

Article

Identification of Cyclone Exposure Regimes in the Southwestern Caribbean Offshore Using Self-Organizing Maps: Implications for Risk-Informed Offshore Wind Planning

Mario Eduardo Carbonó dela Rosa ^{1,2,*}, Adalberto Ospino-Castro ^{1,*}, Carlos Robles-Algarín ³,
Diego Restrepo-Leal ⁴ and Victor Olivero-Ortiz ³

¹ Departamento de Energía, Universidad de la Costa, Calle 58 No. 55-66, Barranquilla 080002, Colombia

² Escuela de Ciencias Básicas Ingeniería y Tecnología, Universidad Nacional Abierta y a Distancia, Calle 14 Sur #14-23, Bogotá 111411, Colombia

³ Facultad de Ingeniería, Universidad del Magdalena, Carrera 32 No. 22-08, Santa Marta 470004, Colombia; croblesa@unimagdalena.edu.co (C.R.-A.); volivero@unimagdalena.edu.co (V.O.-O.)

⁴ Facultad de Ingenierías, Universidad Cooperativa de Colombia, Santa Marta 470002, Colombia; diego.restrepoleal@campusucc.edu.co

* Correspondence: mario.carbono@unad.edu.co (M.E.C.d.R.); aospino8@cuc.edu.co (A.O.-C.)

Abstract

Offshore wind development in the Colombian Caribbean requires characterization of tropical cyclone exposure to support climate-resilient planning, yet objective regime classifications for this basin remain limited. This study develops an unsupervised Self-Organizing Map (SOM) framework using the NOAA HURDAT2 best-track database for 1970–2023, comprising 668 six-hourly observations from 103 storms and nine physically interpretable features describing position, intensity, kinematics, and seasonality. A 4×4 SOM combined with hierarchical Ward clustering identified four cyclone exposure regimes: R0, late-season slow north-eastward disturbances near San Andrés (11.5%); R1, fast westward mature tropical storms across the central Caribbean (62.0%); R2, slow north-westward disturbances off Central America (21.9%); and R3, rare major hurricanes crossing the central Caribbean (4.6%, seven storms; historical mean recurrence interval 7.7 years). R3 exhibited mean maximum sustained winds of 67.8 m/s and a minimum central pressure of 938.5 hPa, with peak winds reaching 79.7 m/s. The results provide a complementary historical hazard characterization layer for offshore wind planning, supporting risk-informed spatial and operational assessment. However, the identified regimes do not represent wind-resource conditions or probabilistic design return periods and should therefore be integrated with site-specific wind, wave, and structural assessments.

Keywords: Self-Organizing Maps; tropical cyclones; unsupervised learning; hazard characterization; Atlantic Caribbean



Academic Editor: Adrian Ilinca

Received: 11 August 2026

Revised: 8 September 2026

Accepted: 9 September 2026

Published: 11 September 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Offshore wind energy has consolidated as a central pillar of the global energy transition, with global installed capacity exceeding 68 GW by 2023 and projections indicating that it could reach 2000 GW by 2050 [1]. Although the industry has historically been concentrated in the North and Baltic Seas, recent technical and policy developments have extended its geographical frontier towards tropical and sub-tropical basins, including the U.S. Gulf of Mexico, the East China Sea, the Bay of Bengal, and the Caribbean Sea [2]. In Colombia, the national Offshore Wind Roadmap released by the Ministry of Mines and Energy estimated

a technical resource potential of approximately 109 GW along the Caribbean coast [3], and subsequent techno-economic assessments suggest that the Caribbean offshore wind resource could satisfy a substantial fraction of the country's projected electricity demand by 2039 [4,5]. Pilot project areas have been identified off the departments of Bolívar, Atlántico, Magdalena and the archipelago of San Andrés, Providencia and Santa Catalina, where the Seaflower Biosphere Reserve hosts marine areas of high ecological value [6].

Unlike most tropical regions, the Colombian Caribbean combines a persistent low-level jet with relatively shallow bathymetry, conditions that are favourable for offshore wind development. However, the same basin is also periodically traversed by tropical cyclones that originate in the Main Development Region of the Atlantic and propagate westward into the Caribbean Sea [7]. While the western Caribbean has historically been less affected than the Greater Antilles or the U.S. Gulf Coast, major hurricanes including Joan (1988), Lenny (1999), Beta (2005), Iota (2020) and, more recently, Beryl (2024) have demonstrated that the region is not exempt from intense cyclone impacts [8]. Recent analyses indicate that Providencia lost approximately 98% of its infrastructure during Hurricane Iota in 2020, illustrating the vulnerability of Colombian offshore and island territories to extreme cyclonic events [9].

Tropical cyclones pose a multifaceted threat to offshore wind infrastructure through extreme sustained winds, gusts, wave loading, storm surge and rapid changes in wind direction [10]. The design of offshore wind turbines is governed by the IEC 61400-1 standard, which defines design load cases for different turbine classes characterized by reference wind speeds (V_{ref}) of 50 m/s (Class I), 42.5 m/s (Class II) and 37.5 m/s (Class III) at hub height [11]. Recent fragility analyses indicate that probabilities of turbine yielding and buckling under 20-year return-period tropical cyclones may increase by approximately 37% and 13%, respectively, in a non-stationary climate, with regional increases reaching up to 51% in the U.S. Atlantic offshore lease areas [12]. Comparable findings have been reported for the U.S. East Coast and for floating offshore wind mooring systems in tropical cyclone environments [13]. These findings highlight the importance of robust cyclone hazard characterization as part of climate-informed risk assessment for offshore wind project planning and design [14].

Despite recent advances in cyclone hazard assessment, the Colombian Caribbean offshore still lacks a data-driven framework capable of objectively identifying recurrent cyclone exposure patterns from historical observations. Previous studies have primarily relied on spectral analyses, kernel density estimation of cyclone tracks, or composite hazard indicators, which effectively describe temporal variability or spatial occurrence but treat cyclone characteristics independently and do not explicitly resolve the multivariate relationships among intensity, storm kinematics, geographical location, and seasonal variability [15]. As offshore wind planning increasingly requires integrated climate-risk information to support resilient infrastructure development, methodologies capable of discovering recurrent exposure regimes directly from historical observations have become increasingly relevant. In this context, unsupervised learning offers the possibility of identifying physically consistent cyclone regimes without imposing predefined hazard classes, thereby providing an objective basis for regional hazard characterization and supporting future multi-criteria offshore wind planning frameworks [16]. Such planning should also account for wind-farm layout and wake interactions when assessing offshore wind performance [17].

Self-Organizing Maps (SOMs) are an unsupervised neural-network technique that projects high-dimensional data onto a low-dimensional, topology-preserving grid [18]. SOMs have been extensively used in synoptic climatology, oceanography, and renewable-energy meteorology to identify recurring atmospheric and oceanic patterns without imposing a priori assumptions on the number or shape of the underlying classes [19,20].

Recent applications include the identification of coastal wind regimes relevant to off-shore wind power production in the United States, the classification of synoptic wind regimes near Kuwait, the analysis of typhoon tracks and associated flood hazards, and the automated identification of large-scale meteorological patterns associated with cyclogenesis [21]. Comparative studies have shown that SOM provide a viable complement to K-means for tropical cyclone track classification in South Korea [22]. These applications indicate that SOM are particularly well suited to multivariate, continuous-but-overlapping phenomena such as tropical cyclones, where physical regimes are not sharply separated but transition gradually [23].

HURDAT2, the Atlantic Hurricane Database 2nd generation maintained by the U.S. National Hurricane Center, provides the most complete and quality-controlled record of tropical cyclone best-track information for the Atlantic basin, including latitude, longitude, maximum sustained wind, and minimum central pressure at six-hourly intervals since 1851 [24]. HURDAT2 is, by construction, a record of extreme weather events and does not represent the mean wind regime at any location. Consequently, HURDAT2 should not be used to estimate wind resource, wind power density, or annual energy production, all of which require continuous reanalysis time series such as ERA5 or MERRA-2, or dedicated products such as the Global Wind Atlas [25]. Using HURDAT2 maximum sustained winds as a proxy for ambient wind resource would lead to a substantial conceptual misinterpretation, because event-based extreme winds do not represent the long-term wind climate required for resource assessment. Within its intended scope, however, HURDAT2 offers a physically consistent, multivariate description of tropical cyclone behaviour that can support hazard characterization and risk-informed planning [26].

Building on these considerations, this study develops and evaluates an unsupervised learning framework based on SOMs to identify and characterize historical cyclone exposure regimes in the Colombian Caribbean offshore using HURDAT2, providing a complementary hazard characterization layer for climate-resilient offshore wind infrastructure planning. The specific contributions are: (i) a reproducible SOM-based workflow (MiniSom, Ward hierarchical clustering, multi-metric selection) applied to a curated set of nine physically interpretable HURDAT2-derived features for the 1970–2023 period; (ii) the identification of four physically interpretable cyclone exposure regimes characterized by distinct intensity, kinematics, geographical position, and seasonality; (iii) a within-storm regime transition matrix that captures the temporal dynamics of regime membership along individual cyclone tracks; and (iv) an explicit articulation of how the resulting regimes complement rather than replace wind-resource assessments based on reanalysis or wind atlases for offshore wind planning in the Colombian Caribbean.

2. Materials and Methods

2.1. Study Area

The study area corresponds to the south-western sector of the Caribbean Basin delimited by 7–16° N and 67–82° W; see Figure 1. This window encompasses the Colombian Caribbean offshore domain, including the continental shelf off the departments of La Guajira, Magdalena, Atlántico and Bolívar, the insular domain of the Seaflower Biosphere Reserve around the archipelago of San Andrés, Providencia and Santa Catalina, and the south-western Caribbean open waters where tropical cyclones transit towards Central America [27]. The boundaries are meteorologically coherent and were chosen to capture all tropical cyclones whose tracks may generate extreme environmental conditions relevant to Colombian offshore infrastructure, while excluding systems confined to the central Atlantic Main Development Region or to the Greater Antilles that do not enter the region of interest [28].

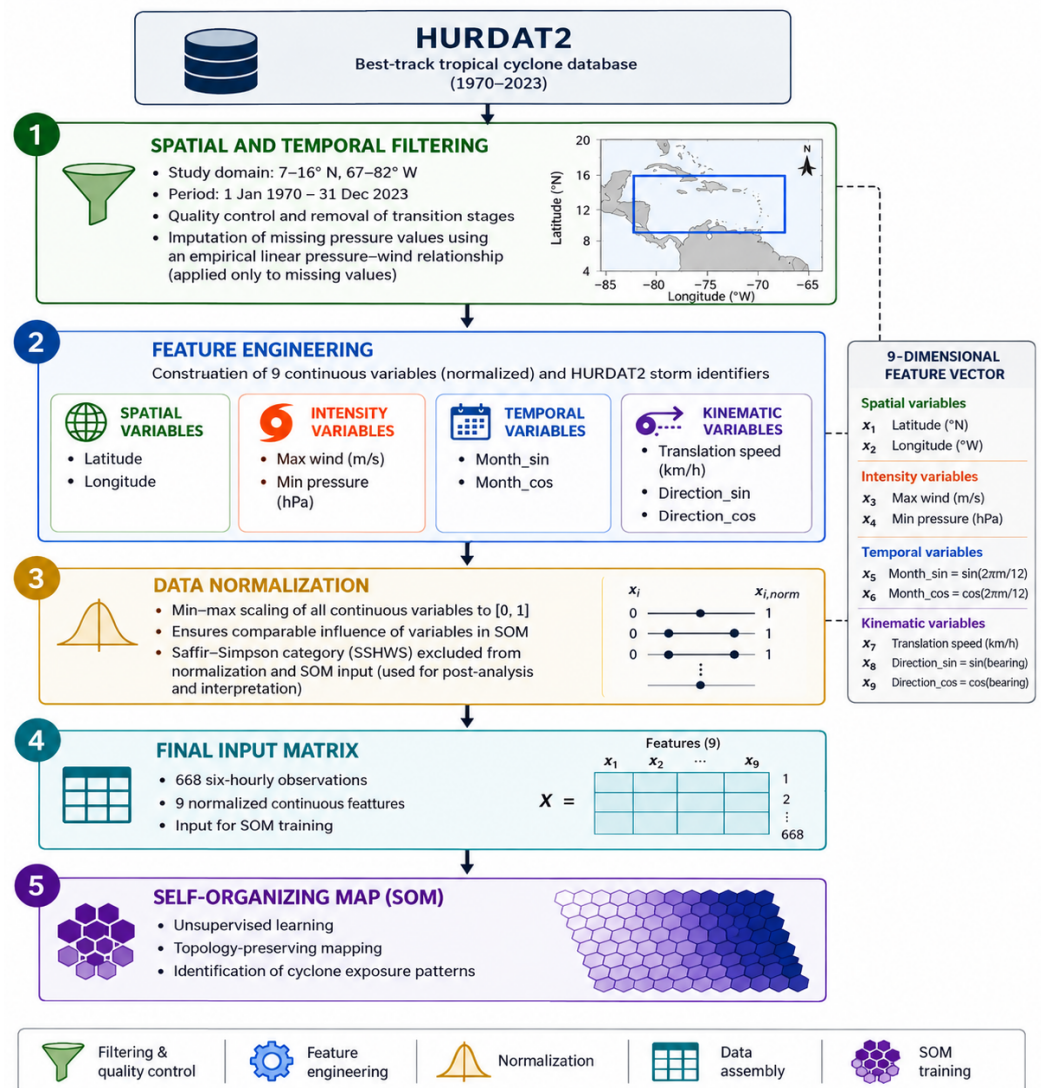


Figure 1. Overview of the HURDAT2 data preprocessing workflow used to construct the normalized feature matrix for Self-Organizing Map (SOM) training. The workflow includes spatial and temporal filtering, feature engineering, min–max normalization, and assembly of the final nine-dimensional input feature vector. The blue box indicates the study domain (7–16° N, 67–82° W) used for spatial filtering of HURDAT2 cyclone observations.

The selected window is wider than the Colombian Exclusive Economic Zone. This choice is deliberate: restricting the analysis to the exclusive economic zone would exclude cyclones transiting through the broader south-western Caribbean that can produce extreme winds, waves, and storm surge affecting Colombian waters [29]. The boundaries are identical to those of the precursor study of the same research group, ensuring direct comparability between the unsupervised classification presented here and the spectral and kernel-density results previously reported.

2.2. HURDAT2 Database and Data Preprocessing

The Atlantic Hurricane Database (HURDAT2), maintained by the U.S. National Hurricane Center (NHC) in collaboration with the Atlantic Oceanographic and Meteorological Laboratory (AOML), was used as the exclusive source of historical tropical cyclone observations for this study. HURDAT2 provides six-hourly best-track records for each Atlantic tropical cyclone since 1851, including latitude, longitude, system status (tropical depression, tropical storm, hurricane, extra-tropical, disturbance), maximum sustained 1-min surface

wind speed (in knots), and minimum central sea-level pressure (in hPa). The database undergoes continuous post-season reanalysis, incorporating updated aircraft reconnaissance, satellite retrievals, in situ observations, and historical record revisions to provide the most reliable best-track representation of tropical cyclone evolution currently available for the Atlantic basin [30].

The analysis period spans 1970–2023, providing a 54-year historical record that encompasses the progressive development of satellite-based tropical cyclone observations, including the establishment of operational geostationary meteorological satellite observations during the 1970s and subsequent improvements in spatial, temporal, and observational coverage. Second, the 54-year record encompasses multiple phases of interannual and multidecadal climate variability associated with the El Niño–Southern Oscillation (ENSO) and the Atlantic Multidecadal Oscillation (AMO), both of which modulate Caribbean tropical cyclone activity [31]. To ensure a robust and internally consistent dataset for unsupervised learning, the original HURDAT2 archive underwent a sequential preprocessing workflow comprising spatial and temporal filtering, feature engineering, and feature normalization prior to SOM training. This workflow was designed to preserve the physical consistency of cyclone observations while generating a standardized multivariate feature space suitable for topology-preserving clustering.

Because the observational basis of Atlantic tropical cyclone best-track records has evolved over time, particularly with the increasing availability of satellite observations and improvements in reconnaissance and post-storm analysis, temporal consistency was considered explicitly in the present study. HURDAT2 represents a continuously revised best-track record that integrates available observational information through retrospective analysis and reanalysis. The nine variables used in the SOM are either directly reported in HURDAT2 or deterministically derived from its six-hourly position, intensity, and timing information; no external reanalysis variables were introduced into the clustering. To assess whether temporal differences in the observational environment substantially affected the identified regimes, a sensitivity analysis was performed by comparing the reference 1970–2023 solution with independently trained SOM–Ward solutions for the 1970–1996 and 1997–2023 subperiods, using the same feature definitions, normalization procedure, SOM configuration, and clustering strategy. The resulting regime assignments were compared using the Adjusted Rand Index (ARI). This analysis was considered a robustness assessment and does not assume strict observational homogeneity across the complete historical record.

2.2.1. Spatial and Temporal Filtering

Since the objective of this study is to characterize historical cyclone exposure regimes relevant to the Colombian Caribbean offshore, a spatial and temporal filtering procedure was applied to isolate only those best-track records capable of influencing the study region. This approach ensures that the subsequent unsupervised classification is based exclusively on cyclone observations representative of the regional hazard environment, while excluding systems whose trajectories remained outside the area of potential influence on offshore infrastructure.

Six-hourly best-track records were first filtered according to the geographical limits of the study area (7–16° N, 67–82° W) and the analysis period (1 January 1970–31 December 2023). This procedure retained all cyclone positions occurring within the regional domain regardless of storm intensity or developmental stage, thereby preserving the complete evolution of each system while traversing the Colombian Caribbean.

Records with missing maximum sustained wind speed were excluded because maximum sustained wind is a primary intensity descriptor required for the multivariate charac-

terization. For records with valid maximum sustained wind but missing minimum central pressure, the missing pressure values were estimated using the following empirical linear pressure–wind relationship:

$$P_{min} = 1010 - 0.55V_{max}, \quad (1)$$

where P_{min} is the estimated minimum central pressure (hPa) and V_{max} is the maximum sustained wind speed (kt). The relationship was applied exclusively to missing pressure values, and observed HURDAT2 pressure values were not modified. This procedure affected 152 of the 668 regional track-point observations (22.75%), while no observations were removed because of missing maximum sustained wind. The uncertainty associated with the empirical relationship was assessed by comparing estimated and observed pressure values for records with available HURDAT2 pressure measurements. This comparison yielded a mean absolute error (MAE) of 12.46 hPa and a root mean square error (RMSE) of 13.23 hPa. A sensitivity analysis was also performed by excluding all observations with imputed pressure and repeating the SOM–Ward analysis to assess whether the identified exposure regimes, particularly the extreme R3 regime, were sensitive to pressure imputation. Extratropical, subtropical, and dissipating stages were intentionally preserved whenever they occurred within the study domain, allowing the complete life-cycle transitions of individual cyclones to be represented during the subsequent regime identification process. After applying the filtering and quality-control procedures, the primary dataset comprised 668 six-hourly best-track records corresponding to 103 individual tropical cyclones.

2.2.2. Feature Engineering

Feature engineering aimed to transform the original HURDAT2 best-track records into a compact yet physically representative feature space suitable for unsupervised learning. Following the principle of parsimony, variables were selected based on their physical interpretability, complementary information content, and direct derivation from HURDAT2. Rather than maximizing the number of predictors, the feature selection strategy sought to preserve the principal dimensions governing tropical cyclone behavior while limiting unnecessary redundancy and avoiding variables requiring external datasets. Consequently, the final SOM input vector comprises nine continuous variables describing four complementary dimensions of cyclone dynamics: geographical position, storm intensity, translational kinematics, and seasonal variability. Together, these variables provide a physically interpretable representation of cyclone evolution while maintaining computational efficiency and reproducibility. Table 1 summarizes the selected variables together with their physical interpretation and derivation.

Geographical position was represented by latitude (°N) and longitude (°W) as reported by HURDAT2. Intensity was represented by maximum sustained wind speed (max_wind_ms , converted from knots to m/s using the factor 0.5144) and minimum central pressure ($min_pressure_mb$, in hPa). The two variables are physically coupled through the pressure–wind relationship, but were retained jointly because they provide complementary information on cyclone intensity and structure [32].

Seasonality was encoded through the sine and cosine of the month ($month_sin = \sin(2\pi \cdot m/12)$, $month_cos = \cos(2\pi \cdot m/12)$, where m is the calendar month). This circular encoding preserves the continuity between December and January, avoiding the artificial discontinuity that would arise from using the month as a linear integer [33].

Table 1. HURDAT2-derived features used as input to the Self-Organizing Map.

Feature	Symbol	Origin/Formula	Physical Role in SOMs
Latitude	latitude	HURDAT2 native	Meridional position of the cyclone centre
Longitude	longitude	HURDAT2 native	Zonal position of the cyclone centre
Maximum sustained wind	max_wind_ms	HURDAT2 native; converted from knots to m/s	Cyclone intensity (1-min sustained wind)
Minimum central pressure	min_pressure_mb	HURDAT2 native; empirical pressure–wind relationship applied to missing values (152/668 observations, 22.75%)	Complementary intensity descriptor
Month (sine component)	month_sin	$\sin(2\pi \cdot m/12)$	Circular seasonality
Month (cosine component)	month_cos	$\cos(2\pi \cdot m/12)$	Circular seasonality
Translation speed	translation_speed_kmh	Haversine distance/6 h (central difference)	Cyclone translational kinematics
Movement direction (sine)	movement_direction_sin	$\sin(\text{bearing})$	Circular heading (avoids 0°/360° discontinuity)
Movement direction (cosine)	movement_direction_cos	$\cos(\text{bearing})$	Circular heading

Cyclone translational speed was estimated from consecutive six-hourly best-track positions using the Haversine great-circle distance divided by the corresponding time interval. The same great-circle distance formulation was used to calculate cumulative storm track length within the study domain from the sequence of consecutive six-hourly positions [34].

Storm heading was computed as the geodetic bearing between consecutive track positions and encoded through sine and cosine transformations to preserve the circular nature of directional data and avoid discontinuities at the 0°/360° boundary.

The Saffir–Simpson Hurricane Wind Scale (SSHWS) category was derived exclusively for post-clustering interpretation of the identified regimes. Because SOM training relies on continuous Euclidean distances, the categorical SSHWS classification was intentionally excluded from the input feature space to prevent distortions in neuron competition and topology preservation.

Collectively, the selected variables provide a multivariate representation of cyclone behavior that simultaneously captures where storms occur, how intense they become, how they move, and when they develop. This integrated representation constitutes the feature space used for SOM training in the subsequent stage of the proposed framework.

2.2.3. Data Normalization

Prior to SOM training, all continuous variables were normalized to the [0, 1] interval using min–max scaling to ensure that variables with different numerical ranges contributed equally to the Euclidean distance calculations governing neuron competition [35]. Min–max normalization was selected because it preserves the relative distribution and geometric relationships among observations while preventing high-magnitude variables from disproportionately influencing the SOM learning process. This approach is widely adopted in topology-preserving neural networks, where maintaining proportional distances between observations is essential for reliable clustering and map organization [36]. Because the

Saffir–Simpson Hurricane Wind Scale represents an ordinal classification rather than a continuous physical variable, it was intentionally excluded from the normalization procedure and from the SOM input space. Instead, it was reserved exclusively for the post hoc interpretation and physical characterization of the identified cyclone exposure regimes. The complete preprocessing workflow adopted to generate the normalized feature matrix used for SOM training is summarized in Figure 1.

2.3. Self-Organizing Map Framework

Self-Organizing Maps (SOMs) were selected because tropical cyclone observations exhibit continuous transitions between physical states rather than forming well-separated clusters. Unlike conventional partitioning algorithms such as K-means, SOMs simultaneously perform dimensionality reduction and topology-preserving clustering, enabling neighbouring neurons to represent gradual changes in cyclone characteristics. This property makes SOMs particularly suitable for representing multivariate atmospheric phenomena whose boundaries are inherently diffuse rather than discrete [37].

The SOM algorithm projects a multidimensional feature space onto a two-dimensional lattice while preserving the topological relationships among neighboring observations. During training, each input vector is assigned to its Best Matching Unit (BMU), whose weight vector and surrounding neurons are iteratively adjusted through a Gaussian learning function with progressively decreasing learning rate and influence radius [38]. This competitive learning process produces an ordered representation of the input space in which cyclone observations with similar characteristics are mapped to adjacent regions of the lattice.

SOM training was implemented using the MiniSom library (version 2.3.6). A two-stage training procedure was used. First, candidate lattice sizes from 4×4 to 8×8 were screened using deterministic random-weight initialization, batch training, 500 training iterations, and a fixed random seed (`random_seed = 42`) [39]. The candidate configuration minimizing the normalized composite score defined in Equation (2) was retained. The selected lattice was then retrained for the final analysis using PCA-based weight initialization (`som.pca_weights_init`), batch training, 2000 training iterations, and the same fixed random seed. The Gaussian neighbourhood function was used throughout, with the initial sigma and learning rate set according to the specified MiniSom configuration [40].

SOM Size Selection

Candidate lattice dimensions were selected to balance representation capacity and model interpretability. Smaller lattices may oversimplify the multidimensional structure of cyclone observations, whereas excessively large lattices increase model complexity and reduce the interpretability of the subsequent hierarchical clustering. The evaluated range (4×4 – 8×8) therefore provides an appropriate compromise for the size of the available dataset while remaining consistent with recommendations reported in previous climatological SOM applications.

The SOM lattice size was selected through a systematic grid search over five candidate grids (4×4 , 5×5 , 6×6 , 7×7 , and 8×8). For each candidate, the Quantization Error (QE) and Topographic Error (TE) were computed. QE was used to evaluate the representation accuracy of each candidate lattice, whereas TE assessed the preservation of neighborhood relationships within the projected feature space [41]. Together, these complementary metrics quantify representation fidelity and topological organization, two key aspects of SOM quality assessment. To enable their combined use in the lattice-selection criterion, QE and TE were normalized to a common scale in this study, thereby preventing differences in

numerical magnitude from disproportionately influencing the composite score [42]. The normalized composite score S was calculated as:

$$S = \frac{QE}{QE_{min}} + \frac{TE}{TE_{min}} \quad (2)$$

where QE_{min} and TE_{min} denote the minimum QE and TE values, respectively, observed across the candidate lattice sizes. The composite score provides a joint criterion that balances representation accuracy and topological preservation, thereby avoiding the selection of a configuration based on either metric alone. To assess the sensitivity of the subsequent regime identification to SOM lattice size, the complete SOM–Ward procedure was repeated for the candidate 4×4 , 5×5 , 6×6 , 7×7 , and 8×8 configurations using the same feature normalization, PCA-based initialization, training parameters, and fixed random seed. For this sensitivity analysis, the number of regimes was fixed at $k = 4$ to isolate the effect of SOM lattice size from the selection of the regime count. The resulting observation-level regime assignments were compared with the reference 4×4 solution using the Adjusted Rand Index (ARI), while the relative regime frequencies and physical characteristics were examined to determine whether the main cyclone-exposure patterns were preserved across SOM configurations. The lattice size minimizing the composite score (S) was selected as the optimal SOM configuration for subsequent hierarchical clustering. The complete optimization procedure adopted for lattice selection is illustrated in Figure 2.

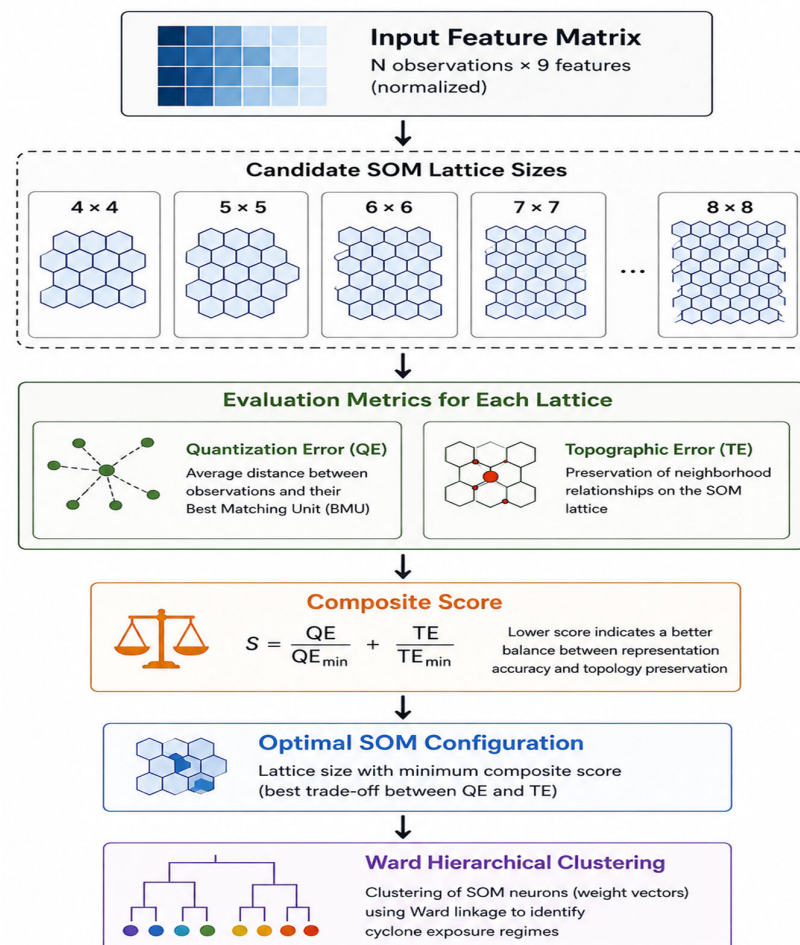


Figure 2. SOM optimization framework based on multi-metric evaluation. Candidate lattice sizes were assessed through Quantization Error (QE) and Topographic Error (TE), and the optimal topology was selected using a normalized composite score before applying Ward hierarchical clustering.

2.4. Hierarchical Clustering and Regime Identification

Although the SOM organizes cyclone observations according to their topological similarity, neighboring neurons do not necessarily correspond to physically distinct exposure regimes. Therefore, a second clustering stage was applied to aggregate neighboring neurons into a reduced number of interpretable cyclone exposure regimes. Agglomerative hierarchical clustering with Ward linkage was selected because it minimizes within-cluster variance while preserving the topological organization learned during SOM training [43]. Ward linkage minimizes the within-cluster variance at each merge and is well suited to SOM neurons, which are already organized on a smooth topological grid. The clustering was performed on the neuron weight vectors (i.e., the prototypes learned by the SOM), not directly on the raw observations, following the established two-stage SOM–hierarchical clustering workflow.

The hierarchical clustering was performed exclusively on the SOM neuron weight vectors rather than on the original observations, following the well-established two-stage SOM–hierarchical clustering framework. This strategy combines the topology-preserving representation generated by the SOM with the interpretability of hierarchical clustering, while reducing the sensitivity of conventional agglomerative clustering to noise, outliers, and local data irregularities present in the original feature space.

Selection of the Number of Regimes

Selecting an appropriate number of cyclone exposure regimes is essential for balancing descriptive parsimony with physical interpretability. Excessive partitioning may produce fragmented regimes with limited physical distinction, whereas an insufficient number of regimes may obscure relevant differences in cyclone characteristics. Accordingly, candidate partitions with k ranging from 2 to 7 were evaluated using the Silhouette Score and Davies–Bouldin Index (DBI) on the active SOM neuron weight vectors after Ward hierarchical clustering. Following the clustering of SOM neurons, each six-hourly observation was assigned to the regime corresponding to its Best Matching Unit (BMU). The two indices were considered jointly because they provide complementary information on cluster separation and compactness, and neither was used as an absolute acceptance threshold. The final number of regimes was selected based on the combined statistical evidence and the physical coherence of the resulting cyclone-exposure patterns. In particular, solutions producing highly imbalanced or physically indistinguishable regimes were not retained.

The stability of the regime configurations was further assessed through cyclone-level bootstrap resampling. Complete trajectories of the 103 cyclones were used as resampling units to preserve the temporal dependence among consecutive six-hourly observations. For each of 500 bootstrap replicates, 103 cyclones were resampled with replacement, and the 4×4 SOM was retrained using the same normalization, PCA initialization, training parameters, and 2000 iterations as in the primary analysis. Ward hierarchical clustering was then applied for $k = 2–7$. The resulting regime labels were propagated to the complete set of 668 original observations and compared with the reference partition using the Adjusted Rand Index (ARI), providing a sensitivity assessment of the candidate regime configurations to sampling variability.

This procedure combines complementary measures of cluster separation and compactness and provides an explicit, reproducible criterion for balancing statistical quality and physical interpretability. Consequently, it reduces the likelihood of selecting either overly simplified partitions dominated by a few extreme observations or excessively fragmented solutions lacking physical significance, as may occur when relying on a single validation criterion and avoids two failure modes of single-criterion selection: the degenerate partition produced by maximizing the Silhouette Score alone (typically separating a

small subset of extreme observations from the bulk of the data) and the over-segmentation produced by minimizing the Davies–Bouldin index alone (which may produce physically uninterpretable partitions at large k). This two-stage procedure combines complementary measures of clustering quality because the Silhouette Score evaluates cluster cohesion and separation, whereas the Davies–Bouldin index quantifies cluster compactness relative to inter-cluster similarity.

Once the optimal partition was determined, each six-hourly cyclone observation inherited the regime label associated with its corresponding Best Matching Unit (BMU). Because the 668 observations represent sequential six-hourly positions from 103 individual cyclones, the observations are not statistically independent and storms with longer trajectories within the study domain contribute a larger number of observations to the point-level analysis and, consequently, have greater representation in the SOM training. To assess the sensitivity of the identified regimes to this unequal contribution, a deterministic storm-balanced sensitivity analysis was performed in addition to the primary point-level analysis. All 668 observations were retained, but each observation was assigned a weight inversely proportional to the number of observations contributed by its corresponding cyclone. For an observation i belonging to cyclone s , the storm-balanced weight was defined as:

$$w_i = \frac{\frac{1}{n_s}}{\frac{1}{N} \sum_{j=1}^N \left(\frac{1}{n_{s_j}} \right)}, \quad (3)$$

where n_s denotes the number of six-hourly observations contributed by cyclone s , and N is the total number of observations. The normalization ensures a mean observation weight of one. Consequently, the total contribution of each cyclone is equalized according to:

$$\sum_{i \in s} w_i = \frac{N}{S}. \quad (4)$$

where S is the number of cyclones. The storm-balanced weights were subsequently used as a sensitivity analysis of regime occurrence, while the original SOM–Ward partition was retained as the primary clustering solution. Weighted regime frequencies were calculated by assigning each cyclone an equal total contribution, thereby reducing the influence of storms with longer trajectories. This analysis was used to assess whether the relative occurrence of the identified regimes was substantially affected by unequal numbers of six-hourly observations per storm. The storm-level predominant regime was also retained as an independent event-based characterization, in which each cyclone contributes once to the regime distribution. The six-hourly point-level analysis remained the primary representation of instantaneous cyclone exposure.

Following regime identification, labels were also propagated to the storm level for descriptive analyses of cyclone evolution and regional exposure patterns. For storm-level descriptive statistics, the nominal duration of each cyclone within the study domain was calculated from the number of available six-hourly observations as:

$$D_s = (n_s - 1) \times 6h, \quad (5)$$

where n_s is the number of observations available for cyclone s within the study domain. This metric represents the nominal temporal span implied by the six-hourly sampling sequence rather than the complete lifetime of the cyclone. Track length within the study domain was calculated as the cumulative Haversine great-circle distance between consecutive six-hourly cyclone positions belonging to the same storm. Thus, for cyclone s .

$$L_s = \sum_{i=0}^{n_s-1} d_H(P_i, P_{i+1}) \quad (6)$$

where d_H denotes the Haversine distance between consecutive positions and n_s is the number of available observations within the study domain. This metric represents the cumulative spatial distance travelled by the cyclone within the analyzed domain rather than the total length of its complete lifetime track.

2.5. Model Evaluation

Model performance was evaluated using four complementary metrics that jointly assess representation accuracy, topological preservation, and clustering quality within the normalized feature space. Quantization Error (QE) measures the average Euclidean distance between each observation and its Best Matching Unit (BMU), providing an estimate of representation accuracy. Topographic Error (TE) quantifies the preservation of neighborhood relationships during projection onto the SOM lattice. Clustering quality was subsequently assessed using the Silhouette Score, which evaluates cluster cohesion and separation, together with the Davies–Bouldin Index, which measures cluster compactness relative to inter-cluster similarity [44]. These metrics provide complementary information on SOM representation fidelity, topological preservation, and internal cluster structure. They were used to evaluate the quality of the selected configuration and to compare candidate regime partitions, rather than as independent evidence of statistical robustness.

Computational reproducibility was ensured through deterministic model initialization, including a fixed random seed (RANDOM_SEED = 42) consistently applied to both NumPy and MiniSom, PCA-based weight initialization and batch training, which removes dependence on the order in which observations are presented during sequential updates.

2.6. Scope and Limitations of the Data Source

HURDAT2 is a best-track database specifically designed to document the historical evolution of tropical cyclones in the Atlantic basin. Consequently, it provides physically consistent information for cyclone hazard characterization but is not intended to represent the long-term ambient wind climate or the offshore wind resource at a given location. Accordingly, the proposed framework should not be interpreted as a resource assessment methodology and does not estimate parameters such as wind resource, annual energy production, levelized cost of electricity, or optimal wind-farm locations. These analyses require continuous atmospheric reanalysis products or dedicated mesoscale numerical simulations rather than event-based cyclone records.

Instead, the proposed framework provides a complementary hazard characterization layer describing the historical spatial, temporal, and dynamical behavior of tropical cyclones within the Colombian Caribbean offshore. This information is intended to support climate-resilient offshore wind planning by incorporating cyclone exposure as an additional decision variable alongside conventional wind resource assessments derived from ERA5, MERRA-2, the Global Wind Atlas, or dedicated mesoscale simulations [45]. Accordingly, the engineering implications discussed in Sections 4 and 5 should be interpreted within a risk-informed planning framework, where historical cyclone exposure complements conventional resource assessment to support more resilient offshore wind development, rather than serving as a direct indicator of energy production potential.

2.7. Use of Generative Artificial Intelligence Tools

Generative Artificial Intelligence (GenAI) was used solely to assist in the generation and visual design of Figures 1 and 2. The scientific methodology, data processing, feature selection, model development, statistical analyses, results, and interpretation were developed and verified by the authors independently of the GenAI tool. The generated figures were reviewed and edited by the authors to ensure consistency with the methodology and results reported in the manuscript.

3. Results

3.1. SOMs and Clustering Metrics

The grid search over SOM lattice sizes identified the 4×4 configuration (16 neurons) as the optimal compromise between Quantization Error (QE) and Topographic Error (TE). During this screening stage, each candidate lattice was trained for 500 iterations using deterministic random-weight initialization, and the composite score defined in Equation (2) was minimized for the 4×4 configuration. The corresponding grid-search TE was 0.2620, compared with 0.5045 for the 8×8 lattice; see Figure 3. As expected, QE decreased with increasing lattice size, reflecting greater representational resolution, whereas TE varied across configurations, indicating differences in topological preservation. The 4×4 lattice was therefore retained for the final analysis.

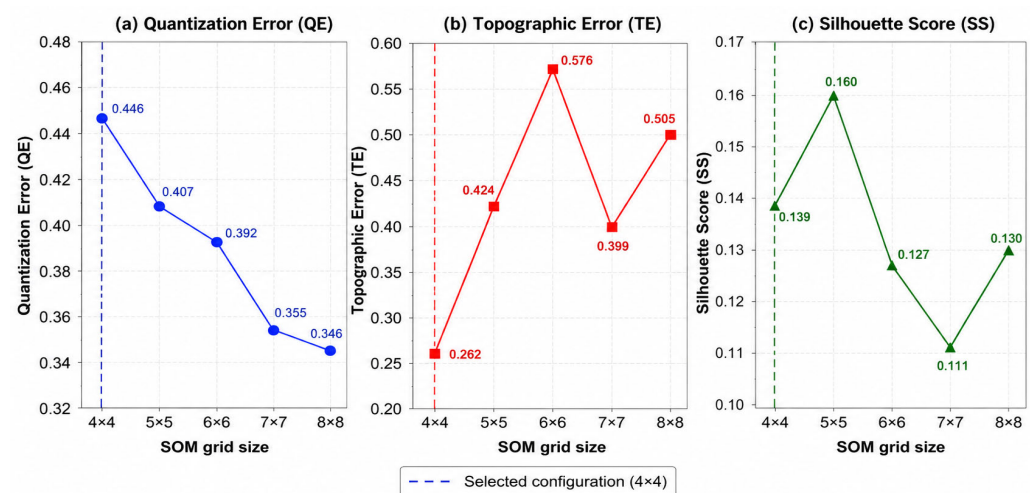


Figure 3. Evaluation of candidate SOM lattice sizes (4×4 – 8×8) using. The blue, red, and green dashed vertical lines in panels (a), (b), and (c), respectively, indicate the selected 4×4 SOM configuration. The 4×4 lattice was selected as the optimal configuration based on the minimum normalized composite score, providing the adopted balance between representation accuracy and topological preservation.

The selected 4×4 SOM was subsequently retrained using PCA-based weight initialization and 2000 training iterations, yielding a final QE of 0.5043 and TE of 0.3533; see Table 2. The latter values correspond to the final SOM configuration used for hierarchical clustering and regime identification. In the present dataset, the TE of 0.3533 indicates that approximately 35% of the observations do not preserve adjacent first- and second-BMU relationships. This result indicates imperfect topological preservation and is consistent with the presence of overlapping and gradually varying cyclone exposure states.

To assess the sensitivity of the identified regime structure to SOM lattice size, the complete SOM–Ward procedure was repeated for the 4×4 , 5×5 , 6×6 , 7×7 , and 8×8 configurations while maintaining ($k = 4$) and the same normalization, PCA-based initialization, training parameters, and random seed. The resulting observation-level partitions were compared with the reference 4×4 solution using the Adjusted Rand Index (ARI). The ARI values were 0.661, 0.334, 0.554, and 0.615 for the 5×5 , 6×6 , 7×7 , and 8×8 configurations, respectively, indicating moderate agreement for most larger lattices, although some sensitivity in individual observation assignments was observed. Importantly, the larger configurations preserved distinct cyclone-exposure patterns, including a high-intensity cluster reaching the same maximum wind speed of 79.73 m/s observed in the reference solution. These results indicate that the principal exposure structure is reasonably robust to SOM lattice size, while the 4×4 configuration provides a parsimonious representation

with an adequate balance between representational resolution, topological organization, and the complexity of the subsequent regime analysis.

Table 2. Self-Organizing Map configuration and quality metrics for the selected 4×4 lattice and the four-regime partition. QE: Quantization Error; TE: Topographic Error; Sil: Silhouette Score; DB: Davies–Bouldin index.

Parameter	Value
SOM lattice size	4×4
Number of neurons	16
Training iterations	2000
Weight initialization	PCA
Training mode	Batch
Random seed	42
Quantization Error (QE)	0.5043
Topographic Error (TE)	0.3533
Number of regimes (k)	4
Silhouette Score (Sil)	0.2153
Davies–Bouldin index (DB)	0.9901

Evaluation of the candidate partitions from $k = 2$ to $k = 7$ showed that the Silhouette Score and Davies–Bouldin index favoured different partition structures, indicating that no single internal validity criterion provided an unambiguous optimum. The $k = 2$ solution produced the highest Silhouette Score (0.2475), but resulted in a highly unbalanced 88.5/11.5% partition and therefore provided limited physical discrimination among cyclone-exposure states. In contrast, the $k = 6$ solution yielded the lowest Davies–Bouldin index (0.8789), but generated a more fragmented partition with weaker physical interpretability. The $k = 4$ solution provided an intermediate configuration, with a Silhouette Score of 0.2153 and a Davies–Bouldin index of 0.9901, while preserving four physically distinguishable exposure patterns associated with differences in cyclone intensity, location, seasonality, and movement characteristics.

A temporal sensitivity analysis was also performed to evaluate whether changes in the observational environment across the 1970–2023 record substantially affected the identified regime structure. The SOM–Ward procedure was independently applied to the 1970–1996 and 1997–2023 subperiods using the same nine feature definitions, normalization procedure, 4×4 SOM configuration, PCA-based initialization, training parameters, and $k = 4$. Compared with the reference 1970–2023 partition, the resulting regime assignments yielded ARI values of 0.356 for 1970–1996 and 0.467 for 1997–2023, indicating moderate temporal agreement. Thus, although individual observation assignments varied between periods, the multivariate regime structure showed a recurrent pattern across the two temporal subsets. These results support the use of the full 1970–2023 record for the primary analysis while acknowledging that residual temporal variability in the underlying best-track observations cannot be completely excluded.

The storm-balanced sensitivity analysis yielded regime frequencies of 15.1%, 53.9%, 26.5%, and 4.5% for R0–R3, respectively, compared with 11.5%, 62.0%, 21.9%, and 4.6% in the primary point-level analysis, indicating that the relative prominence of the regimes changes moderately under storm balancing, while the four-regime structure remains consistent. To further assess the robustness of the candidate partitions, a cyclone-level bootstrap analysis with 500 replicates was performed. The $k = 4$ solution showed moderate stability, with a mean Adjusted Rand Index (ARI) of 0.303 and a median ARI of 0.290. Although the $k = 5$ – 7 configurations exhibited higher bootstrap agreement, they produced increasingly fragmented exposure patterns without a corresponding improvement in physical inter-

pretability. Therefore, $k = 4$ was retained as the most parsimonious compromise between internal clustering quality, stability, and physical interpretability. The moderate Silhouette and DBI values also indicate that the regimes should be interpreted as partially overlapping exposure states rather than as sharply separated cyclone populations.

A sensitivity analysis was also conducted by excluding the 152 observations with imputed minimum central pressure values. The resulting $k = 4$ solution retained a distinct extreme-intensity regime, with 29 of the 31 observations originally assigned to R3 remaining associated with the corresponding extreme cluster. Importantly, none of the 31 observations originally assigned to R3 contained an imputed pressure value in the primary dataset. The persistence of this extreme-intensity pattern after excluding all imputed-pressure observations therefore indicates that the identification of R3 was not driven by pressure imputation.

3.2. Cyclone Exposure Regimes

Regime R0: Late-season slow north-eastward disturbances near San Andrés

Regime R0 groups 77 six-hourly observations (11.5% of the total) belonging to 22 unique storms, with a historical mean recurrence interval of 2.45 years. Among these, 15 storms are predominantly associated with R0 based on their majority regime assignment. These storms have a mean duration of approximately 28.8 h (1.20 days) within the study domain and a mean track length of approximately 260.6 km. The mean maximum sustained wind is 23.2 m/s (approximately 45 kt), corresponding to tropical-storm intensity, while the mean minimum central pressure is 994.0 hPa, with the lowest value in this regime reaching 959 hPa. The translational kinematics are distinctive: the mean translation speed is 10.3 km/h, and the mean bearing is 37° (north-eastward), indicating systems moving away from the Colombian Caribbean towards the central Atlantic. Geographically, R0 observations cluster near the archipelago of San Andrés and Providencia (mean position 14.5° N, 72.8° W). The seasonal distribution peaks in November (33 observations), with secondary occurrences in October (25) and December (8). The combination of weak intensity, slow translation, late-season occurrence, and north-eastward heading is consistent with systems in the late stage of the Atlantic hurricane season that recurve towards the north-east near the south-western Caribbean.

Regime R1: Fast westward tropical-storm-dominated exposure across the central Caribbean

Regime R1 is the dominant partition, grouping 414 observations (62.0%) from 66 unique storms, with the shortest historical mean recurrence interval of the four regimes (0.82 years, i.e., near-annual historical occurrence). Among these, 59 storms are predominantly associated with R1. These storms have a mean duration of approximately 39.6 h (1.65 days) within the study domain and a mean track length of approximately 806.3 km, consistent with their sustained westward propagation across the central Caribbean. The mean maximum sustained wind is 21.1 m/s (approximately 41 kt), corresponding to tropical-storm intensity. However, the peak intensity within R1 reaches 77.2 m/s (approximately 150 kt), indicating that this regime also includes observations from storms that attained Category 5 hurricane intensity. The mean minimum central pressure is 997.6 hPa. The kinematics are the most distinctive feature of R1: the mean translation speed is 22.4 km/h, approximately twice that of R0 and R2, and the mean bearing is 268° (due west), indicating a classical east-to-west transit across the Caribbean Sea. Geographically, R1 observations occupy the central Colombian Caribbean (mean position 13.7° N, 74.5° W). The seasonal distribution peaks in September (108 observations) and August (102), in close agreement with the climatological peak of the Atlantic hurricane season. R1

may be interpreted as the dominant historical cyclone-exposure regime in the Colombian Caribbean offshore.

Regime R2: Slow north-westward disturbances off Central America

Regime R2 groups 146 observations (21.9%) from 36 unique storms, with a historical mean recurrence interval of 1.50 years. Among these, 25 storms are predominantly associated with R2. These storms have the shortest mean duration within the study domain among the four regimes, approximately 19.4 h (0.81 days), and a mean track length of approximately 114.9 km, indicating comparatively short spatial exposure within the study region. It is the weakest regime in terms of mean intensity, with a mean maximum sustained wind of 19.1 m/s (approximately 37 kt), corresponding to low-end tropical-storm intensity, and a mean minimum central pressure of 998.0 hPa; the lowest pressure within R2 is 964 hPa. The kinematics resemble those of R0 in translation speed (10.9 km/h, slow) but differ in heading: the mean bearing is 328° (north-westward), indicating systems moving towards the coast of Central America. Geographically, R2 observations cluster in the western Caribbean (mean position 13.3° N, 79.8° W), off the coasts of Nicaragua and Honduras. The seasonal distribution peaks in October (64 observations) and November (58), late in the Atlantic season. This pattern is consistent with the late-season south-western Caribbean conditions that favor the formation of brief disturbances near the Central American coast.

Regime R3: Major hurricanes crossing the central Caribbean

Regime R3 is the rarest but most intense partition, grouping 31 observations (4.6%) from 7 unique storms, with a historical mean recurrence interval of 7.71 years. Four of these storms are predominantly associated with R3. These storms have a mean duration of approximately 34.5 h (1.44 days) within the study domain, with a mean track length of approximately 463.7 km. The mean maximum sustained wind is 67.8 m/s (approximately 132 kt), corresponding to Category 4 intensity on the Saffir–Simpson Hurricane Wind Scale, while the peak wind within R3 reaches 79.7 m/s (approximately 155 kt), corresponding to Category 5 intensity. The mean minimum central pressure is 938.5 hPa, with the lowest value at 916 hPa. The dominant SSHWS category is Category 5. The kinematics indicate fast translation (18.6 km/h) and a west-north-westward heading (mean bearing 295°), traversing the central Caribbean in a relatively narrow corridor. Geographically, R3 observations are located slightly north of R1 (mean position 14.3° N, 72.3° W), within the 13 – 17° N corridor that has been previously described as the principal pathway of major Caribbean hurricanes. The seasonal distribution is concentrated in October (13 observations) and September (12), consistent with the climatological peak of major Atlantic hurricane activity. Given the small number of storms predominantly assigned to R3, this duration should be interpreted as a descriptive statistic of historical cyclone exposure within the study domain. The sensitivity analysis excluding imputed pressure observations preserved the extreme-intensity R3 pattern, supporting that its identification was not an artifact of pressure imputation.

The separation of R3 from the dominant R1 regime is therefore not attributable to intensity alone. R3 combines an extreme intensity–pressure state with a distinct geographical position, west-north-westward propagation, and late-season concentration, indicating that the SOM–Ward partition captures a multivariate exposure configuration rather than a univariate severity gradient. The physical characteristics of the four cyclone exposure regimes are summarized in Table 3.

Table 3. Physical characteristics of the four cyclone exposure regimes (HURDAT2, 1970–2023). N storms exposed denotes cyclones with at least one observation in the regime, whereas N storms predominant denotes the predominant regime assignment. Frequency (%) represents the fraction of six-hourly observations, while storm-balanced frequency (%) is calculated using inverse storm-length weights so that each cyclone contributes equally. Historical mean recurrence interval is calculated from the number of storms exposed during the 54-year study period; mean duration and track length refer to storms predominantly associated with each regime.

Regime	N Obs.	Freq. (%)	Storm-Balanced Freq. (%)	N Storms	Historical Mean Recurrence Interval (Year)	N Storms Predominant	Mean Duration Within Domain (h)	Mean Track Length Within Domain (km)	Vmax (m/s)	Pmin (hPa)	Translation (km/h)	Bearing (°)	SSHWS Dom.
R0	77	11.5	15.1	22	2.45	15	28.8	260.6	23.2	994.0	10.3	37 (NE)	TD; mod. month Nov
R1	414	62.0	53.9	66	0.82	59	39.6	806.3	21.1	997.6	22.4	268 (W)	TS; mod. month Sep
R2	146	21.9	26.5	36	1.50	25	19.4	114.9	19.1	998.0	10.9	328 (NW)	TD; mod. month Oct
R3	31	4.6	4.5	7	7.71	4	34.5	463.7	67.8	938.5	18.6	295 (WNW)	Cat-5; mod. month Oct

Because the six-hourly observations are temporally dependent and storms contribute different numbers of observations, regime frequencies were additionally evaluated using storm-balanced weights. Under equal storm contribution, the frequencies changed from 11.5%, 62.0%, 21.9%, and 4.6% for R0–R3 to 15.1%, 53.9%, 26.5%, and 4.5%, respectively. Thus, the dominant R1 regime remains the most frequent configuration, although its relative contribution decreases by 8.1 percentage points, mainly reflecting the greater representation of long-track storms in the unweighted six-hourly sample. In contrast, the frequency of the extreme R3 regime remains essentially unchanged (4.6% versus 4.5%). At the storm level, based on the predominant regime assigned to each of the 103 cyclones, R0, R1, R2, and R3 account for 14.6%, 57.3%, 24.3%, and 3.9% of storms, respectively. These results indicate that unequal numbers of observations per cyclone affect the estimated relative frequency of the intermediate regimes, but do not alter the identification of R1 as the dominant exposure configuration or R3 as a rare extreme regime.

3.3. Spatial Distribution of Regimes

The spatial distribution of the four regimes, depicted by the color-coded cyclone tracks in Figure 4 and the smoothed spatial density of observations in Figure 5, shows that the SOM-derived regimes exhibit spatially differentiated and physically interpretable exposure patterns. Regime R1 tracks traverse the central Colombian Caribbean along an east-to-west corridor approximately between 12° N and 15° N, consistent with the dominant westward propagation of tropical waves and cyclones in the basin. Regime R3 tracks follow a similar but slightly northward and narrower corridor, traversing the region between approximately 13° N and 17° N and crossing the longitudinal band of the San Andrés and Providencia archipelago. Regime R0 observations are concentrated in the immediate vicinity of the archipelago, while R2 observations cluster off the Central American coast, predominantly west of 80° W.

The smoothed spatial density of observations (Figure 5) further quantifies the geographical footprint of each regime. R1 presents a broad density maximum over the central Colombian Caribbean, consistent with its dominant frequency (62.0% of observations). R3, although contributing only 4.6% of observations, produces a localized density maximum in the 13–16° N, 70–75° W band, a region that overlaps with the prospective offshore wind development areas off the departments of Bolívar and Atlántico. R0 and R2 produce peripheral density maxima near San Andrés and off Central America, respectively, with limited

direct overlap with the principal continental-shelf offshore wind interest zones, although R0 affects the “Seaflower area” that has been considered for offshore wind development around the archipelago.

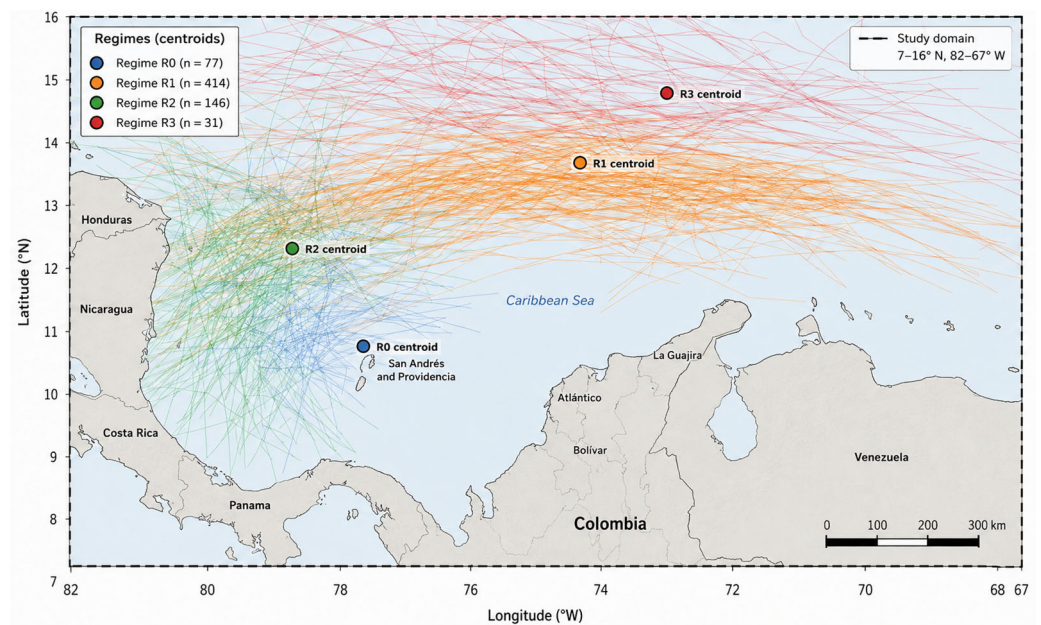


Figure 4. Historical tropical cyclone tracks in the Colombian Caribbean offshore domain from HURDAT2 (1970–2023), colored according to the four SOM-derived cyclone exposure regimes. Large markers indicate the mean spatial centroid of each regime. The dashed bounding box delineates the study domain (7–16° N, 67–82° W); major geographic features are shown for spatial reference.

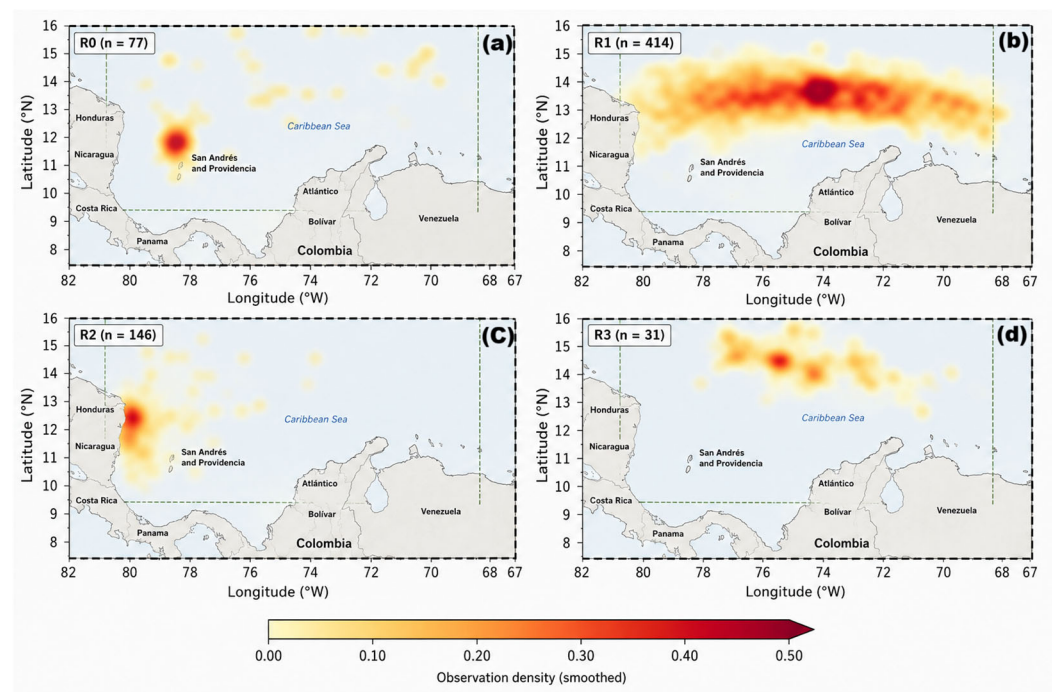


Figure 5. Smoothed Gaussian spatial density of six-hourly cyclone observations for the four SOM-derived exposure regimes: (a) R0, (b) R1, (c) R2, and (d) R3. All panels use the same geographic extent, projection, and density scale to enable direct spatial comparison. The green dashed box delineates the study domain (7–16° N, 67–82° W).

3.4. Physical Differentiation of Regimes

The boxplots of the six main physical descriptors (maximum sustained wind, minimum central pressure, translation speed, latitude, longitude, and movement direction) by regime confirm the differentiation described qualitatively in Section 3.2; see Figure 6. Regime R3 is clearly separated from the other three regimes in maximum sustained wind and minimum central pressure, with wind intensities that are relevant to the extreme-load conditions considered in offshore wind design. Regime R1 is distinguished from R0 and R2 primarily by its translation speed (median approximately 22 km/h versus approximately 10–11 km/h). R0 and R2 share weak intensity but differ in their geographical position (R0 near San Andrés; R2 off Central America) and movement direction (R0 north-eastward; R2 north-westward). The separation between R0 and R2 in movement direction is consistent with their distinct geographical footprints, as north-eastward movement near San Andrés and north-westward movement off Central America correspond to physically different late-season configurations of the basin-scale circulation.

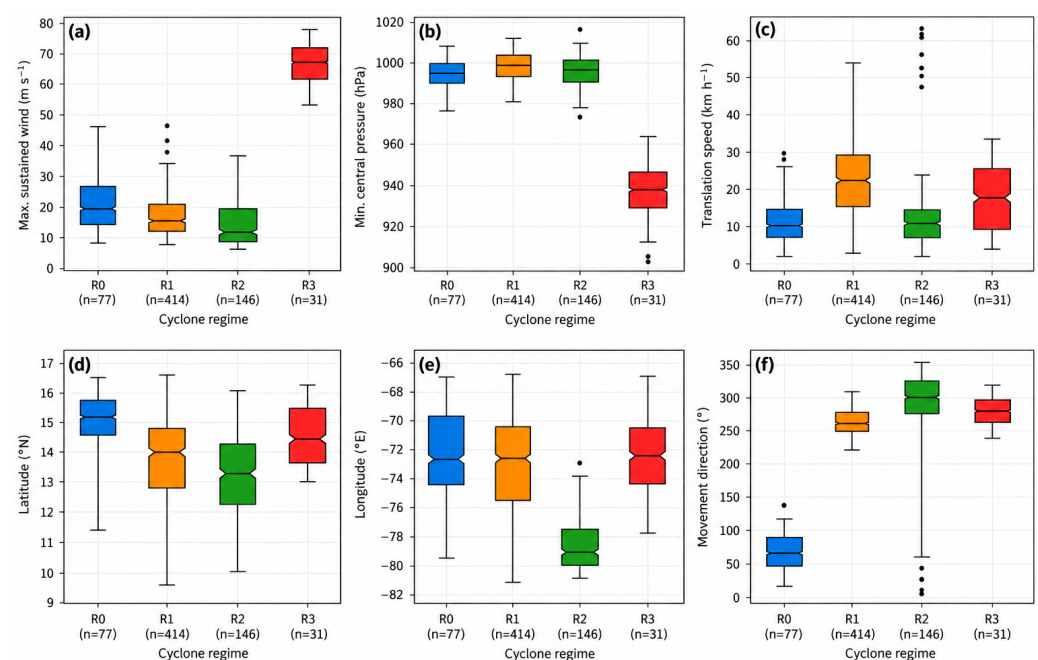


Figure 6. Distribution of physical cyclone descriptors across the four SOM-derived regimes: (a) maximum sustained wind; (b) minimum central pressure; (c) translation speed; (d) latitude; (e) longitude; and (f) movement direction (bearing). Boxplots show the interquartile range, median, and $1.5 \times \text{IQR}$ whiskers; regime sample sizes are indicated below the regime labels.

The Saffir–Simpson composition of the storms assigned to each regime (Figure 7) provides an additional descriptive characterization of the intensity distribution associated with each regime. The concentration of high peak-intensity categories in R3 is consistent with its elevated wind speeds and reduced central pressures identified in Figure 6. R0 and R2 are characterized predominantly by tropical-depression and low-end tropical-storm conditions. This distinction highlights that regime membership reflects a multivariate combination of intensity, location, kinematics, and seasonality rather than peak wind intensity alone.

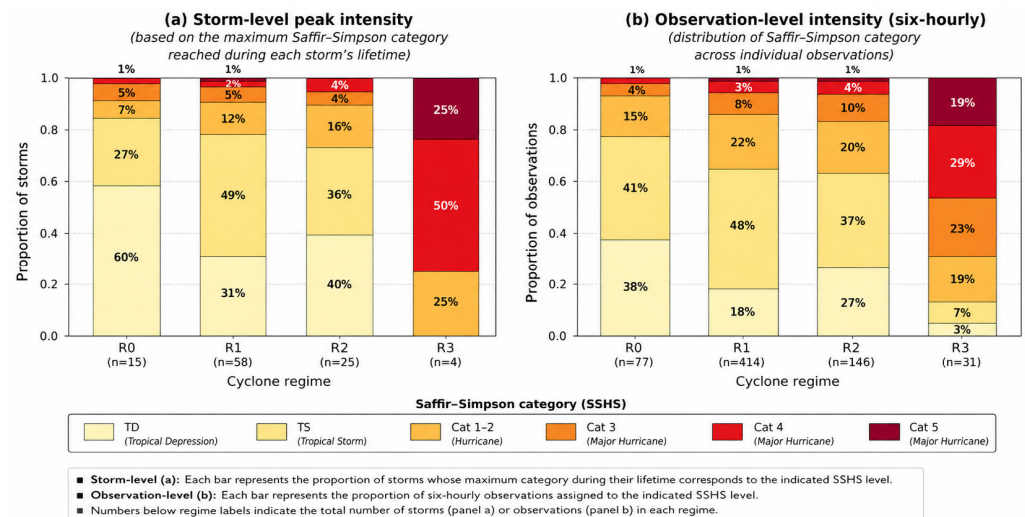


Figure 7. Saffir–Simpson Hurricane Wind Scale composition across the four SOM-derived regimes: (a) storm-level peak intensity, based on the maximum SSHWS category reached during each cyclone's lifetime; and (b) observation-level intensity, based on the SSHWS category assigned to individual six-hourly observations. This distinction highlights the difference between lifetime peak intensity and instantaneous regime-associated intensity.

3.5. Seasonal Distribution

The seasonal distribution of cyclone observations across the four SOM-derived regimes is presented in Figure 8, while the corresponding monthly frequencies and regime-specific seasonal statistics are summarized in Table 4. R1 exhibits the highest concentration of observations during August and September, with 102 and 108 observations, respectively, broadly coinciding with the climatological peak of Atlantic tropical cyclone activity. R3 is comparatively concentrated in September and October, with 12 and 13 observations, respectively, whereas R0 and R2 show later-season maxima, particularly in October and November, with 25 and 33 observations for R0 and 64 and 58 observations for R2, respectively. These contrasting seasonal profiles indicate that regime membership varies not only in intensity, location, and kinematics, but also in the time of year in which specific exposure patterns are most frequently observed.

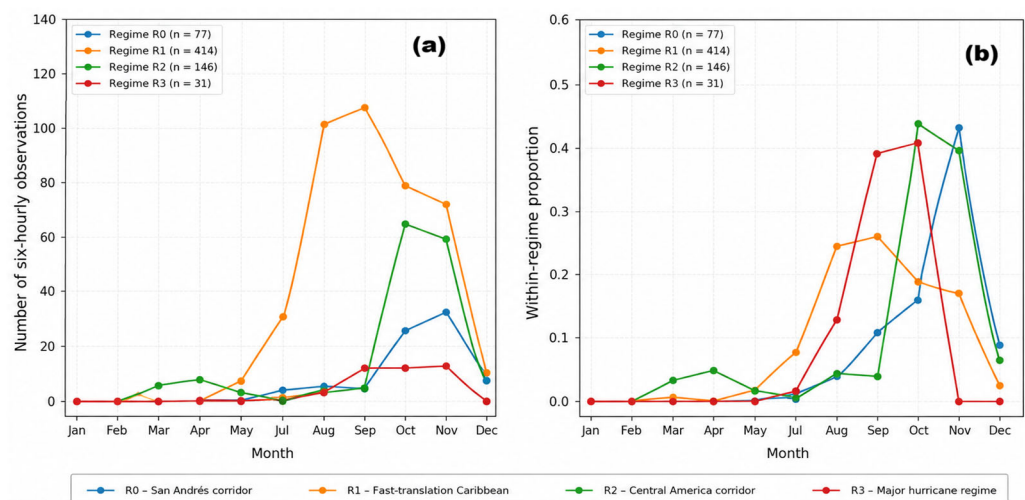


Figure 8. Seasonal distribution of cyclone observations across the four SOM-derived regimes: (a) absolute monthly counts of six-hourly observations; and (b) within-regime normalized monthly proportions, with each regime profile summing to unity.

Table 4. Monthly distribution of six-hourly observations by regime (HURDAT2, 1970–2023). Months with zero observations for all regimes (January–March) are omitted.

Regime	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
R0	0	0	0	2	6	3	25	33	8
R1	4	0	8	33	102	108	78	71	10
R2	6	8	3	0	3	4	64	58	0
R3	0	0	0	2	4	12	13	0	0

This seasonal differentiation provides an additional dimension for interpreting cyclone exposure and may support future assessments of seasonal operational and maintenance constraints for offshore wind infrastructure.

3.6. Interannual Variability

The interannual distribution of cyclone occurrences across the four SOM-derived regimes over 1970–2023 is shown in Figure 9. The figure reveals pronounced year-to-year variability, with periods of relatively high cyclone occurrence, including 1995, 2005, 2017, and 2020, interspersed with years of limited or no occurrence within the study domain. R1 accounts for the largest absolute interannual variability, consistent with its dominant contribution to the overall observation and storm counts, whereas R3 occurs episodically and is concentrated in a comparatively small number of years. The observed temporal variability provides climatic context for the regime distribution and is broadly consistent with the documented modulation of Atlantic tropical cyclone activity by large-scale climate variability, including the Atlantic Multidecadal Oscillation and El Niño–Southern Oscillation [46]. These temporal correspondences, however, should not be interpreted as evidence of a causal relationship between AMO/ENSO variability and the occurrence of the SOM-derived regimes.

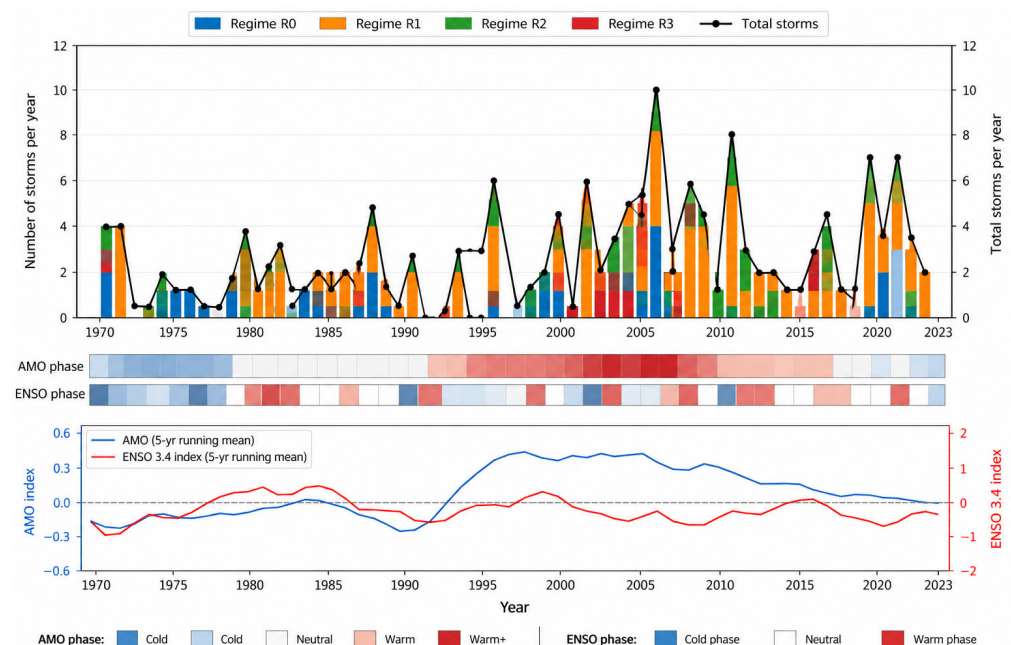


Figure 9. Interannual variability of cyclone occurrence across the four SOM-derived regimes during 1970–2023. Stacked bars show the annual number of unique cyclones with at least one six-hourly observation assigned to each regime; a cyclone may contribute to multiple regimes within the same year. The accompanying climate-index information provides contextual reference for the observed temporal variability and does not imply a causal relationship between AMO/ENSO variability and regime occurrence.

Regime R1 exhibits the largest absolute interannual variability, consistent with its dominant frequency. Regime R3 occurs infrequently and shows temporal clustering in a small number of years, some of which coincide with warm AMO conditions and weak La Niña episodes. Given the small number of R3 storms, these temporal correspondences are suggestive of a possible association rather than evidence of systematic AMO—ENSO control. The episodic occurrence of R3 nevertheless highlights the potential relevance of considering spatial and temporal diversification in future offshore wind risk assessments; however, the quantitative effectiveness of such diversification is beyond the scope of this study. The episodic occurrence of R3 is nevertheless relevant for portfolio-level risk management: spatial diversification of offshore wind projects across zones dominated by different regimes may mitigate the aggregate risk of cyclone-induced interruptions, although a quantitative assessment of this diversification effect is beyond the scope of the present study and is identified as a direction for future work.

3.7. Within-Storm Regime Transitions

The within-storm regime transition matrix (Table 5, Figure 10) characterizes the temporal dynamics of regime membership during individual cyclone events. The diagonal of the matrix is strongly dominant, with R1 → R1 accounting for 340 transitions, R2 → R2 for 109, R0 → R0 for 54, and R3 → R3 for 24. This dominance indicates substantial short-term regime persistence, as consecutive six-hourly observations from the same storm are more frequently assigned to the same regime than to different regimes. The 24 R3 → R3 transitions indicate repeated persistence of the R3 regime across consecutive observations, but they should be interpreted as transition events rather than as independent R3 observations or individual storms. Because transition counts represent links between consecutive observations rather than independent events, these values are interpreted as descriptive indicators of within-storm persistence rather than as independent transition frequencies.

Table 5. Within-storm regime transition matrix (counts). Rows represent source regimes and columns represent destination regimes. Each entry denotes the number of transitions between consecutive six-hourly observations belonging to the same cyclone. Diagonal elements represent consecutive observations remaining within the same regime.

From \ To	R0	R1	R2	R3
R0	54	6	1	0
R1	9	340	11	2
R2	0	6	109	1
R3	0	2	0	24

Off-diagonal transitions provide information on changes in regime membership during cyclone evolution. The most informative transition involving the extreme regime is R1 → R3 (2 cases), indicating two instances in which consecutive six-hourly observations from the same cyclone changed from the R1 to the R3 exposure regime within the study region. These transitions identify a potentially important pathway toward extreme cyclone exposure, but the transition matrix alone does not establish the physical cause of the regime change or demonstrate intensification into a major hurricane. Similarly, the R1 → R2 (11 cases) and R1 → R0 (9 cases) transitions indicate changes in regime membership that require examination of the underlying intensity, kinematic, and spatial variables for physical interpretation. The R1 → R3 transition is rare in absolute terms but relevant from a risk-screening perspective because it identifies instances in which a cyclone moved from the dominant R1 exposure state to the extreme R3 state. This dynamic information

complements static spatial and climatological analyses by explicitly representing changes in regime membership along individual cyclone tracks.

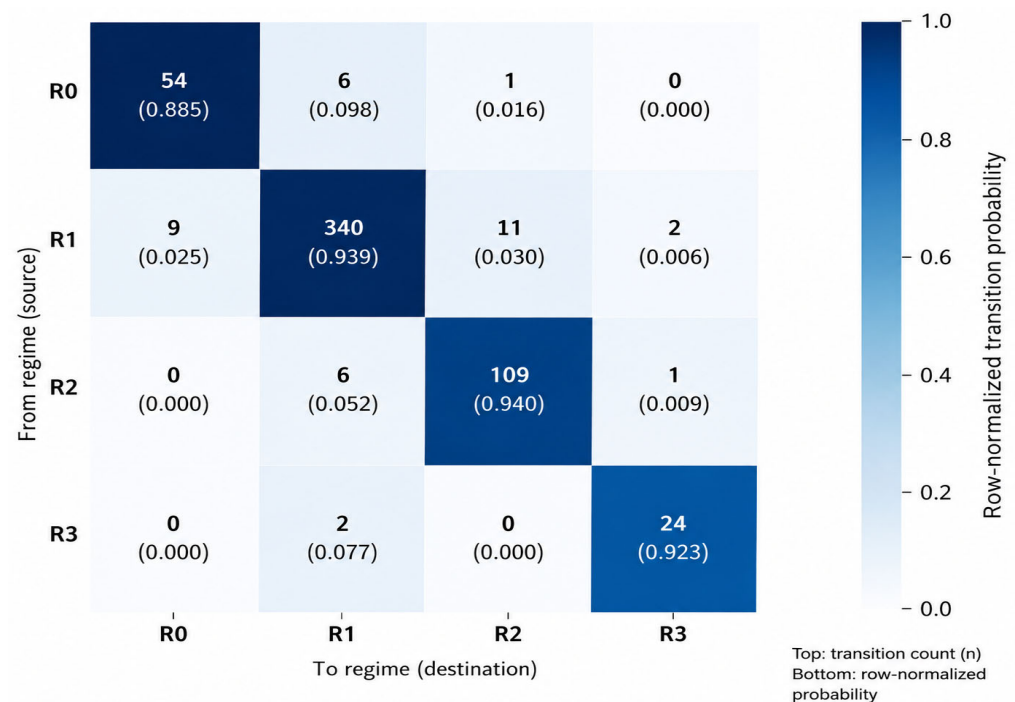


Figure 10. Within-storm regime transition matrix based on consecutive six-hourly observations from the same cyclone. Cell colors represent row-normalized transition probabilities, while cell labels indicate the corresponding number of transitions. Diagonal elements quantify within-regime persistence, whereas off-diagonal elements represent observed changes in regime membership.

4. Discussion

4.1. Physical Interpretation of the Four Regimes

The four cyclone exposure regimes identified by the SOM–Ward workflow are physically interpretable and consistent with the established climatology of the Caribbean basin. Regime R1, the dominant partition, captures the canonical westward propagation of mature tropical storms across the central Caribbean during the peak of the Atlantic hurricane season. Its high translation speed (~22 km/h), westward heading, and broad geographical footprint over the central Colombian Caribbean are compatible with the steering flow associated with the subtropical ridge and the Caribbean Low-Level Jet, both of which favor east-to-west propagation during the peak season.

Regime R3, although rare, corresponds to the major-hurricane pathway of the central Caribbean. Its mean position slightly north of R1, its west-northwestward heading, and its concentration in September–October are all consistent with the well-documented corridor of major Atlantic hurricanes traversing the Greater Caribbean between approximately 13° N and 17° N. The peak wind values within R3 (up to 79.7 m/s) and the minimum pressure of 916 hPa are compatible with intense Category 5 systems such as those historically affecting the Colombian insular territory. The historical mean recurrence interval of approximately 7.7 years estimated for R3 is consistent with the infrequent historical occurrence of major Caribbean hurricanes in the study domain. The temporal distribution of R3 events also coincides in some years with warm AMO conditions, suggesting a possible association that requires further statistical assessment.

Regimes R0 and R2 share weak intensity but differ in their geographical and kinematic signatures. R0, located near San Andrés and moving north-eastward, is consistent with late-season systems that form in the south-western Caribbean and recurve towards the

north-east as the subtropical ridge weakens. R2, located off Central America and moving north-westward, is consistent with the late-season south-western Caribbean disturbances that have been previously described in the regional literature. The November peak of R0 and the October–November peak of R2 are compatible with the southward migration of the Intertropical Convergence Zone during the final weeks of the Atlantic season, which modifies the basin-scale circulation and favors localized cyclogenesis near Central America.

4.2. Comparison with Previous Studies

The present study is methodologically complementary to the predecessor study of the same research group, which applied Fast Fourier Transform (FFT) and Continuous Wavelet Transform (CWT) to sea-surface-temperature and cyclone time series, complemented by two-dimensional kernel density estimation (KDE) of historical track density and an aggregated Wind Risk Index. The spectral approach identified dominant periodicities at annual and ENSO (2–7 year) scales, a post-1995 shift in spectral energy associated with the positive AMO phase, and a thermodynamically consistent energy corridor along 12–16° N. The present study does not repeat the spectral, KDE, or Wind Risk Index analyses; rather, it adds a multivariate, unsupervised classification of cyclone exposure that is not available from those approaches. The two studies are therefore complementary: the spectral analysis characterizes temporal periodicities and trends, while the SOM-based classification characterizes multivariate regimes of cyclone exposure. This complementarity is consistent with the broader literature on SOM applications in synoptic climatology, where SOM-derived regimes are typically used alongside, rather than in replacement of, spectral or principal-component analyses.

Compared with previous SOM applications to tropical cyclone tracks in other basins, the present framework differs in three respects. First, it uses nine physically interpretable HURDAT2-derived features rather than track-shape descriptors alone, allowing regimes to be characterized by intensity and kinematics in addition to trajectory. Second, it evaluates candidate regime partitions using the Silhouette Score and Davies–Bouldin index jointly, complemented by cyclone-level bootstrap stability analysis and physical interpretability. This combined assessment avoids relying on a single internal validity criterion and provides a more robust basis for selecting a parsimonious regime configuration. Third, it includes an explicit within-storm transition matrix, which captures the temporal dynamics of regime membership and is not available in static track-clustering analyses.

These methodological differences shift the interpretation from identifying geometrically similar tracks toward characterizing recurring multivariate exposure states. Consequently, the resulting regimes retain information on intensity, kinematics, spatial position, and seasonality that would be less explicit in classifications based primarily on track geometry, while the transition analysis provides an additional event-level perspective on changes in regime membership.

4.3. Implications for Climate-Resilient Offshore Wind Planning

The cyclone exposure regimes identified in this study provide a complementary hazard characterization layer for offshore wind infrastructure planning in the Colombian Caribbean. Three main implications can be drawn from the results, while explicitly acknowledging that the present study does not estimate wind resource, annual energy production, or levelized cost of electricity.

First, the spatial differentiation of regimes allows the basin to be zoned according to its dominant cyclone exposure profile. Areas dominated by R1 are characterized by more frequent but generally lower-intensity cyclone exposure, whereas areas with historical R3 exposure warrant particular attention in subsequent assessments of extreme wind, wave,

structural, foundation, and mooring loads associated with major hurricanes. The overlap between the R3 density maximum (13–16° N, 70–75° W) and the prospective offshore wind development areas off Bolívar and Atlántico is particularly relevant for preliminary siting decisions.

Second, the historical mean recurrence intervals estimated for each regime (0.82 years for R1, 1.50 years for R2, 2.45 years for R0, and 7.71 years for R3) provide descriptive measures of historical regime occurrence derived from the 54-year observational record. These values are not probabilistic return periods, exceedance probabilities, or estimates of design-level recurrence. They may support preliminary historical hazard characterization and risk screening, but they should not replace site-specific extreme-value analyses required for offshore wind design. In particular, the R3 regime, associated with seven distinct storms during the 1970–2023 period and peak winds of up to 79.7 m/s, indicates that extreme cyclone exposure should be considered where this regime is historically relevant. However, the present analysis does not provide sufficient information to determine turbine design classes or establish site-specific design loads. Instead, R3 should be regarded as a screening-level indicator that warrants subsequent assessment of extreme wind, wave, storm-surge, structural, foundation, and mooring loads under the applicable offshore wind design standards.

A practical pathway for integrating the identified regimes into future engineering assessments is to use the regime assignments as conditioning classes for site-specific environmental characterization. Cyclone-track observations can be combined with high-resolution atmospheric and oceanographic datasets to estimate regime-conditioned distributions of extreme wind speed, wind direction, significant wave height, peak wave period, and associated directional characteristics. These environmental states can subsequently be incorporated into coupled wind–wave–structure simulations and evaluated within the applicable site-specific IEC design load cases. Because HURDAT2 does not provide wave parameters or turbine-level atmospheric measurements, these design-oriented quantities cannot be derived directly within the present framework without introducing additional datasets and modelling assumptions. This integration would therefore extend the present basin-scale exposure characterization toward quantitative engineering assessment while preserving the role of the SOM framework as a preliminary hazard-screening tool.

Third, the within-storm transition matrix (Table 5) provides information that may be relevant to operational risk management. The R1 → R3 transition, although rare (2 cases), identifies instances in which consecutive six-hourly observations from the same cyclone changed from the dominant R1 exposure regime to the extreme R3 regime. Although the transition matrix does not by itself establish the physical mechanism underlying these changes, such transitions may provide useful event-based information for subsequent analyses of cyclone evolution and operational risk management. Their direct implications for preventive shutdown, yaw alignment, pitch-feathering, or other turbine-control actions require dedicated dynamic and aeroelastic analyses. The seasonal concentration of each regime may also support preliminary planning of installation and major maintenance windows, subject to site-specific meteorological and oceanographic constraints.

The four regimes should not be interpreted as a substitute for site-specific hazard assessments based on synthetic tropical cyclone models, mesoscale numerical weather prediction, or coupled wave–wind–structure simulations. They constitute a preliminary, basin-scale characterization that can guide the scope and focus of subsequent, higher-resolution studies.

4.4. Limitations and Implications

Several limitations should be acknowledged. First, HURDAT2 is a best-track product and is subject to known uncertainties, particularly for intensity estimates of historical storms before the modern reconnaissance era. Although the 1970–2023 period was selected to mitigate pre-satellite-era biases, changes in observational practices, intensity-estimation algorithms, and re-analysis procedures over the 54-year record may introduce residual inhomogeneities. Second, the SOM is sensitive to initialization and parametrization; while PCA initialization, batch training, and a fixed random seed ensure reproducibility, alternative configurations may yield slightly different partitions, and the sensitivity of the four-regime solution to these choices has not been exhaustively explored. The feature space also implicitly assigns relative weight to the four physical dimensions according to the number of component variables included in each dimension. Although min–max normalization places all individual variables on a common scale, it does not enforce equal weighting among conceptual feature groups. Consequently, alternative group-level weighting schemes may produce somewhat different regime structures and should be evaluated in future sensitivity analyses. The primary SOM analysis is conducted at the six-hourly track-point level, so the 668 observations represent repeated measurements from 103 individual cyclones rather than independent events. Although the storm-balanced sensitivity analysis reduces the influence of unequal storm representation, temporal dependence among consecutive observations remains inherent to the trajectory-based formulation. Consequently, the identified regimes should be interpreted primarily as recurrent cyclone-exposure states at the track-point level rather than as statistically independent storm populations.

The storm-balanced sensitivity analysis further shows that the relative frequency of the regimes is partly influenced by unequal storm-track lengths. The reduction in R1 from 62.0% to 53.9% indicates that its high point-level frequency is partly associated with the greater representation of longer trajectories, whereas the increase in R0 and R2 reflects the greater contribution of shorter-duration systems when each cyclone is weighted equally. Importantly, R3 remains virtually unchanged (4.6% versus 4.5%), supporting the interpretation that the extreme regime is not an artifact of over-representation by a small number of long-track storms. Accordingly, the results are interpreted primarily in terms of instantaneous six-hourly exposure, while storm-level frequencies provide an event-based complementary perspective.

The temporal evolution of tropical cyclone observing systems also represents a limitation of the historical record, particularly for position and intensity estimates in earlier decades. Although the temporal sensitivity analysis yielded moderate agreement with the reference regime structure (ARI = 0.356 for 1970–1996 and 0.467 for 1997–2023), residual observational inhomogeneity and temporal variability cannot be completely excluded.

The 4×4 lattice was selected on the basis of a composite score balancing QE and TE; alternative selection criteria (e.g., focusing primarily on TE, or considering the Bayesian Information Criterion) might yield different lattice sizes and, consequently, different regime structures. The sensitivity analysis across SOM lattice sizes nevertheless showed that the principal four-regime structure was reasonably robust, with ARI values of 0.661, 0.334, 0.554, and 0.615 for the 5×5 , 6×6 , 7×7 , and 8×8 configurations, respectively, relative to the 4×4 reference solution. The Silhouette Score (0.2153) indicates moderate internal separation, reflecting substantial overlap among the identified cyclone regimes. This overlap is consistent with the continuous nature of cyclone exposure states and implies that regime boundaries should be interpreted as soft rather than strictly discrete. The historical occurrence metrics reported in Table 3 are derived from a 54-year record and are subject to sampling uncertainty, particularly for the rare R3 regime ($n = 7$ storms); they should therefore be interpreted as historical mean recurrence intervals rather than

probabilistic return-period estimates. Finally, the framework does not account for non-stationarity associated with climate change, which has been projected to increase tropical cyclone intensity and to alter the frequency of intense storms in the Atlantic basin; the regimes identified here characterize historical exposure and may evolve under future climate scenarios.

4.5. Future Research Directions

Four main directions for future research emerge from the present study. First, the framework can be extended to other tropical basins exposed to offshore wind development, including the South China Sea, the Bay of Bengal, and the U.S. Gulf of Mexico, using IBTrACS or regional best-track archives. Comparative analyses across basins would test the robustness of the regime-based approach and may reveal basin-specific regime structures. Second, the cyclone exposure regimes identified here can be integrated with continuous reanalysis-based wind-resource assessments (ERA5, MERRA-2) and with the Global Wind Atlas, within a multi-criteria siting framework that combines hazard, resource, bathymetry, grid proximity, and competing-use constraints. Third, the within-storm transition matrix provides a basis for future development of stochastic state-transition models, including Markov-chain formulations, to simulate cyclone exposure over the lifetime of offshore wind projects and subsequently estimate potential downtime and economic losses. Fourth, the framework can be extended to incorporate climate-change projections from synthetic tropical cyclone models downscaled from CMIP6 scenarios, allowing the four-regime classification to be re-evaluated under non-stationary conditions and supporting climate-resilient design specifications aligned with the IEC 61400-1 standard.

5. Conclusions

This study developed and evaluated an unsupervised learning framework based on Self-Organizing Maps to identify and characterize historical cyclone exposure regimes in the Colombian Caribbean offshore, using the NOAA HURDAT2 best-track database for the 1970–2023 period. The framework combines a 4×4 SOM trained on nine physically interpretable HURDAT2-derived features (latitude, longitude, maximum sustained wind, minimum central pressure, translation speed, circularly encoded movement direction, and circular seasonality) with hierarchical Ward clustering of the 16 SOM neurons. Candidate regime partitions were evaluated using the Silhouette Score and Davies–Bouldin index jointly, complemented by cyclone-level bootstrap stability analysis and physical interpretability. The configuration yielded a Quantization Error of 0.5043, a Topographic Error of 0.3533, a Silhouette Score of 0.2153, and a Davies–Bouldin index of 0.9901. These metrics characterize the representation accuracy, topological preservation, and internal separation of the selected four-regime partition obtained from 668 six-hourly observations corresponding to 103 unique storms.

Four physically interpretable cyclone exposure regimes were identified. R0 (11.5% of observations, 22 storms, historical mean recurrence interval 2.45 years) corresponds to late-season slow north-eastward disturbances near San Andrés and Providencia. R1 (62.0%, 66 storms, historical mean recurrence interval 0.82 years) is the dominant regime and corresponds to fast westward mature tropical storms traversing the central Colombian Caribbean during the peak of the Atlantic season. R2 (21.9%, 36 storms, historical mean recurrence interval 1.50 years) corresponds to slow north-westward disturbances off Central America in the late season. R3 (4.6%, 7 storms, historical mean recurrence interval 7.71 years) corresponds to major hurricanes crossing the central Caribbean with peak winds up to 79.7 m/s and minimum central pressure of 916 hPa. The four regimes differ in intensity, kinematics, geographical position, seasonality, and historical occurrence frequency, and

their spatial footprints are physically coherent with the established climatology of the basin. The storm-balanced sensitivity analysis produced moderate changes in regime frequencies but preserved the four-regime structure, supporting the robustness of the identified exposure patterns to unequal cyclone representation.

The within-storm regime transition matrix further extends the static regime classification by characterizing temporal changes in regime membership along individual cyclone tracks, including two R1 $\rightarrow$ R3 transitions. These transitions identify instances in which consecutive six-hourly observations from the same cyclone changed from the dominant R1 regime to the extreme R3 regime. Although the transition matrix does not by itself demonstrate physical intensification into a major hurricane, it provides a dynamic characterization that complements static spatial and climatological analyses by identifying changes in cyclone exposure state along individual tracks.

The cyclone exposure regimes identified in this study constitute a complementary hazard characterization layer for climate-resilient offshore wind infrastructure planning in the Colombian Caribbean. They may inform preliminary hazard screening and spatial prioritization for subsequent offshore wind assessments, including site-specific evaluation of extreme wind, wave, structural, foundation, and mooring loads under the applicable design standards. The results do not estimate wind resource, annual energy production, or levelized cost of electricity, and should be integrated with reanalysis-based wind-resource assessments (ERA5, MERRA-2, Global Wind Atlas) within a multi-criteria planning framework. The methodology is transferable to other tropical basins exposed to offshore wind development and can be extended to incorporate climate-change projections from synthetic tropical cyclone models.

Author Contributions: Conceptualization, M.E.C.d.R. and A.O.-C.; Methodology, M.E.C.d.R., V.O.-O. and C.R.-A.; Software, M.E.C.d.R. and C.R.-A.; Validation, M.E.C.d.R., A.O.-C., C.R.-A. and D.R.-L.; Formal Analysis, M.E.C.d.R., C.R.-A. and D.R.-L.; Investigation, M.E.C.d.R., A.O.-C., C.R.-A. and D.R.-L.; Resources, A.O.-C. and C.R.-A.; Data Curation, M.E.C.d.R. and V.O.-O.; Writing—original draft preparation, M.E.C.d.R.; Writing—review and editing, M.E.C.d.R., A.O.-C., V.O.-O., C.R.-A. and D.R.-L.; Visualization, M.E.C.d.R., V.O.-O. and C.R.-A.; Supervision, A.O.-C. and C.R.-A.; Project Administration, M.E.C.d.R. and A.O.-C.; Funding Acquisition, A.O.-C. and C.R.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the “Ministerio de Ciencia, Tecnología e Innovación Fondo Francisco José de Caldas” and the Agencia Nacional de Hidrocarburos (ANH), through the Contrato de Financiamiento de Recuperación Contingente No. 112721-053-2025, within the project “Herramientas tecnológicas y prototipos para caracterizar, identificar y optimizar el potencial eólico en zonas costeras de Colombia” (Project Code 111983).

Data Availability Statement: The HURDAT2 database used in this study is publicly available from the U.S. National Hurricane Center at <https://www.nhc.noaa.gov/data/> (accessed on 1 October 2025) [23]. The Python (version 3.11.9) code (som_cyclone_regime_analysis.py) and the intermediate CSV outputs supporting the reported results are available from the corresponding author upon reasonable request and will be deposited in a public repository upon acceptance of the manuscript.

Acknowledgments: The authors would like to acknowledge the support provided by Universidad de la Costa, Universidad del Magdalena, and Universidad Nacional Abierta y a Distancia de Colombia in the development of this research. The authors also gratefully acknowledge the data and resources provided by the National Oceanic and Atmospheric Administration (NOAA) and the National Hurricane Center (NHC). This work was supported by the Ministerio de Ciencia, Tecnología e Innovación (Minciencias) through Call 951. During the preparation of this manuscript, the authors used ChatGPT (OpenAI, GPT-5.5 Instant; August 2026) solely to assist in the generation and visual design of Figures 1 and 2. The authors reviewed and edited the generated figures and take full responsibility for the final content of the manuscript.

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

Abbreviations

The following abbreviations are used in this manuscript:

AMO	Atlantic Multidecadal Oscillation
BMU	Best Matching Unit
CWT	Continuous Wavelet Transform
DB	Davies–Bouldin index
EEZ	Exclusive Economic Zone
ENSO	El Niño–Southern Oscillation
ERA5	Fifth-generation ECMWF Atmospheric Reanalysis
FFT	Fast Fourier Transform
HURDAT2	Atlantic Hurricane Database 2nd generation
IEC	International Electrotechnical Commission
KDE	Kernel Density Estimation
MERRA-2	Modern-Era Retrospective analysis for Research and Applications, Version 2
PCA	Principal Component Analysis
QE	Quantization Error
SOM	Self-Organizing Map

References

1. von Jouanne, A.; Agamloh, E.; Yokochi, A. A Review of Offshore Renewable Energy for Advancing the Clean Energy Transition. *Energies* **2025**, *18*, 4798. [\[CrossRef\]](#)
2. Ospino-Castro, A.; Robles-Algarín, C.; Vásquez Capacho, J.W. Balancing Resource Potential and Investment Costs in Offshore Wind Projects: Evidence from Northern Colombia. *Energies* **2025**, *18*, 6003. [\[CrossRef\]](#)
3. Castro, A.O.; Robles-Algarín, C.; Hernández-Callejo, L.; Muñoz Maldonado, Y.; Cordero, A.M. Feasibility Analysis of Offshore Wind Power Projects in the Caribbean Region of Colombia: A Case Study Using FAHP–GIS. *Sustainability* **2023**, *15*, 16620. [\[CrossRef\]](#)
4. Gkeka-Serpetsidaki, P.; Skinitis, G.; Tournaki, S.; Tsoutsos, T. A Review of the Sustainable Siting of Offshore Wind Farms. *Sustainability* **2024**, *16*, 6036. [\[CrossRef\]](#)
5. Valencia, J.; Yepes, J.; Mejía, J.F.; Builes-Jaramillo, A.; Salas, H.D. Influence of Tropical Easterly Waves on the ChocoJet during the 2019 OTREC Campaign. *J. Hydrometeorol.* **2024**, *25*, 325–337. [\[CrossRef\]](#)
6. Cerón, W.L.; Kayano, M.T.; Andreoli, R.V.; Avila-Diaz, A.; de Souza, I.P.; Souza, R.A.F. Pacific and atlantic multidecadal variability relations with the choco and caribbean low-level jets during the 1900–2015 period. *Atmosphere* **2021**, *12*, 1120. [\[CrossRef\]](#)
7. Bhaskaran, S.; Verma, A.S.; Goupee, A.J.; Bhattacharya, S.; Nejad, A.R.; Shi, W. Comparison of Extreme Wind and Waves Using Different Statistical Methods in 40 Offshore Wind Energy Lease Areas Worldwide. *Energies* **2023**, *16*, 6935. [\[CrossRef\]](#)
8. Wang, X.H.; Khor, C.S.; Ng, J.H.; Ung, S.K.; Fazlizan, A.; Wong, K.H. A Review of Natural Hazards’ Impacts on Wind Turbine Performance, Part 2: Earthquakes, Waves, Tropical Cyclones, and Thunderstorm Downbursts. *Energies* **2026**, *19*, 385. [\[CrossRef\]](#)
9. Edwards, E.C.; Holcombe, A.; Brown, S.; Ransley, E.; Hann, M.; Greaves, D. Evolution of floating offshore wind platforms: A review of at-sea devices. *Renew. Sustain. Energy Rev.* **2023**, *183*, 113416. [\[CrossRef\]](#)
10. Xu, N.; Yang, B.; Ren, J. Classification of Tropical Cyclone Tracks in the Northwest Pacific Based on the SD-K-Means Model. *Appl. Sci.* **2025**, *15*, 6160. [\[CrossRef\]](#)
11. IEC 61400-1:2019; Wind Energy Generation Systems Part 1: Design Requirements. IEC: Geneva, Switzerland, 2019.
12. Lipari, S.; Balaguru, K.; Rice, J.; Feng, S.; Xu, W.; Berg, L.K.; Judi, D. Amplified threat of tropical cyclones to US offshore wind energy in a changing climate. *Commun. Earth Environ.* **2024**, *5*, 755. [\[CrossRef\]](#)
13. Barthelmie, R.J.; Dantuono, K.E.; Renner, E.J.; Letson, F.L.; Pryor, S.C. Extreme wind and waves in U.S. east coast offshore wind energy lease areas. *Energies* **2021**, *14*, 1053. [\[CrossRef\]](#)
14. Ma, K.T.; Huang, W.Y.; Wu, K.Y.; Ivanov, G. Wind Farm Design with 15 MW Floating Offshore Wind Turbines in Typhoon Regions. *J. Mar. Sci. Eng.* **2025**, *13*, 687. [\[CrossRef\]](#)
15. Song, T.; Li, Y.; Meng, F.; Xie, P.; Xu, D. A Novel Deep Learning Model by BiGRU with Attention Mechanism for Tropical Cyclone Track Prediction in the Northwest Pacific. *J. Appl. Meteorol. Climatol.* **2022**, *61*, 3–12. [\[CrossRef\]](#)
16. Yan, Y.; Nanding, N.; Li, X.; Shi, Y.; Chen, B. Assessment and application of tropical cyclone clustering in the South China Sea. *Sci. Rep.* **2025**, *15*, 1762. [\[CrossRef\]](#) [\[PubMed\]](#)

17. Li, T.; Yang, B.; Yang, Q.; Zhang, M.; Hu, W.; Guo, K. Wind farm layout optimization with multi-type wind turbines for minimizing levelized cost of energy. *Renew. Energy* **2026**, *270*, 124386. [CrossRef]
18. Fan, X.; Zhang, S.; Xue, X.; Jiang, R.; Fan, S.; Kou, H. An Improved Self-Organizing Map (SOM) Based on Virtual Winning Neurons. *Symmetry* **2025**, *17*, 449. [CrossRef]
19. Zhang, Z.; Gu, Y.; Wang, Z.; Luo, S.; Sun, S.; Wang, S.; Feng, G. Application of the Self-Organizing Map Method in February Temperature and Precipitation Pattern over China: Comparison between 2021 and 2022. *Atmosphere* **2023**, *14*, 1182. [CrossRef]
20. Kempitiya, T.; Alahakoon, D.; Osipov, E.; Kahawala, S.; De Silva, D. A Two-Layer Self-Organizing Map with Vector Symbolic Architecture for Spatiotemporal Sequence Learning and Prediction. *Biomimetics* **2024**, *9*, 175. [CrossRef] [PubMed]
21. Zaman, T.; Juliano, T.W.; Hawbecker, P.; Astitha, M. On Predicting Offshore Hub Height Wind Speed and Wind Power Density in the Northeast US Coast Using High-Resolution WRF Model Configurations during Anticyclones Coinciding with Wind Drought. *Energies* **2024**, *17*, 2618. [CrossRef]
22. Yoon, J.S.; Park, S.; Chung, I.U. Interpretational Pitfalls in SOM-Based Clustering: A Case Study of Extreme Cold Events in South Korea. *Atmosphere* **2026**, *17*, 44. [CrossRef]
23. Wang, L.; Wan, B.; Zhou, S.; Sun, H.; Gao, Z. Forecasting tropical cyclone tracks in the northwestern Pacific based on a deep-learning model. *Geosci. Model Dev.* **2023**, *16*, 2167–2179. [CrossRef]
24. Boussiou, L.; Zeng, C.; Guénais, T.; Bertsimas, D. Hurricane Forecasting: A Novel Multimodal Machine Learning Framework. *Weather Forecast* **2022**, *37*, 817–831. [CrossRef]
25. Zekeik, Y.; OrtizBevia, M.J.; Alvarez-Garcia, F.J.; Haddi, A.; El Mourabit, Y.; RuizdeElvira, A. Long-Term Assessment of Morocco's Offshore Wind Energy Potential Using ERA5 and IFREMER Wind Data. *J. Mar. Sci. Eng.* **2024**, *12*, 460. [CrossRef]
26. Corkran, R.; Trepanier, J.; Brown, V. Spatiotemporal Climatology of Georgia Tropical Cyclones and Associated Rainfall. *J. Mar. Sci. Eng.* **2024**, *12*, 1693. [CrossRef]
27. Rueda-Bayona, J.G.; Guzmán, A.; Eras, J.J.C.; Silva-Casarín, R.; Bastidas-Arteaga, E.; Horrillo-Caraballo, J. Renewables energies in Colombia and the opportunity for the offshore wind technology. *J. Clean. Prod.* **2019**, *220*, 529–543. [CrossRef]
28. Asthana, T.; Krim, H.; Sun, X.; Roheda, S.; Xie, L. Atlantic hurricane activity prediction: A machine learning approach. *Atmosphere* **2021**, *12*, 455. [CrossRef]
29. Hernández-Delgado, E.A.; Alejandro-Camis, P.; Cabrera-Beauchamp, G.; Fonseca-Miranda, J.S.; Gómez-Andújar, N.X.; Gómez, P.; Guzmán-Rodríguez, R.; Olivo-Maldonado, I.; Suleimán-Ramos, S.E. Stronger Hurricanes and Climate Change in the Caribbean Sea: Threats to the Sustainability of Endangered Coral Species. *Sustainability* **2024**, *16*, 1506. [CrossRef]
30. Landsea, C. *The Revised Atlantic Hurricane Database (HURDAT2)*; The National Hurricane Center: Miami, FL, USA, 2022.
31. Landsea, C.W.; Franklin, J.L. Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Weather Rev.* **2013**, *141*, 3576–3592. [CrossRef]
32. Tian, J.; Ooka, R.; Lee, D. Multi-scale solar radiation and photovoltaic power forecasting with machine learning algorithms in urban environment: A state-of-the-art review. *J. Clean. Prod.* **2023**, *426*, 139040. [CrossRef]
33. Rashid, L.B.; Shuja, S.Z.; Rehman, S. Machine Learning Forecasting of Direct Solar Radiation: A Multi-Model Evaluation with Trigonometric Cyclical Encoding. *Forecasting* **2025**, *7*, 58. [CrossRef]
34. Kuttippurath, J.; Akhila, R.S.; Martin, M.V.; Girishkumar, M.S.; Mohapatra, M.; Balan Sarojini, B.; Mogensen, K.; Sunanda, N.; Chakraborty, A. Tropical cyclone-induced cold wakes in the northeast Indian Ocean. *Environ. Sci. Atmos.* **2022**, *2*, 404–415. [CrossRef]
35. Törönen, P.; Kolehmainen, M.; Wong, G.; Castrén, E. Analysis of gene expression data using self-organizing maps. *FEBS Lett.* **1999**, *451*, 142–146. [CrossRef] [PubMed]
36. Alshehri, H.S.; Bajaber, F. Improving Data Aggregation in IoT Sensor Networks Using Self-Organizing Maps and Firefly Optimization Algorithm. *Sensors* **2025**, *25*, 7107. [CrossRef] [PubMed]
37. Guan, S.; Jin, F.F.; Tian, J.; Lin, I.I.; Pun, I.F.; Zhao, W.; Huthnance, J.; Xu, Z.; Cai, W.; Jing, Z.; et al. Ocean internal tides suppress tropical cyclones in the South China Sea. *Nat. Commun.* **2024**, *15*, 3903. [CrossRef] [PubMed]
38. Huang, H.; Yang, J.; Zeng, H.; Wang, Y.; Xiao, L. Self-Organizing Maps-Assisted Variational Autoencoder for Unsupervised Network Anomaly Detection. *Symmetry* **2025**, *17*, 520. [CrossRef]
39. Vettigli, G. MiniSom, a Minimalistic and Numpy Based Implementation of the Self Organizing Maps Giuseppe. 2021. Available online: <https://github.com/JustGlowing/minisom/> (accessed on 9 June 2026).
40. Chaudhary, V.; Bhatia, R.S.; Ahlawat, A.K. A novel Self-Organizing Map (SOM) learning algorithm with nearest and farthest neurons. *Alex. Eng. J.* **2014**, *53*, 827–831. [CrossRef]
41. Gelao, V.; Fornasaro, S.; Briguglio, S.C.; Mattiussi, M.; De Martin, S.; Astel, A.M.; Barbieri, P.; Lichen, S. Self-Organizing Maps: An AI Tool for Identifying Unexpected Source Signatures in Non-Target Screening Analysis of Urban Wastewater by HPLC-HRMS. *Toxics* **2024**, *12*, 113. [CrossRef] [PubMed]
42. Matic, F.; Kalinić, H.; Vilibić, I. Interpreting Self-Organizing Map errors in the classification of ocean patterns. *Comput. Geosci.* **2018**, *119*, 9–17. [CrossRef]

43. Zhang, T.; Lan, X.; Feng, J. A Progressive Clustering Approach for Buildings Using MST and SOM with Feature Factors. *ISPRS Int. J. Geoinf.* **2025**, *14*, 103. [[CrossRef](#)]
44. Wang, J.; Zhang, Z.; Yue, S. A Validity Index for Clustering Evaluation by Grid Structures. *Mathematics* **2025**, *13*, 1017. [[CrossRef](#)]
45. dela Rosa, M.E.; Ospino-Castro, A.; Robles-Algarín, C.; Restrepo-Leal, D.; Olivero-Ortiz, V. Integrated Multi-Scale Spectral Framework for Tropical Cyclone Dynamics: Implications for Offshore Wind Energy Resilience in the Atlantic Caribbean Basin. *Energies* **2026**, *19*, 2473. [[CrossRef](#)]
46. Szczapiński, A.; Bednorz, E.; Czernecki, B. Variability in the Occurrence of Tropical and Extratropical Cyclones in the Atlantic Ocean and Its Climatic and Hydrological Determinants. *Atmosphere* **2023**, *14*, 312. [[CrossRef](#)]

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.