **1998**, *98*, 91–99.
26. Yuan, C.; Yang, H. Research on K-Value Selection Method of K-Means Clustering Algorithm. *J.* **2019**, *2*, 226–235. [[CrossRef](#)]
27. Bechtel, B.; Daneke, C. Classification of Local Climate Zones Based on Multiple Earth Observation Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1191–1202. [[CrossRef](#)]
28. Grupo de Climatología y Agrometeorología-Subdirección de Metereología, IDEAM. Clasificaciones Climaticas Colombia. In Proceedings of the Segundo Congreso Nacional del Clima, Bogotá, Colombia, 3–5 August 2011.
29. IDEAM. *Evolución Del Índice de Confort Térmico Por Periodos 1971–2000. Segunda Comunicación Nacional ante la Convención Marco de las Naciones Unidas sobre Cambio Climático*; Instituto de Hidrología, Meteorología y Estudios Ambientales—IDEAM: Bogotá, Colombia, 2010.
30. Ministerio de Ambiente y Desarrollo Sostenible. *Criterios Ambientales Para El Diseño y Construcción de Vivienda Urbana*; Ministerio de Ambiente y Desarrollo Sostenible: Bogotá, Colombia, 2012; ISBN 9789588491585.
31. Gillies, S. *Rasterio Documentat*, 23rd ed.; MapBox: San Francisco, CA, USA, 2019.
32. Brimicombe, C.; Di Napoli, C.; Quintino, T.; Pappenberger, F.; Cornforth, R.; Cloke, H.L. Thermofeel: A Python Thermal Comfort Indices Library. *SoftwareX* **2022**, *18*, 101005. [[CrossRef](#)]
33. Li, H.; Huang, J.; Hu, Y.; Wang, S.; Liu, J.; Yang, L. A New TMY Generation Method Based on the Entropy-Based TOPSIS Theory for Different Climatic Zones in China. *Energy* **2021**, *231*, 120723. [[CrossRef](#)]
34. Tadić, L.; Bonacci, O.; Brleković, T. An Example of Principal Component Analysis Application on Climate Change Assessment. *Theor. Appl. Climatol.* **2019**, *138*, 1049–1062. [[CrossRef](#)]
35. Jolliffe, I.T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [[CrossRef](#)]
36. Kaiser, H.F. The Application of Electronic Computers to Factor Analysis. *Educ. Psychol. Meas.* **1960**, *20*, 141–151. [[CrossRef](#)]
37. Xiong, J.; Yao, R.; Grimmond, S.; Zhang, Q.; Li, B. A Hierarchical Climatic Zoning Method for Energy Efficient Building Design Applied in the Region with Diverse Climate Characteristics. *Energy Build.* **2019**, *186*, 355–367. [[CrossRef](#)]
38. Dinh, D.T.; Fujinami, T.; Huynh, V.N. Estimating the Optimal Number of Clusters in Categorical Data Clustering by Silhouette Coefficient. In *Knowledge and Systems Sciences. KSS 2019; Communications in Computer and Information Science*; Springer: Singapore, 2019; Volume 1103, pp. 1–17.
39. Brusco, M.J.; Steinley, D. A Comparison of Heuristic Procedures for Minimum Within-Cluster Sums of Squares Partitioning. *Psychometrika* **2007**, *72*, 583–600. [[CrossRef](#)]

40. Semahi, S.; Benbouras, M.A.; Mahar, W.A.; Zemmouri, N.; Attia, S. Development of Spatial Distribution Maps for Energy Demand and Thermal Comfort Estimation in Algeria. *Sustainability* **2020**, *12*, 6066. [[CrossRef](#)]
41. Gupta, R.; Mathur, J.; Garg, V. Assessment of Climate Classification Methodologies Used in Building Energy Efficiency Sector. *Energy Build.* **2023**, *298*, 113549. [[CrossRef](#)]
42. Oliveira, A.; Lopes, A.; Niza, S. Local Climate Zones in Five Southern European Cities: An Improved GIS-Based Classification Method Based on Copernicus Data. *Urban. Clim.* **2020**, *33*, 100631. [[CrossRef](#)]
43. Bechtel, B.; Alexander, P.J.; Böhner, J.; Ching, J.; Conrad, O.; Feddema, J.; Mills, G.; See, L.; Stewart, I. Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities. *ISPRS Int. J. Geoinf.* **2015**, *4*, 199–219. [[CrossRef](#)]
44. Abbasi, F.; Bazgeer, S.; Kalehbasti, P.R.; Oskoue, E.A.; Haghighat, M.; Kalehbasti, P.R. New Climatic Zones in Iran: A Comparative Study of Different Empirical Methods and Clustering Technique. *Theor. Appl. Climatol.* **2022**, *147*, 47–61. [[CrossRef](#)]
45. Sathiaraj, D.; Huang, X.; Chen, J. Predicting Climate Types for the Continental United States Using Unsupervised Clustering Techniques. *Environmetrics* **2019**, *30*, e2524. [[CrossRef](#)]
46. Balogun, I.A.; Daramola, M.T. The Outdoor Thermal Comfort Assessment of Different Urban Configurations within Akure City, Nigeria. *Urban. Clim.* **2019**, *29*, 100489. [[CrossRef](#)]
47. Attia, S.; Lacombe, T.; Rakotondramiarana, H.T.; Garde, F.; Roshan, G.R. Analysis Tool for Bioclimatic Design Strategies in Hot Humid Climates. *Sustain. Cities Soc.* **2019**, *45*, 8–24. [[CrossRef](#)]
48. Daemei, A.B.; Eghbali, S.R.; Khotbehsara, E.M. Bioclimatic Design Strategies: A Guideline to Enhance Human Thermal Comfort in Cfa Climate Zones. *J. Build. Eng.* **2019**, *25*, 100758. [[CrossRef](#)]
49. Li, Z.; Feng, X.; Fan, X.; Sun, J.; Fang, Z. Effect of Direct Solar Projected Area Factor on Outdoor Thermal Comfort Evaluation: A Case Study in Shanghai, China. *Urban. Clim.* **2022**, *41*, 101033. [[CrossRef](#)]
50. Anderson, R.; Bayer, P.E.; Edwards, D. Climate Change and the Need for Agricultural Adaptation. *Curr. Opin. Plant Biol.* **2020**, *56*, 197–202. [[CrossRef](#)] [[PubMed](#)]
51. Kogo, B.K.; Kumar, L.; Koech, R. Climate Change and Variability in Kenya: A Review of Impacts on Agriculture and Food Security. *Environ. Dev. Sustain.* **2021**, *23*, 23–43. [[CrossRef](#)]
52. Attia, S.; Eleftheriou, P.; Xenii, F.; Morlot, R.; Ménézo, C.; Kostopoulos, V.; Betsi, M.; Kalaitzoglou, I.; Pagliano, L.; Cellura, M.; et al. Overview and Future Challenges of Nearly Zero Energy Buildings (NZEB) Design in Southern Europe. *Energy Build.* **2017**, *155*, 439–458. [[CrossRef](#)]
53. Santos-Herrero, J.M.; Lopez-Guede, J.M.; Flores-Abascal, I. Modeling, Simulation and Control Tools for NZEB: A State-of-the-Art Review. *Renew. Sustain. Energy Rev.* **2021**, *142*, 110851. [[CrossRef](#)]
54. Belussi, L.; Barozzi, B.; Bellazzi, A.; Danza, L.; Devitofrancesco, A.; Fanciulli, C.; Ghellere, M.; Guazzi, G.; Meroni, I.; Salamone, F.; et al. A Review of Performance of Zero Energy Buildings and Energy Efficiency Solutions. *J. Build. Eng.* **2019**, *25*, 100772. [[CrossRef](#)]
55. Dnp Colombia, Potencial Mundial De La Vida. *Bases Del Plan Nacional de Desarrollo 2022–2026*; Departamento Nacional de Planeación: Bogotá, Colombia, 2022.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.