JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2025JH000865

XAI-GPI: An Interpretable and Adaptive Machine Learning Genesis Index for Tropical Cyclones

Filippo Dainelli[1], Jorge Pérez-Aracil[2], Guido Ascenso[1,3], Enrico Scoccimarro[3], Matteo Giuliani[1,3], Sancho Salcedo Sanz[2], and Andrea Castelletti[1,3]

[1]Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, Italy, [2]Department of Signal Processing and Communications, Universidad de Alcalá, Madrid, Spain, [3]CMCC Foundation—Euro-Mediterranean Center on Climate Change, Italy

Key Points:

- A machine learning framework improves tropical cyclogenesis detection employing feature selection to identify the most predictive features
- The framework outperforms traditional Genesis Potential Indices in capturing the interannual variability of tropical cyclone activity
- The approach advances understanding of cyclogenesis by pinpointing key drivers in the North Atlantic and Northeastern Pacific

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Castelletti,
andrea.castelletti@polimi.it

Citation:

Dainelli, F., Pérez-Aracil, J., Ascenso, G., Scoccimarro, E., Giuliani, M., Sanz, S. S., & Castelletti, A. (2025). XAI-GPI: An interpretable and adaptive machine learning genesis index for tropical cyclones. *Journal of Geophysical Research: Machine Learning and Computation*, *2*, e2025JH000865. https://doi.org/10.1029/2025JH000865

Received 4 JUL 2025
Accepted 10 DEC 2025

Author Contributions:

Conceptualization: Filippo Dainelli, Jorge Pérez-Aracil, Guido Ascenso, Enrico Scoccimarro, Matteo Giuliani, Sancho Salcedo Sanz, Andrea Castelletti
Data curation: Filippo Dainelli
Formal analysis: Filippo Dainelli
Funding acquisition: Matteo Giuliani, Andrea Castelletti
Investigation: Filippo Dainelli
Methodology: Filippo Dainelli, Jorge Pérez-Aracil, Sancho Salcedo Sanz

Abstract Tropical cyclones are extreme weather events that threaten the lives of millions of people worldwide and cause widespread damage. Accurately detecting the genesis of tropical cyclones (TCG), particularly its interannual variability across ocean basins, remains a major challenge. This work aims to improve annual TCG estimation while simultaneously identifying key formation drivers and providing explanatory insights into their influence on cyclone occurrences and interannual variability. The approach combines spatial clustering for dimensionality reduction, ensemble-based feature selection, and interpretability via SHapley Additive exPlanations (SHAP). Based on the results of the framework, we propose a Machine Learning-based alternative to conventional Genesis Potential Indices (GPIs) for the estimation of basin-wide aggregated numbers of TCG: the XAI-GPI. Unlike fixed-form GPIs such as the Emanuel and Nolan GPI, our index identifies and explains key predictors of TCG variability directly from the data. Applied across six tropical basins, the XAI-GPI outperforms conventional GPIs in capturing observed year-to-year variability while offering physical interpretability. SHAP analysis reveals how drivers such as vertical shear, relative humidity, and ENSO-related indices modulate interannual fluctuations across the basins, while also identifying less intuitive predictors that act as proxies for large-scale variability. This explanatory capacity provides a transparent and scalable method for advancing cyclone genesis estimation in a changing climate, positioning XAI-GPI as a next-generation, explainable GPI for use in both projections and attribution contexts.

Plain Language Summary Tropical cyclones (TCs), also known as hurricanes or typhoons in their most intense versions, can have devastating impacts on communities, infrastructure, and ecosystems. Predicting when and where these storms will form remains a major scientific challenge especially in understanding why some years there are more or fewer cyclones. In this study, we present a new approach using machine learning to better detect and explain the conditions that lead to the formation of TCs across six ocean regions around the world. Our method combines climate and environmental data and identifies which factors are most important for predicting tropical cyclone formation. We also use a tool, called SHapley Additive exPlanations values, to explain how these factors influence TCs formation. Our method is more accurate than traditional tools in capturing year-to-year changes in tropical cyclone activity especially in the North Atlantic and Northeastern Pacific. These improvements can help scientists not only improve forecast accuracy but also better understand the physical processes that drive tropical cyclone formation.

1. Introduction

Every year, approximately 80–90 tropical cyclones (TCs) form worldwide (Walsh et al., 2016). Landfalling TCs rank among the costliest and deadliest natural hazards (Mendelsohn et al., 2012) due to their strong winds, heavy precipitation, and storm surges that cause coastal flooding. Accurately predicting TC activity and frequency across timescales, from short-term forecasts to long-term projections, is therefore crucial. In recent decades, advances in numerical weather prediction models and the increasing availability of high-quality observational data have steadily improved the short-term forecast of tropical cyclone genesis (TCG) (Knapp et al., 2010; Robertson et al., 2020). However, comparable progress in seasonal and climate-scale forecasts has lagged behind mainly due to the high computational cost of running large ensembles of high-resolution climate models over long periods (Cavicchia et al., 2023). To overcome these limitations, empirical Genesis Potential Indices (GPIs) have been developed. GPIs relate the likelihood of TCG to large-scale environmental variables—both dynamic and thermodynamic—combined through empirically derived formulas. Several versions exist (Ascenso et al., 2023;

Software: Filippo Dainelli, Jorge Pérez-Aracil
Supervision: Andrea Castelletti
Validation: Filippo Dainelli
Visualization: Filippo Dainelli, Jorge Pérez-Aracil, Guido Ascenso, Enrico Scoccimarro, Matteo Giuliani, Sancho Salcedo Sanz, Andrea Castelletti
Writing – original draft: Filippo Dainelli
Writing – review & editing: Jorge Pérez-Aracil, Guido Ascenso, Enrico Scoccimarro, Matteo Giuliani, Sancho Salcedo Sanz, Andrea Castelletti

Emanuel, 2003; Gray, 1979; Tippett et al., 2011; Wang & Murakami, 2020) designed to capture the relationship between climate variability and TC activity in both observations and models. Typically derived from monthly or seasonal mean fields, GPIs are better suited to capture climatological patterns than short-term fluctuations, and their agreement with observed cyclogenesis is expected to be the best at the seasonal timescale (Camargo, Sobel, et al., 2007). In fact, GPIs reproduce the spatial distribution and seasonal cycle of observed TCG events reasonably well (Jihoon et al., 2019; Yu et al., 2018).

Despite their usefulness, GPIs have key limitations. They often fail to capture the absolute frequency of TCs, both globally and at the basin scale (Sobel et al., 2021), and yield inconsistent projections under global warming—some indicating increases, others decreases, even when they behave similarly under present-day conditions (Camargo et al., 2014; Lee et al., 2020). GPIs are also sensitive to the data sets used for calibration, making them difficult to apply across different models or observations without ad hoc rescaling (Camargo, Emanuel, & Sobel, 2007). Furthermore, the choice of predictors is typically empirical not grounded in a unified theory of cyclogenesis, limiting interpretability and transferability (Sobel et al., 2021). As such, although GPIs remain useful for assessing spatial and seasonal TCG patterns, they struggle to capture interannual variability and offer robust projections, highlighting the need for improved methods and a deeper physical understanding of TCG processes (Cavicchia et al., 2023; Menkes et al., 2012).

In parallel, artificial intelligence (AI) is reshaping weather and climate science. Recent studies have shown its potential to improve forecasting skill (Bi et al., 2022; Kurth et al., 2023; Lang et al., 2024; Price et al., 2025) and enhance understanding of extreme events (Camps-Valls et al., 2025; Materia et al., 2024; Salcedo-Sanz et al., 2024). Machine learning (ML) and deep learning (DL) approaches have proven effective for detecting extremes in large climate data sets (Salcedo-Sanz et al., 2024). In the context of tropical cyclones, several works have applied ML to extend the concept of GPIs. For instance, Qian et al. (2022) developed a machine learning framework to project future changes in TCG frequency finding a robust reduction under warming scenarios. Fu et al. (2023) employed convolutional neural networks (CNNs) trained with transfer learning on large ensembles of high-resolution climate simulations and reanalyses using data at a seasonal temporal resolution. Their framework was applied to predictions, historical reconstructions, and climate projections demonstrating the flexibility of deep learning approaches for studying tropical cyclone activity. Similarly, Tong et al. (2025) used deep learning architectures to simulate TCG in the Northwest Pacific, showing that data-driven approaches can provide alternative formulations of genesis indices. These studies illustrate the potential of ML-based approaches to overcome some of the limitations of empirical GPIs but also highlight ongoing challenges in terms of interpretability and transferability across basins. In this regard, eXplainable AI (XAI) methods are increasingly being used to interpret complex models and uncover the physical drivers behind their predictions (Bommer et al., 2024; Dramsch et al., 2025; Yang et al., 2024), offering a way to bridge predictive skill with physical understanding.

Building on these developments and this line of research, we propose a robust and interpretable ML-based framework to build a novel index for seasonal TCG count detection: the XAI-GPI. The approach is designed to minimize predictor collinearity and support physical interpretation of TCG-climate relationships across major ocean basins. We employ a wrapper-based feature selection framework to identify the most informative variables for training a neural network (NN) model. SHapley Additive exPlanations (SHAP) are then used to quantify the contribution of each variable to the model output and explore their role in driving interannual variability in TC frequency. The feature selection process significantly reduces the input dimensionality without compromising performance improving model transparency and interpretability. Compared to standard GPIs, our model achieves better skill in estimating basin-aggregated TCG counts across all tropical regions. Traditional GPIs, such as the Emanuel and Nolan GPI, rely on a fixed set of predictors and predefined functional forms, combining dynamic and thermodynamic variables selected based on empirical evidence and expert judgment. Although effective in reproducing large-scale spatial and seasonal patterns, this rigid formulation limits their robustness and generalizability across basins and data sets. In contrast, our framework starts from a broad candidate pool of clustered environmental fields and climate indices and uses an optimization-based feature selection process to identify the most relevant predictors for each basin. This adaptive strategy reduces redundancy, captures specific regional dynamics, and improves robustness. By embedding SHAP analysis into the modeling workflow, the framework also ensures physical interpretability of the selected predictors. XAI-GPI thus provides a generalizable and more informative alternative to conventional GPIs particularly well-suited to capturing interannual variability and diagnosing the environmental drivers of TCG under nonstationary climate conditions. Beyond empirical indices, our approach also advances over existing ML-based GPIs, which have largely focused on predictive accuracy. By

Table 1
Candidate Variables for the Feature Selection Process

Variable name	Units	Acronym	Group
Absolute vorticity at 850 hPa	s^{-1}	abs_vo850	Atmosphere
Maximum potential intensity	ms^{-1}	Mpi	Atmosphere and ocean
Mean sea level pressure	hPa	Msl	Atmosphere
Relative humidity at 700 hPa	%	r700	Atmosphere
Sea surface temperature	°C	Sst	Ocean
Relative vorticity at 850 hPa	s^{-1}	vo850	Atmosphere
Vertical wind shear between 850 and 200 hPa	ms^{-1}	vws_850-200	Atmosphere
Vertical velocity at 500 hPa	Pas^{-1}	W	Atmosphere
Atlantic Meridional Mode	–	AMM	Climate indices
East Central Tropical Pacific sea surface temperature	°C	Niño3.4	Climate indices
North Atlantic Oscillation	–	NAO	Climate indices
Pacific Decadal Oscillation	–	PDO	Climate indices
Pacific North American Index	–	PNA	Climate indices
Southern Oscillation Index	–	SOI	Climate indices
Tropical Northern Atlantic Index	°C	TNA	Climate indices
Tropical Southern Atlantic Index	°C	TSA	Climate indices
Western Pacific Index	–	WPI	Climate indices
Month	–	Month	–

explicitly integrating explainability into the model design, XAI-GPI bridges estimation and interpretation ensuring skillful estimates of TCG counts and deeper insights into the environmental drivers that govern cyclone activity across different tropical basins. This basin-specific perspective enhances our understanding of how TCG variability manifests regionally, offering valuable information for both scientific research and climate services. The dual ability to achieve predictive performance while uncovering physically meaningful relationships highlights the importance of interpretability in evolving GPIs from diagnostic tools into explanatory frameworks for physical drivers of TC variability.

2. Data Description

In this work, we use three types of data: (a) monthly gridded maps of atmospheric and oceanic variables, (b) monthly time series of climate indices, and (c) observed counts of tropical cyclone genesis (TCG) events. The first two form the pool of candidate predictors for the feature selection process (Table 1), whereas TCG counts serves the target variable. All data sets cover the period from January 1980 to December 2021. We also include the month of the year as a candidate variable to control for the influence of the seasonal cycle during feature selection, ensuring that other predictors are selected based on their physical relevance rather than just for their seasonal periodicity. We apply our framework to six tropical basins (Figure 1) confined by the 40° parallels to the north and south. Each tropical basin has specific borders as determined by the definitions provided by Fudeyasu et al. (2014). For the North Atlantic and the Northeast Pacific, we apply these boundaries solely to crop the fields of the atmospheric and oceanic variables. However, for the TCG events in these areas, we use a segmented line along the Central America isthmus to appropriately divide the regions. For the other basins, the domain boundaries were used both for processing the fields of spatially distributed variables and for counting TCG events.

Atmospheric and oceanic predictors are derived from the ERA5 reanalysis data set (Hersbach et al., 2020) with a spatial resolution of 2.5° × 2.5°. The selection of candidate variables is guided by three main criteria. First, we choose variables known to influence TC formation and considered in previous GPI studies (Ascenso et al., 2023; Camargo, Sobel, et al., 2007; Camargo et al., 2014; Gray, 1968; Menkes et al., 2012; Meng & Garner, 2023; Tippett et al., 2011; Vu et al., 2021; Walsh et al., 2020; Wang, 2012; Zhao et al., 2012; Zong & Wu, 2023) to

10.1029/2025JH000865

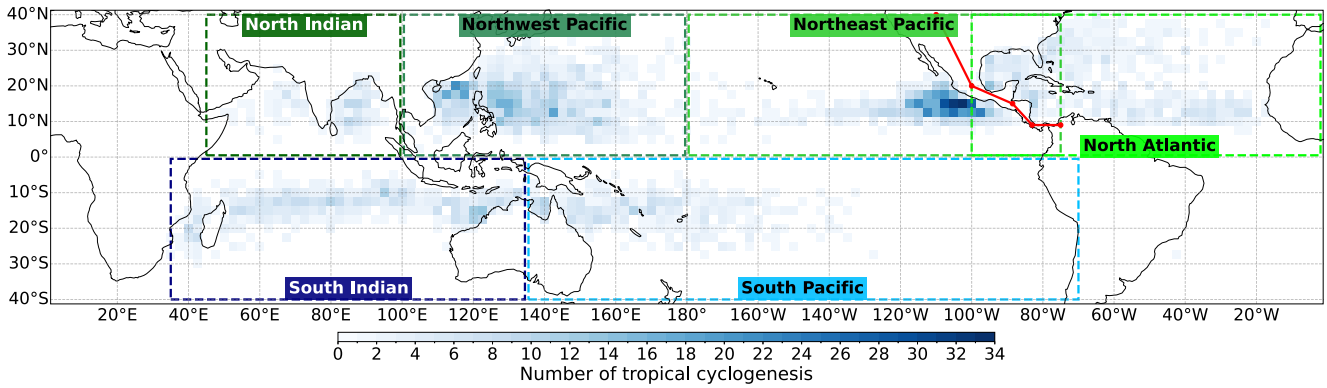

Figure 1. Geographical distribution of TCG events from 1980 to 2021, as recorded in the International Best Track Archive for Climate Stewardship (IBTrACS) data set (Knapp et al., 2010). The color scale indicates the number of events per grid cell. Rectangles mark the boundaries of the six tropical basins used in the study (North Atlantic [0°N, 100°W, 40°N, 0°E], Northeast Pacific [0°N, 180°W, 40°N, 75°W], Northwest Pacific [0°N, 100°E, 40°N, 180°E], North Indian [0°N, 45°E, 40°N, 100°E], South Indian [40°S, 35°E, 0°N, 135°E], and South Pacific [40°S, 135°E, 0°N, 70°W]). Red lines separate the North Atlantic and Northeast Pacific for the purpose of cyclone assignment.

ensure comparability with previous studies. Second, we include both primary and derived variables to assess whether feature selection favors added-value transformations or retains the original predictors. For instance, including both sea surface temperature and maximum potential intensity, which is partially computed from SST, helps evaluate whether MPI improves skill beyond its SST component. The MPI represents a theoretical estimate of the maximum sustained wind speed a TC could reach in a given environment. It was formulated by Emanuel (1986) and later modified by Bister and Emanuel (1998). Third, we balance the number of candidate features with the number of training examples available in our data set to ensure statistical robustness. We include a set of climate indices obtained from the NOAA Physical Sciences Laboratory (NOAA Physical Sciences Laboratory, 2023) selected based on their known or hypothesized links to TC activity and their frequent inclusion in previous studies on the subject (Camargo, Emanuel, & Sobel, 2007; Chylek & Lesins, 2008; Irwin III & Davis, 1999; Jiang & Li, 2021; Kang et al., 2019; Lee et al., 2021; Li et al., 2015, 2023; Liu et al., 2024; Patricola et al., 2014; Pérez-Alarcón et al., 2021; Song & Klotzbach, 2019; Vimont & Kossin, 2007; Wang & Chan, 2002; Xie et al., 2005; Zhao et al., 2023). Although no direct relationship with TC activity has been established, we also include the Western Pacific Index (WPI) due to its relevance to Pacific moisture transport (Park et al., 2018), which might provide new insights into tropical cyclone activity. The monthly number of TCG events within each basin is drawn from the IBTrACS (version v04r01) (Knapp et al., 2010), which provides 3-hourly best-track records. We define a TCG event as the first timestamp when a system's maximum sustained wind speed exceeds 34 knots. Figure 1 illustrates the spatial distribution of events in our data set aggregated by grid point on a 2.5° × 2.5° grid.

3. Methodology

We propose a machine learning (ML) framework for detecting tropical cyclone genesis (TCG) activity that is both robust and interpretable. This is the first application to combine spatial clustering, evolutionary feature selection, and explainable artificial intelligence (XAI) to construct a data-driven alternative to traditional Genesis Potential Indices (GPIs), which we refer to as the XAI-GPI. By leveraging the strengths of these complementary techniques, the framework adaptively identifies specific predictors for each basin from a broad set of candidates, enhances the understanding of how these predictors influence TCG frequency, and improves the detection of interannual variability in TCG activity. The framework is composed of five main components, each described in detail in Appendix A:

- Spatial clustering: Atmospheric and oceanic variables are spatially aggregated into regional time series using K-means clustering (MacQueen, 1967) reducing dimensionality while retaining relevant variability. We explore between 5 and 12 clusters and compute area-weighted spatial means for each cluster. These time series, along with the climate indices, form the input data set for feature selection.
- Feature selection: We adopt a wrapper method that combines and evolutionary algorithm, the probabilistic coral reef optimization algorithm with substrate layers (PCRO-SL) (Pérez-Aracil et al., 2023; Salcedo-

Table 2
Pearson Correlation Coefficient (R) Between the Detrended Anomalies of Annual Observations of Numbers of TCGs and the Detrended Anomalies of Annual Estimates Produced by the Models and GPIs

Basin	FS	noFS	ENGPI	oGPI
Globally aggregated	**0.546**	0.441	0.139	0.186
Northeast Pacific	**0.759**	0.600	0.684	0.601
Northwest Pacific	0.385	**0.408**	−0.037	0.066
North Atlantic	**0.690**	0.493	0.329	0.033
North Indian	0.339	**0.343**	−0.005	−0.030
South Indian	0.591	**0.603**	−0.141	−0.229
South Pacific	0.354	**0.417**	0.119	0.209

Note. Bold values indicate the highest R for each basin across all methods.

Sanz, 2017) with a regression model to identify the most informative subset of features for each basin. This evolutionary strategy promotes sparse solutions, limiting collinearity between selected features. Different cluster numbers and regression models are tested to ensure stability.

- Screening of selection settings: We evaluate multiple feature selection configurations and retain the most stable one used to define the final set of input variables.
- Model training: A feed-forward neural network (NN) with one hidden layer of 64 units is trained on the selected features. The same architecture is used for all basins and evaluated against benchmarks.
- Interpretation: We employ SHapley Additive exPlanations (SHAP) values (Lundberg & Lee, 2017) to attribute predictions to each features analyzing the learned relationships between predictors and TCG activity.

The resulting model is benchmarked against several baselines. These include two established GPIs, the Emanuel and Nolan GPI (ENGPI) (Emanuel & Nolan, 2004), and the optimized GPI (oGPI) (Ascenso et al., 2023), a refinement of ENGPI. The ENGPI is arguably the most studied GPI in the literature (Bruyère et al., 2012; Camargo, Sobel, et al., 2007; Camargo, Emanuel, & Sobel, 2007; Camargo et al., 2009, 2014; Jihoon et al., 2019; Meng & Garner, 2023; Menkes et al., 2012; Song et al., 2015; Tippett et al., 2011; Wang & Murakami, 2020; Yu et al., 2018; Zhang et al., 2020). We also include a neural network with the same architecture and hyperparameters as the XAI-GPI but trained on the full set of candidate predictors, denoted as *NoFS*. To evaluate performance, we analyze the detrended anomalies of annual TCG counts. This choice reflects our focus on interannual variability, a known challenge for existing GPIs, which were originally designed to reproduce the seasonal cycle and the spatial distribution of genesis locations (Camargo, Sobel, et al., 2007; Cavicchia et al., 2023). Detrending helps avoid inflating Pearson correlation due to long-term trends and enables a clearer assessment of year-to-year fluctuations. We first validate global-scale model outputs by aggregating the predictions across all basins and comparing them with GPIs and observations. We then perform a basin-by-basin comparison of detrended anomalies. Additional evaluations at monthly and seasonal time scales are provided in the Supporting Information S1. After evaluating model performance, we provide a detailed analysis of the final selected features and a physical interpretation of their relationships with the target variable. SHAP value plots are used to interpret each feature's contribution revealing both the strength and direction of its influence. Finally, we investigate specific years with anomalously high or low TC activity to evaluate how well the model captures shifts in the underlying drivers of interannual variability.

4. Results and Discussion

We first assess the performance of the proposed models and benchmark indices using the Pearson correlation coefficient between detrended annual anomalies of observed and estimated tropical cyclone genesis (TCG) counts. Results are presented globally and for each tropical basin (Table 2) with corresponding time series plots shown in Figure 2 for the global domain and the North Atlantic. Sections S2–S6 in Supporting Information S1 provide detailed results for other basins as well as additional analyses based on monthly and seasonal anomalies.

The performance analysis is followed by an examination of the selected features across basins with a focus on the North Atlantic. This basin is selected for detailed discussion due to its prominence in the literature and the particularly compelling results obtained by our approach. Spatial characteristics of the selected features and their relevance to TCG variability are discussed in Section 4.2, whereas feature attribution results using SHAP values are presented in Section 4.3. Corresponding analyses for the other five basins are provided in the Supporting Information S1.

Additionally, we performed supplementary experiments in which climate indices and teleconnection were excluded from the pool of candidate features. These tests were designed to assess whether the information conveyed by climate modes is already embedded in the environmental predictors. Although the main focus of this study remains on the configuration considering all the features, a summary of the additional results is presented in Tables S4 and S5 of Supporting Information S1. These findings indicate that excluding climate indices generally leads to a reduction in model skill, thus supporting their inclusion in the candidate feature set. The observed

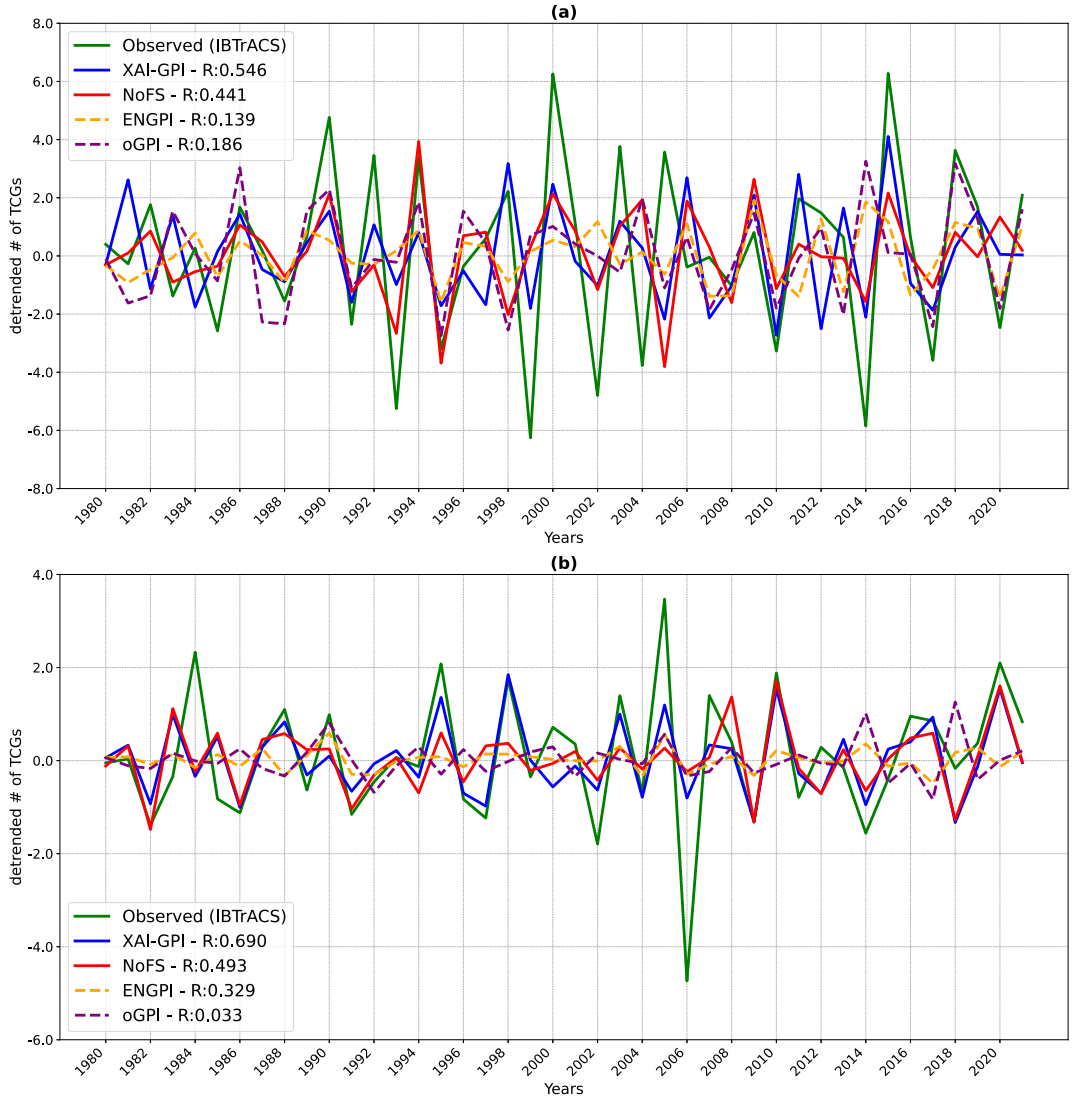

Figure 2. Detrended annual anomalies of observed and model-estimated TCG counts (1980–2021) for (a) the global domain and (b) over the North Atlantic basin.

differences in the final selected features, despite the absence of climate indices in the original selection, are likely caused by changes in the composition of the candidate feature pool. The optimization algorithm used (PCRO-SL) explores a rugged objective surface where modifying the candidate set can influence feature interactions, collinearity patterns, and the location of local optima. Since the final selected features are determined based on their presence in at least 75% of the top 20% performing models, even minor shifts in feature importance rankings can affect whether a given predictor crosses the selection threshold.

4.1. Detrended Annual Anomalies Analysis

At the global level, the neural networks trained on the selected features outperform all benchmarks in detecting the number of TCGs per year, including the *NoFS* models and the two GPIs (Table 2, first row). As shown in Figure 2a, the XAI-GPI models closely track the interannual fluctuations observed in TCG counts, capturing peaks and troughs more effectively than the flatter, low-amplitude responses produced by the GPIs. At the basin scale, XAI-GPI outperform the literature GPIs in all six regions when evaluating the aggregated TCG counts (Table 2, rows 2–7). In the Northeast Pacific and North Atlantic basins, they also outperform *NoFS* models achieving the highest overall skill. Figure 2b illustrates this for the North Atlantic where the XAI-GPI captures

Table 3
Final Selected Features for Each Basin

North Atlantic	North Indian	South Indian
abs_vo850_cluster1	abs_vo850_cluster3	mpi_cluster4
mpi_cluster4	abs_vo850_cluster4	r700_cluster5
mpi_cluster5	sst_cluster6	sst_cluster2
mpi_cluster6	sst_cluster7	vo850_cluster2
msl_cluster1	vo850_cluster4	vo850_cluster6
msl_cluster3	vo850_cluster5	w_cluster4
sst_cluster1	vws850-200_cluster1	NAO
vws850-200_cluster2	w_cluster1	PNA
vws850-200_cluster3	w_cluster3	TSA
vws850-200_cluster6	w_cluster7	WPI
w_cluster1		
w_cluster4		
w_cluster5		
Niño3.4		
NAO		
SOI		
month		
Northeast Pacific	**Northwest Pacific**	**South Pacific**
mpi_cluster7	abs_vo850_cluster9	abs_vo850_cluster4
r700_cluster5	vo850_cluster10	abs_vo850_cluster7
vo850_cluster2		r700_cluster6
vo850_cluster7		sst_cluster1
vws850-200_cluster7		PDO
w_cluster1		

interannual variability far more effectively than both GPIs, which largely fail to reproduce year-to-year changes. The temporal structure and amplitude of the XAI-GPI estimates align more closely with observations.

In the Northwest Pacific, North Indian, South Indian, and South Pacific basins, *NoFS* models achieve slightly higher correlations than their XAI-GPI counterparts. However, the differences are small and not statistically significant based on the test for dependent correlations proposed by Meng et al. (1992) (p-values: 0.825, 0.979, 0.731, and 0.661, respectively). A possible explanation lies in the relatively low number of selected features in these basins (Table 3), which may have excluded relevant signals. In contrast, *NoFS* models retain access to the full pool of predictors, including potentially informative, but less generalizable, features. Importantly, although XAI-GPI may exhibit marginally lower performance in some basins, they enable feature interpretability advantages that cannot be obtained from models using the full input set. Thus, the trade-off between performance and explainability is minimal and justified by the added value of interpretability especially in a scientific context where understanding the drivers of variability is a key objective. Moreover, the clustering step, while reducing dimensionality and limiting collinearity, can occasionally lead to information loss if the spatial aggregation is too coarse or if clusters do not align with physically coherent patterns. Future work could mitigate this limitation by exploring alternative clustering strategies or expanding the spatial coverage of candidate variables.

Although the XAI-GPI improves interannual skill relative to empirical GPIs across all basins, certain edge cases highlight its limitations. In the North Atlantic and South Indian Oceans (Figure S4 in Supporting Information S1), the model tends to underestimate years of extreme cyclone activity. This may indicate its inability to capture low-frequency variability associated with climate modes not included among the candidate features, such as the Atlantic Multidecadal Variability and the Indian Ocean Dipole. In the Northwest Pacific and South Pacific (Figures S2 and S5 in Supporting Information S1), the correlation coefficients remain modest with several years showing discrepancies between model estimates and observations particularly those influenced by ENSO in the late 1980s and early 1990s. The North Indian (Figure S3 in Supporting Information S1) basin exhibits overall low skill with strong positive anomalies in cyclone activity not reproduced by the model. This likely reflects both the limited sample size and the relatively small interannual variability in absolute terms, which can make correlation estimates highly sensitive to even minor estimation errors.

4.2. Feature Selection

Table 3 summarizes the final set of selected features across the six basins. Vorticity-related predictors specifically clusters of absolute vorticity or relative vorticity at 850 hPa are consistently selected in all basins, with the exception of the North Atlantic, where only one such cluster is included. This widespread selection aligns with previous findings: absolute vorticity has been identified as a key predictor in dynamic Genesis Potential Indices (B. Wang & Murakami, 2020), whereas relative vorticity is recognized as a dominant factor in the transition of precursor disturbances into tropical cyclones (Chen et al., 2018). Clusters of maximum potential intensity (MPI) are selected exclusively in the Northeast Pacific and the North Atlantic, while at least one cluster of SST is selected in all basins with the exception of the two located in the northern Pacific. In contrast, relative humidity at 700 hPa is infrequently selected with only a single cluster chosen in the North Atlantic, Northeast Pacific, and Northwest Pacific. This sparse selection supports previous findings, who highlight that the influence of mid-tropospheric humidity on cyclogenesis is basin-dependent and may exhibit a threshold effect, beyond which increases in humidity yield diminishing predictive value (Bruyère et al., 2012; Dowdy et al., 2012; Sharmila & Walsh, 2017). Mean sea level pressure appears to be relevant only in the North Atlantic where two clusters are selected. Clusters of vertical velocity at 500 hPa are selected in all basins except the Northwest Pacific and the South Pacific. Vertical wind shear between 850 and 200 hPa is particularly influential in the North Atlantic where

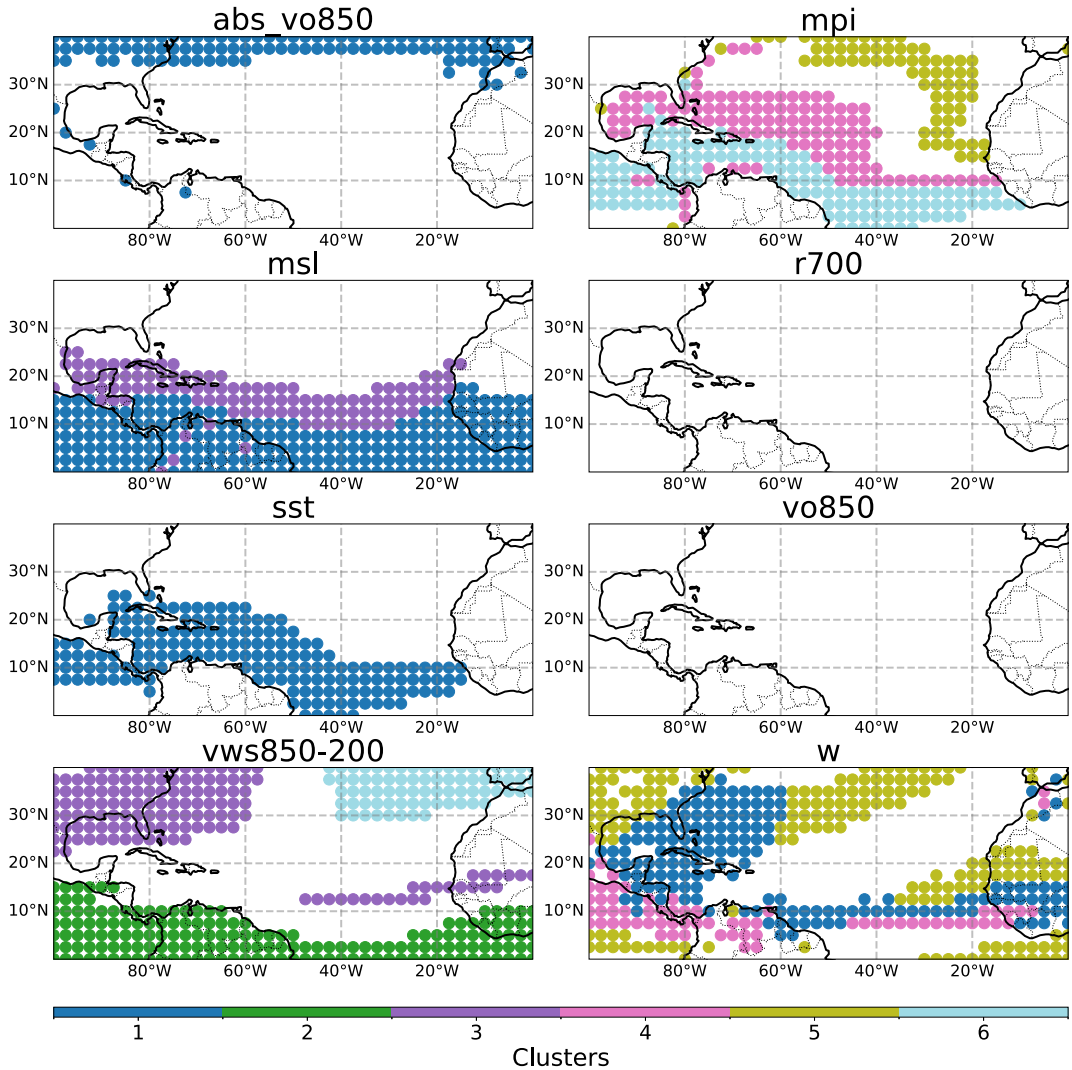

Figure 3. Spatial distribution of atmospheric and oceanic clusters selected by the feature selection framework for the North Atlantic basin. Each panel corresponds to one of the selected variables: *abs_vo850* (absolute vorticity at 850 hPa) cluster 1; *mpi* (maximum potential intensity) clusters 4, 5, and 6; *msl* (mean sea level pressure) clusters 1 and 3; *sst* (sea surface temperature) cluster 1; *vws850-200* (vertical wind shear between 850 and 200 hPa) clusters 2, 3, and 6; and *w* (vertical velocity at 500 hPa) clusters 1, 4, and 5.

three clusters are selected. One cluster is also selected in the Northeast Pacific and North Indian basins, whereas no clusters are chosen in the remaining basins. Climate indices and teleconnections are selected in only three basins: the North Atlantic, South Indian, and South Pacific with the North Atlantic Oscillation (NAO) being the only index selected in more than one basin. Other indices, including Niño3.4, SOI, PNA, TSA, WPI, and PDO, are each selected only once, whereas several others are excluded entirely. Regarding the selection of the month of the year, we observe that this feature is selected only in the North Atlantic. This suggests that, in general, the month does not provide significant additional information, as the seasonal signal associated with TCG is already largely captured by the other selected features. Overall, these results suggest that, in most basins, the feature selection algorithm prioritizes local atmospheric and oceanic variables over large-scale climate indices and teleconnection signals in explaining tropical cyclone genesis variability.

A more detailed analysis of the atmospheric and oceanic variable selected for the North Atlantic offers insight into the spatial patterns most strongly associated with tropical cyclone genesis. Figure 3 illustrates the spatial extent of the selected clusters for the eight oceanic and atmospheric key predictors. The three variables with the highest number of selected clusters are MPI, vertical wind shear between 850 and 200 hPa, and vertical velocity at

500 hPa, all factors consistently identified in the literature as central to TCG in this basin. For example, Yip and Yau (2012) developed a methodology using neural networks to project changes in North Atlantic TCG potential over the 21st century identifying vertical wind shear and MPI as dominant controls on TCG. Similarly, Sharmila and Walsh (2017) emphasized the role of wind shear in shaping TCG frequency in the North Atlantic. In our results, the selection of clusters 2 and 6 for vertical wind shear aligns with the findings of Zhao and Li (2019), who showed that wind shear associated with the Madden-Julian Oscillation (MJO) plays a key role in modulating TCG over the eastern part of the basin. Cluster 3, covering the Gulf of Mexico, one of the primary hotspots for TCG in the North Atlantic (see Figure 1) further underscores the spatial precision of the model. Meng and Garner (2023), using a sequential feature selection method, identified vertical wind shear and favorable heating conditions as leading factors influencing TCG. Their results also indicate that, in the North Atlantic, the most influential variables are MPI and SST. This is consistent with our findings, which include the selection of cluster 1 of SST, spanning the southern part of the North Atlantic, and clusters 4 and 6 of MPI covering areas associated with a significant concentration of TCG events. Furthermore, Meng and Garner (2023) highlight upward motion as the most critical variable for seed genesis regardless of basin and timescales, which supports our result showing the selection of three vertical velocity clusters in the North Atlantic. Seeds are defined as weak precursor disturbances of TC, sharing a few characteristics: cyclonic rotation at low to mid-levels, surface pressure lower than the surrounding environment, and the presence of deep convection (Hsieh et al., 2020; Sobel et al., 2021; Sugi et al., 2020). Wang and Murakami (2020) also identified vertical velocity at 500 hPa as one of the most influential factors contributing to TCG when constructing their index. Regarding mean sea level pressure, the clusters selected by the algorithm span much of the region between 0°N and 20°N, which aligns with the primary zones of TCG activity in the basin. In contrast, the selected cluster of absolute vorticity at 850 hPa covers the most northern portion of the domain. This selection suggests that the algorithm finds variability in absolute vorticity on the domain boundaries to be more informative for explaining fluctuations in TCG frequency than the clusters located within regions of higher event density. In contrast, no clusters of relative humidity at 700 hPa and relative vorticity at 850 hPa are selected in this basin. This seems to diverge from Sharmila and Walsh (2017), who identified a significant role for low-level relative vorticity in the Atlantic. Among the climate indices, the selection of Niño3.4 and SOI confirms the dominant influence of ENSO on TCG variability in line with long-standing evidence (Camargo, Emanuel, & Sobel, 2007; Gray, 1984; Gray et al., 1993; Knaff, 1997; Patricola et al., 2014; Tang & Neelin, 2004). In contrast, the selection of the NAO is somewhat less expected, as its influence has traditionally been associated with track and landfall rather than genesis (Xie et al., 2005). Nonetheless, recent studies (Ding et al., 2023) point to emerging SST-mediated teleconnections between the NAO and TCG, supporting its inclusion.

4.3. Explainability and Interpretation

To interpret the influence of each predictor on TCG estimates, we apply the SHAP (SHapley Additive exPlanations) to the XAI-GPI. To analyze the results, we generate beeswarm plots for each fold-specific neural network across the six basins. Figure 4 displays the SHAP results for the three networks trained on data from the North Atlantic. With the exception of cluster 4 of vertical velocity at 500 hPa (*w_cluster4*), feature rankings are consistent across the three folds underscoring the model's stability. The most influential features include cluster 1 of vertical velocity at 500 hPa (*w_cluster1*), cluster 5 of MPI (*mpi_cluster5*), cluster 3 of vertical wind shear between 850 and 200 hPa (*vws850-200_cluster3*), clusters 1 and 3 of mean sea level pressure (*msl_cluster1*, *msl_cluster3*), and cluster 1 of SST (*sst_cluster1*). For *mpi_cluster5*, *vws850-200_cluster3*, *msl_cluster3*, and *sst_cluster1*, the relationships with the target variable align with physical expectations: higher values of *mpi_cluster5* and *sst_cluster1* (red points) are associated with an increase in TCG (positive SHAP value on the *x*-axis), whereas higher values of *vws850-200_cluster3* and *msl_cluster3* are associated with a decrease (negative SHAP value on the *x*-axis). These physically consistent patterns are also observed for MPI clusters 4 and 6. In contrast, higher values of *w_cluster1* are associated with a decrease in the estimated number of TCGs, which may appear counterintuitive given that vertical velocity is typically linked to enhanced convection. However, this relationship is clarified by the correlation between the time series of vertical velocity in cluster 1 and the observed TCG counts in the North Atlantic, which is moderately strong and negative (Pearson correlation = −0.63). This suggests that, from a purely data-driven perspective, the model is learning a statistically consistent but physically counterintuitive relationship. This highlights one of the limitations of our approach: aside from the optional inclusion of a physically informed regression model during the feature selection phase (see Appendix A1), the rest of the framework is fully data-driven. As a result, the neural network trained on the selected features is not

10.1029/2025JH000865

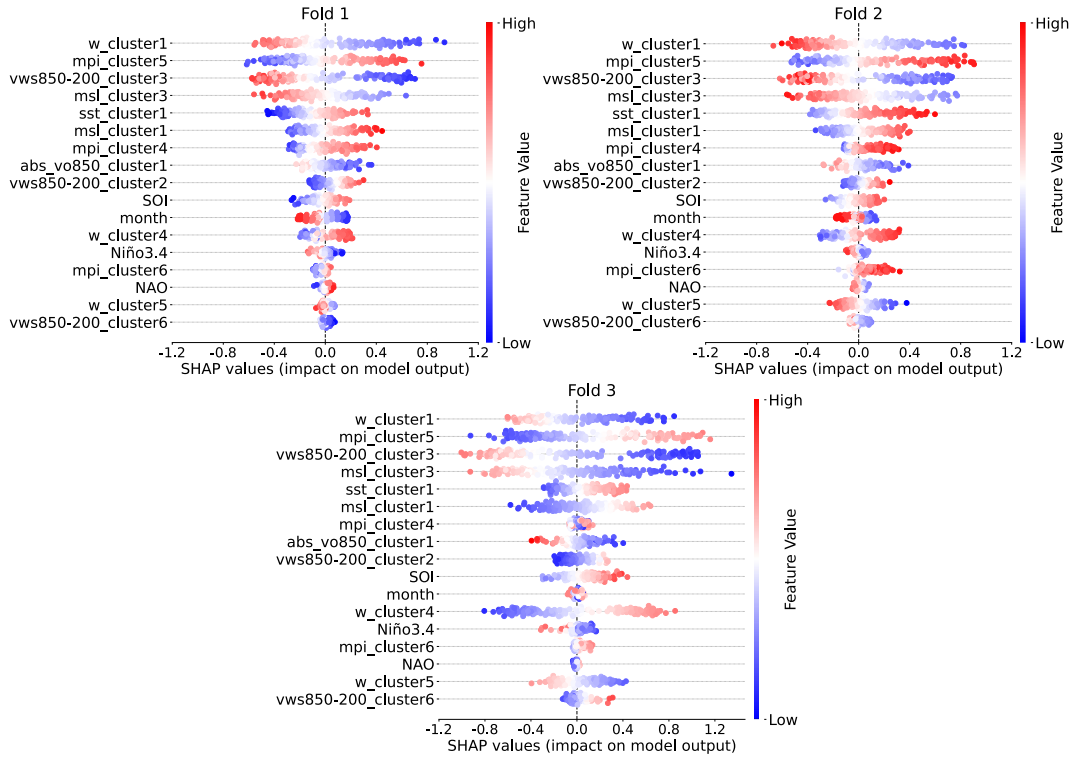

Figure 4. Beeswarm plot of SHAP values for the NN trained on the North Atlantic data. Each point represents the SHAP value of a single feature for a single observation. The spread of points along the *x*-axis reflects the influence of each feature on the model's output: the greater the absolute SHAP value, the stronger the feature's impact on the prediction. The sign of the SHAP values indicates whether a feature contributes positively (increasing the predicted number of TCG value) or negatively (decreasing the predicted value) to the model output. The color of each point reflects the relative magnitude of the feature value with red representing higher values and blue lower ones.

constrained by physical knowledge and instead learns patterns primarily based on statistical associations. This affects the interpretability of the relationships learned by the model, which are not guaranteed to align with the known physical processes driving TCG. In this case, the strong negative correlation between *w_cluster1* and TCG frequency dominates the model's learned behavior, even if it runs counter to establish physical understanding. Conversely, the more intuitive positive association between vertical velocity and cyclone occurrence is observed for *w_cluster4* where higher vertical velocity values lead to increased TCG particularly in the third fold's network. The interpretation of *msl_cluster1* remains less clear, as this region overlaps with key TCG zones but exhibits less consistent SHAP signals. Regarding the climate indices, the influence of both Niño3.4 and SOI on the target aligns with findings in the literature identifying ENSO as an influential driver on TC occurrences in the North Atlantic (Camargo, Emanuel, & Sobel, 2007).

To gain deeper insight into the factors driving interannual variability in TC frequency, we generated a specialized beeswarm plot of SHAP values (Figure 5). This plot focuses on the years in which the three-fold NNs estimated the highest and lowest numbers of cyclogenesis events representing periods of maximum and minimum cyclone activity. The goal of this analysis is to evaluate whether there are notable differences in the SHAP values distributions between years with different levels of TC activity. His analysis isolates key features that differentially influence the model during years of extreme activity. Three features exhibit substantial differences between years of maximum and minimum cyclone activity. For *w_cluster1*, we observe that, across all three folds, the SHAP values for the years of maximum cyclone activity are on average shifted further to the right. This indicates a generally stronger positive impact on the model output. Additionally, lower values of this feature are associated with higher SHAP values, meaning they exert greater influence on the prediction. These differences suggest that vertical velocity at 500 hPa in cluster 1 plays a particularly important role in modulating annual differences in tropical cyclone activity in the North Atlantic. *sst_cluster1* also exhibits noticeable differences in SHAP value distributions between years of maximum and minimum cyclone activity. In this case, the SHAP values for years

10.1029/2025JH000865

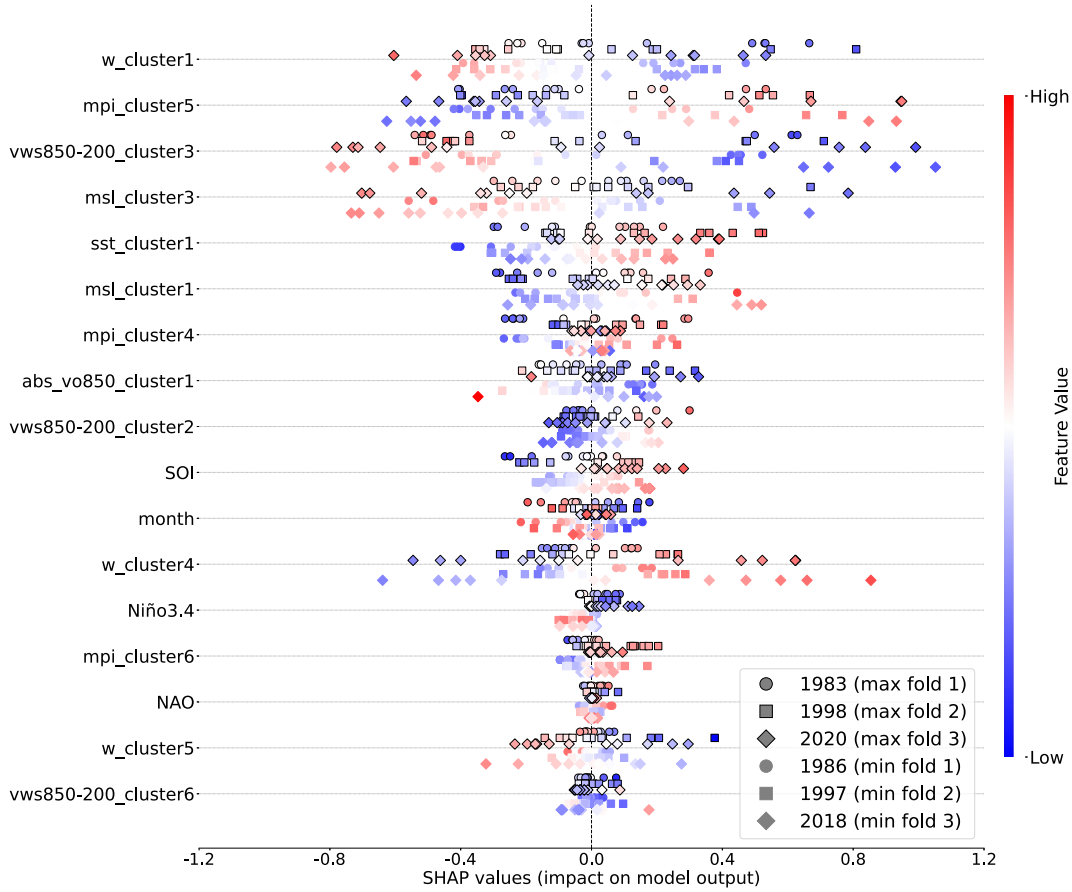

Figure 5. Beeswarm plot of SHAP values for the years in which three-fold NNs predicted maximum and minimum tropical cyclone activity in the North Atlantic. Each point represents the SHAP value of a single feature for an individual observation. The *x*-axis shows the impact of the feature on the model's prediction: larger absolute values indicate greater influence, whereas the sign denotes whether the effect is positive or negative. Point color shows the original feature value with red indicating high values and blue indicating low values. Points plotted above each feature axis correspond to SHAP values from months in years with maximum predicted activity, whereas those below correspond to months in years with minimum predicted activity.

with a higher number of cyclogenesis events are shifted to the right indicating a stronger positive influence on the model output. Furthermore, higher values of this feature correspond to a greater impact on estimates. This pattern suggests that higher average SST in the region of cluster 1 is associated with increased cyclone formation in the North Atlantic. During years when sea surface temperatures are higher over the 10°N–20°N latitude band, stronger cyclone activity is more likely. Niño3.4 also shows differences in SHAP value distributions between years of higher and lower cyclone activity. During years of maximum activity, lower values of the index (indicative of La Niña conditions) are associated with a positive impact on the number of cyclogenesis events. Conversely, in years of minimum activity, higher Niño3.4 values (reflecting El Niño conditions) are linked to a negative impact on TCG. This relationship identified by our model is consistent with previous findings that associate increased tropical cyclone activity in the North Atlantic with La Niña phases and with studies identifying ENSO as a major driver of TC interannual variations (Menkes et al., 2012).

To facilitate comparison, Table 4 summarizes the main characteristics of our proposed index alongside the two GPIs taken as reference from literature.

5. Conclusions

The objectives of this study were to improve the accuracy of annual tropical cyclone genesis (TCG) estimation, identify the most relevant environmental drivers of monthly cyclone formation, interpret the relationships

Table 4
Overview of the Main Characteristics of the Three Genesis Potential Indices (GPIs) Analyzed in This Study: ENGPI, oGPI, and XAI-GPI

	ENGPI	oGPI	XAI-GPI
Inputs	Fixed	Optimized ENGPI variant	Selected clusters and Climate indices
Input selection	Empirically chosen	Like ENGPI, but optimized pressure levels	Fully optimized from large set of candidates
Form	Analytical formula	Analytical formula	ML model (NN)
Derivation	Empirical	Genetic algorithm optimization	Gradient-based optimization
Output	Spatial genesis potential maps	Spatial genesis potential maps	Time series (monthly/annual TCG)
Interpretable?	Yes	Yes	Yes (via SHAP)
Captures interannual variability?	Weak (low skill in some basins)	Moderate	Strong (highest skill)

between predictors and output, and improve our understanding of interannual variability in TCG. To do so, we proposed an ML-based framework for the detection and interpretation of TCG activity across six tropical ocean basins to build a novel Genesis Potential Index (GPI): the XAI-GPI. The framework integrates spatial dimensionality reduction through clustering, feature selection via probabilistic coral reef optimization with substrate layers (PCRO-SL), and model interpretation using SHapley Additive exPlanations (SHAP) values.

The results demonstrate that, when considering basin-aggregated TCG counts, the XAI-GPI outperforms both the ENGPI and oGPI benchmarks in terms of Pearson correlation with observed annual TCG values. The global aggregation achieved the highest correlation and was the top-performing model in three of the six basins. In the other basins, their performance was comparable to models trained on the full feature set (*NoFS*), indicating that our feature selection method preserved the most informative predictors while enhancing interpretability with minimal loss in accuracy. The XAI-GPI also matched the performance of GPIs at monthly and seasonal scales further confirming their robustness across temporal resolutions. At the same time, some edge cases remain: the model underestimates years of extreme activity in the North Atlantic and South Indian basins, shows only modest skill in the Northwest Pacific and South Pacific, and exhibits limited performance in the North Indian Ocean. These limitations highlight the challenge of capturing basin specific extremes. In-depth analyses of the North Atlantic and Northeastern Pacific validated the physical relevance of the selected features and yielded additional insights. In the North Atlantic, the vertical wind shear over the Gulf of Mexico and the northwestern region of the basin strongly influences the output of the model as well as the MPI in the central and western area of the basin and the vertical velocity over the eastern region of the Northeastern Pacific. The latter, sea surface temperature over the Caribbean Sea and the south of the domain, and Niño3.4 play a crucial role in modulating the differences between years of higher and lower TC activity. In the Northeastern Pacific, the cluster of relative humidity covering part of the northern region of the domain emerges as one of the most influential features on the model's output. It also appears to play a crucial role in modulating the differences between years of maximum and minimum cyclone activity. In the remaining basins, some selected clusters had physical meaning and were in accordance with findings from existing literature. In some cases, the relationships learned by the models were counterintuitive particularly for localized clusters or features situated outside the main cyclogenesis regions. The selection of these features can be attributed to their strong negative correlation with the observed number of TCG events. In the remaining basins, the selected clusters are generally consistent with established knowledge of regional dynamics. For instance, vertical shear and mid-level humidity play an expected role in the South Indian Ocean, whereas ENSO-related predictors appear in the South Pacific. At the same time, some predictors appear in locations outside the core genesis regions or show counterintuitive relationships. In most of these cases, their importance stems from strong negative correlations with TCG counts, indicating that they act as proxies for larger-scale patterns rather than direct physical drivers.

These findings underscore the added value of the XAI-GPI framework: beyond improving prediction skill, it allows us to uncover and interpret the basin-specific climate features that modulate interannual variability of cyclone activity. By combining feature selection and SHAP explainability, the model provides a transparent view of the drivers of TCG, highlighting both well-known mechanisms and less intuitive proxy signals. This explanatory capacity is central to the framework's contribution, offering a pathway toward a more physically interpretable understanding of basin-scale cyclone variability.

Despite these promising results, the framework presents several limitations. In some cases, the clustering step produced small or singleton clusters complicating physical interpretation. Incorporating more advanced spatial clustering methods could improve coherence and alignment with known atmospheric patterns but would increase computational complexity and sensitivity to hyperparameters. K-means clustering, though simplistic, provided a stable and interpretable baseline for most applications. A second limitation concerns the spatial coverage of candidate variables. Expanding the geographic domain of predictors may improve feature selection especially in basins like the Northwestern Pacific where few features were ultimately retained. However, such expansion must balance the trade-off between increasing input dimensionality and the limited number of monthly training examples available. Lastly, although GPIs provide spatially explicit maps of TCG probability, our current approach outputs only basin-level time series. As such, our performance comparisons with ENGPI and oGPI refer specifically to the interannual variability of basin-wide activity rather than to spatial genesis patterns. Traditional GPIs therefore remain better suited for examining spatial genesis distributions, whereas XAI-GPI captures interannual variability and can be useful to reproduce the trends of tropical cyclone activity aggregated over each tropical basins in climate projection. Extending the framework to include spatially distributed predictions is a promising direction for future research and would further address the limitations of traditional GPI formulations.

Appendix A: Additional Methodological Details

A1. Spatial Cluster-Optimized Feature Selection (SCO-FS)

The Spatial Cluster-Optimized Feature Selection (SCO-FS) framework combines dimensionality reduction of spatial predictors through clustering with wrapper-based feature selection based on evolutionary optimization.

To reduce spatial dimensionality while preserving relevant temporal variability, we apply the K-means clustering algorithm (MacQueen, 1967) to each atmospheric and oceanic variable. Each cluster is represented by an area-weighted mean time series. This step enables tractable feature selection by reducing the number of spatial features. Given 504 monthly samples (1980–2021), we constrain the total number of candidate features to approximately 100 to avoid overfitting and instability. However, we also explore higher cluster counts (up to 12) as increased granularity often improves model performance. Alongside clustered fields, we include 10 supplementary variables, comprising climate indices, teleconnection patterns, and the month of the year.

Feature selection is performed independently for each basin using the probabilistic coral reef optimization algorithm with substrate layers (PCRO-SL) (Pérez-Aracil et al., 2023; Salcedo-Sanz, 2017), a multistrategy evolutionary method derived from the coral reef optimization (CRO) algorithm (Salcedo-Sanz et al., 2014). PCRO-SL incorporates several search heuristics—Harmony Search, Multipoint Crossover, XOR Operation, and BLX-α Crossover—running concurrently in distinct subpopulations, enabling more effective exploration of the solution space.

The optimization problem of selecting the optimal set of features is formulated as follows. Each solution generated by the PCRO-SL algorithm consists of a set of binary indicators defining whether a feature is selected along with time lag and sequence length parameters. The inclusion of this binary parameter promotes solutions that prioritize minimal information, thereby reducing the potential influence of collinearity between selected features. Since the model is trained on monthly data, we fix time lag and sequence length to 1 and 0, respectively, making the inclusion flag the only active parameter. The fitness of each feature subset is evaluated using a fast-learning regression model. We test three models: linear regression (LR), LightGBM (Ke et al., 2017), and a physically informed variant (PI-LGBM). The objective function is the mean squared error (MSE) on the training data (Section A2). We conduct multiple feature selection simulations to evaluate how the choice of regression model and the number of clusters affect the selection of relevant features. Specifically, we explore multiple values of k (5–12 clusters) to balance spatial detail and feature sparsity. Each additional cluster introduces 8 new features (one per spatial variable). For example, with $k = 8$, we obtain 64 spatial features. For robustness, five independent feature selection runs are performed per combination of regression model and k.

A2. Machine Learning Regression Model

To evaluate each feature selection run, we train a neural network (NN) using the selected features. These models, referred to as *evalNNs*, share the same architecture: a single hidden layer with 64 ReLU-activated units followed

by a linear output layer. Training uses MSE loss, L2 regularization ($\lambda = 0.001$), and the Adam optimizer (Kingma & Ba, 2017) with default parameters ($\beta_1 = 0.9, \beta_2 = 0.999$). Models are trained for 200 epochs with a batch size of 32. The final model for TCG detection adopts the same NN architecture and hyperparameters as the *evalNNs* and is trained on the final feature set selected in the most robust configuration (see Section A3).

A3. Filtering of Feature Selection Final Configuration and Sensitivity Analysis

The filtering process is designed to identify the most robust feature selection configuration. In this process, we evaluate each feature selection simulation according to the mean square error (MSE) of the corresponding *evalNNs*. All simulations are grouped according to two criteria: the regression model used within the PCRO-SL algorithm, and the number of clusters applied during the dimensionality reduction of spatially distributed variables. For each group, we generate violin plots illustrating the distribution of the *evalNNs'* average MSE values across the validation sets and compute the median MSE for comparison. This process results in three figures: Figure A1 displays the violin plots for different cluster counts when LR is used as the regression model; Figure A2 presents the corresponding plots for the LGBM; and Figure A3 shows the results for the PI-LGBM. After generating the violin plots and computing the median MSE values, we compare the different configurations and select the one with the lowest median MSE. In cases where two or more configurations have similar median values, we perform a more detailed comparison of the violin plots to identify the configuration with the more favorable distribution.

Following this procedure, we identify the final feature selection configuration for each basin:

- Northeast Pacific: LGBM with 7 clusters;
- Northwest Pacific: PI-LGBM with 10 clusters;
- North Atlantic: LR with 6 clusters;
- North Indian: PI-LGBM with 7 clusters;
- South Indian: LR with 6 clusters;
- South Pacific: PI-LGBM with 9 clusters.

Once the most robust feature selection configuration is identified through the filtering process, we proceed to determine the final set of selected features. This is done by analyzing, which features are most frequently selected among the best performing models under the chosen configuration. This final selection is based on the assumption that the feature selection configuration resulting in the *evalNNs* with the best average performance is the most suitable for building the final model and investigating the relationships between the selected features and the target variable. However, once the optimal configuration is identified for each basin, only five simulations are

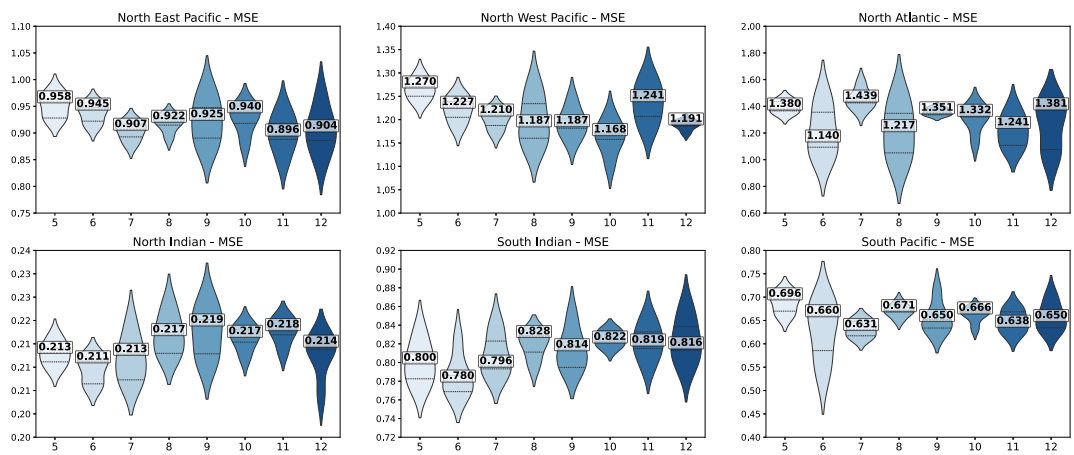

Figure A1. Violin plots showing the probability distributions of mean squared error (MSE) across the different number of clusters, with LR used as the model within the coral reef optimization algorithm. The distributions represent MSE values from feature selection simulations conducted for six different basins based on the performance of the *evalNN* trained using the features selected in each simulation. Each subplot corresponds to a specific basin, and the bold number in the white box within each plot indicates the median MSE value of the distribution.

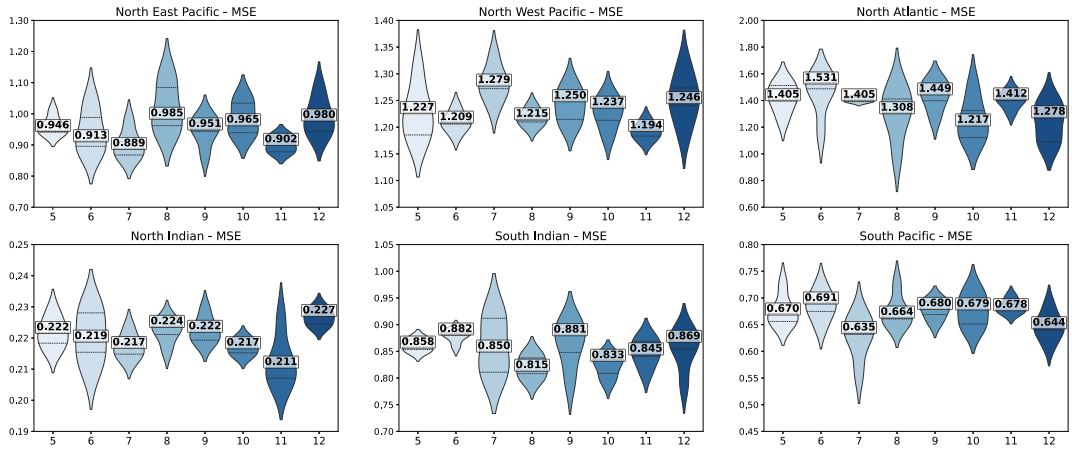

Figure A2. Violin plots showing the probability distributions of mean squared error (MSE) across the different number of clusters with LGBM used as the model within the coral reef optimization algorithm. The distributions represent MSE values from feature selection simulations conducted for six different basins based on the performance of the *evalNN* trained using the features selected in each simulation. Each subplot corresponds to a specific basin, and the bold number in the white box within each plot indicates the median MSE value of the distribution.

available to proceed with the final feature selection, meaning the selection process would rely on a relatively small sample of simulations, potentially limiting its robustness. To address this limitation and ensure a more robust and meaningful determination of the final selected features, we conduct an additional 95 simulations—each including the training of an *evalNN*—bringing the total to 100 simulations under the optimal configuration. The number of additional simulations was chosen as a compromise between achieving a sufficient sample size to have a significant computation of the frequencies of selection for the features among the best-performing runs and maintaining a feasible overall computation time. This expanded set of simulations not only improves the robustness of the final feature selection but also enhances the model's explainability and interpretability.

After completing all additional feature selection simulations, we proceed to identify the final set of selected features for each basin. This is done by ranking all simulations according to the performance of their corresponding *evalNNs*, measured by the Pearson correlation with observations at the annual scale and then dividing

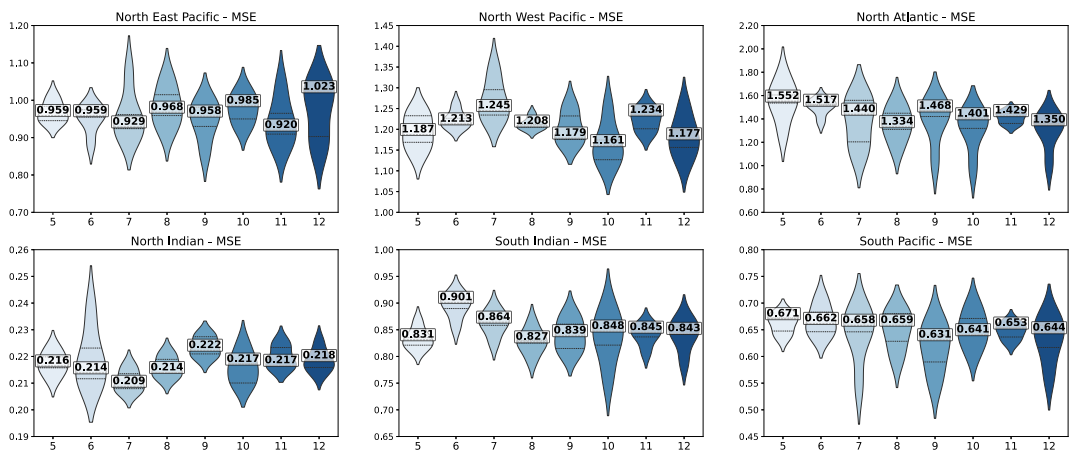

Figure A3. Violin plots showing the probability distributions of mean squared error (MSE) across the different number of clusters with PI-LGBM used as the model within the coral reef optimization algorithm. The distributions represent MSE values from feature selection simulations conducted for six different basins based on the performance of the *evalNN* trained using the features selected in each simulation. Each subplot corresponds to a specific basin, and the bold number in the white box within each plot indicates the median MSE value of the distribution.

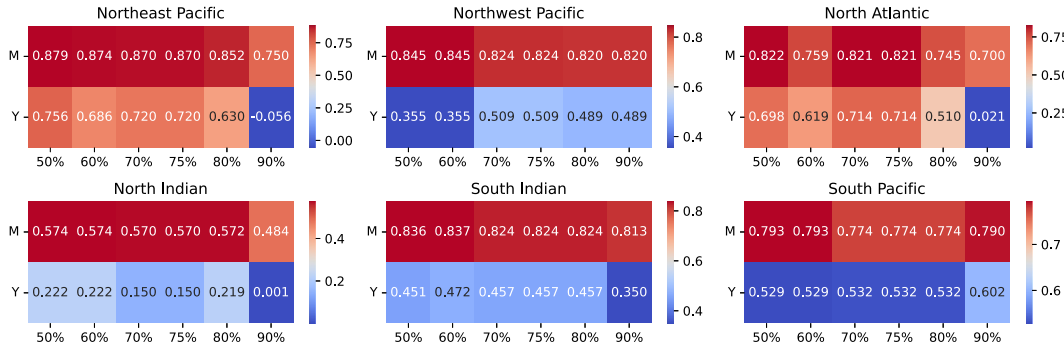

Figure A4. Monthly (M) and annual (Y) Pearson correlation coefficients between observations and the estimates produced by neural networks (NNs) using features selected at different percentage thresholds within the *Top 20%* of the simulation rankings. The *x*-axis indicates the selection thresholds, whereas the *y*-axis represents the two temporal scales. Each subplot corresponds to a specific basin and includes its own colorbar representing correlation values.

them into five performance subgroups: *Top20%*, *TopMid20%*, *Mid20%*, *MidBottom20%*, and *Bottom20%*. Based on the assumption that the best-performing models are also the most suitable for explanation and interpretation, we focus on the features most frequently selected within the *Top 20%* of the simulations rankings. The most frequently selected features are those that the algorithm identifies as having the greatest influence on the target variable. To quantitatively assess the selection frequency, we conducted a sensitivity analysis examining models performance in relation to the percentage of times each feature was selected within the *Top 20%* of the simulations rankings. The models used in the sensitivity analysis are NNs with the same architecture, hyperparameters, and data splitting scheme as the *evalNNs*, but their inputs consist of features selected in at least x% of the top-performing simulations. We evaluate six different selection thresholds: 50%, 60%, 70%, 75%, 80%, and 90%. Model performance is assessed using the Pearson correlation coefficient between the models' estimates and the observations, computed on the validation sets, both at the monthly and annual scales. The goal is to identify the highest possible selection threshold that does not result in a significant decline in performance. Figure A4 reports the monthly and annual Pearson correlation values for each selection threshold across all basins. Based on both the monthly and annual correlation results across all basins, we selected 75% as the common selection threshold.

A4. SHapley Additive exPlanations (SHAP) Values

We employ SHapley Additive exPlanations (SHAP) values (Lundberg & Lee, 2017) to interpret the contribution of each individual feature to the NN output. SHAP is a game-theoretic approach that assigns each feature an importance value based on its marginal contribution to the model's estimation averaged over all possible feature combinations. The method satisfies two key properties: local accuracy, meaning that the sum of SHAP values equals the model prediction for a given instance, and consistency, ensuring that a feature's attribution does not decrease when its influence increases in a newer model. SHAP values are computed for the final trained model and provide insights into whether each feature has a positive or negative effect on TCG prediction. The spread of SHAP values across samples reflects the feature's importance: wider distributions indicate stronger influence.

A5. Data Partitioning

Different components of the framework use different data splits. For clustering, feature selection, and *evalNNs* training, we use a time-based split: data from 1980 to 2013 are used for training and 2014–2021 for testing. Clustering is performed on the training set, and models within PCRO-SL are evaluated using 5-fold CV. For *evalNNs*, we use 3-fold CV to reduce computational time. To limit computational cost, we reduce the number of folds for the *evalNNs*. Although the regression models embedded in the feature selection algorithm are light-weight and can be trained rapidly regardless of the CV configuration, training neural networks is considerably more time-consuming. Using fewer folds for *evalNNs* thus prevents excessive computational burden when evaluating a large number of feature selection simulations. For the final model training, however, we adopt a different partitioning. Specifically, the full data set (1980–2021) is split into three CV folds, each defining a distinct train-test split. This setup allows for longer test periods, which are essential for evaluating model

performance. Model skill is evaluated using the Pearson correlation coefficient between the observed and predicted TCG counts at the annual timescale (Ascenso et al., 2023; Camargo, Emanuel, & Sobel, 2007; Cavicchia et al., 2023; Menkes et al., 2012; Yu et al., 2018). A longer test window ensures more robust correlation estimates and facilitates a consistent evaluation across basins and models.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

We collected the data of the atmospheric and oceanic variables from the *ERA5 reanalysis dataset* (Hersbach et al., 2020), the time series data of the climate indices and teleconnections from the *NOAA Physical Sciences Laboratory* (NOAA Physical Sciences Laboratory, 2023), and the tropical cyclones observations from the *International Best track Archive for Climate Stewardship dataset* (Knapp et al., 2010). The data set is openly available (Dainelli, 2025). The release version v1.0.0 of the eXplainable Artificial Intelligence Genesis Potential Index (XAI-GPI) framework is available via GNU General Public license (Dainelli et al., 2025). The XAI-GPI framework is deployed with TensorFlow and Keras API version 2.14.0 (Abadi et al., 2015; Chollet et al., 2015) available under the Apache license. Figures were made with Matplotlib version 3.10.1 (Hunter, 2007). Geospatial referencing and map visualizations were performed using Cartopy version 0.24.1 (Met Office, 2010–2015) available under the GNU Lesser General Public License. To compute the MPI, we use the `pyPI` Python package (Gilford, 2020, 2021).

Acknowledgments
This research was supported by the EU-funded Climate Intelligence (CLINT) project [Grant agreement ID:101003876; DOI: 10.3030/101003876]. This work used resources of the Deutsches Klimarechenzentrum (DKRZ) Granted by its Scientific Steering Committee (WLA) under project ID bk1318. Open access publishing facilitated by Politecnico di Milano, as part of the Wiley - CRUI-CARE agreement.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Retrieved from https://www.tensorflow.org/(Softwareavailablefromtensorflow.org

Ascenso, G., Cavicchia, L., Scoccimarro, E., & Castelletti, A. (2023). Optimization-based refinement of genesis indices for tropical cyclones. *Environmental Research Communications*, *5*(2), 021001. https://doi.org/10.1088/2515-7620/acb52a

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2022). Pangu-weather: A 3D high-resolution model for fast and accurate global. *Weather and Forecasting*. Retrieved from https://arxiv.org/abs/2211.02556

Bister, M., & Emanuel, K. A. (1998). Dissipative heating and hurricane intensity. *Meteorology and Atmospheric Physics*, *65*(3), 233–240. https://doi.org/10.1007/bf01030791

Bommer, P. L., Kretschmer, M., Hedström, A., Bareeva, D., & Höhne, M. M.-C. (2024). Finding the right Xai method—A guide for the evaluation and ranking of explainable AI methods in climate science. *Artificial Intelligence for the Earth Systems*, *3*(3), e230074. https://doi.org/10.1175/aies-d-23-0074.1

Bruyère, C. L., Holland, G. J., & Towler, E. (2012). Investigating the use of a genesis potential index for tropical cyclones in the north Atlantic basin. *Journal of Climate*, *25*(24), 8611–8626. https://doi.org/10.1175/jcli-d-11-00619.1

Camargo, S. J., Emanuel, K. A., & Sobel, A. H. (2007). Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *Journal of Climate*, *20*(19), 4819–4834. https://doi.org/10.1175/JCLI4282.1

Camargo, S. J., Sobel, A. H., Barnston, A. G., & Emanuel, K. A. (2007). Tropical cyclone genesis potential index in climate models. *Tellus A: Dynamic Meteorology and Oceanography*, *59*(4), 428–443. https://doi.org/10.1111/j.1600-0870.2007.00238.x

Camargo, S. J., Tippett, M. K., Sobel, A. H., Vecchi, G. A., & Zhao, M. (2014). Testing the performance of tropical cyclone genesis indices in future climates using the Hiram model. *Journal of Climate*, *27*(24), 9171–9196. https://doi.org/10.1175/jcli-d-13-00505.1

Camargo, S. J., Wheeler, M. C., & Sobel, A. H. (2009). Diagnosis of the MJO modulation of tropical cyclogenesis using an empirical index. *Journal of the Atmospheric Sciences*, *66*(10), 3061–3074. https://doi.org/10.1175/2009jas3101.1

Camps-Valls, G., Fernández-Torres, M.-Á., Cohrs, K.-H., Höhl, A., Adrianand, C., Pacal, A., et al. (2025). Artificial intelligence for modeling and understanding extreme weather and climate events. *Nature Communications*, *16*(1), 1919. https://doi.org/10.1038/s41467-025-56573-8

Cavicchia, L., Scoccimarro, E., Ascenso, G., Castelletti, A., Giuliani, M., & Gualdi, S. (2023). Tropical cyclone genesis potential indices in a new high-resolution climate models ensemble: Limitations and way forward. *Geophysical Research Letters*, *50*(11), e2023GL103001. https://doi.org/10.1029/2023GL103001

Chen, J.-M., Wu, C.-H., Chung, P.-H., & Sui, C.-H. (2018). Influence of intraseasonal–interannual oscillations on tropical cyclone genesis in the Western north Pacific. *Journal of Climate*, *31*(12), 4949–4961. https://doi.org/10.1175/jcli-d-17-0601.1

Chollet, F., et al. (2015). Keras. https://keras.io

Chylek, P., & Lesins, G. (2008). Multidecadal variability of Atlantic hurricane activity: 1851–2007. *Journal of Geophysical Research*, *113*(D22). https://doi.org/10.1029/2008JD010036

Dainelli, F. (2025). Clusters of atmopsheric and Oceanic variables and teleconnections that are candidate drivers for tropical cyclogenesis. *World Data Center for Climate (WDCC) at DKRZ*. https://doi.org/10.26050/wdcc/clint_tc

Dainelli, F., Pérez-Aracil, J., Ascenso, G., Scoccimarro, E., Giuliani, M., Salcedo-Sanz, S., & Castelletti, A. (2025). An explainable artificial intelligence genesis potential index (Xai-gpi) for tropical cyclone genesis detection. *Zenodo*. https://doi.org/10.5281/zenodo.17913240

Ding, R., Nnamchi, H. C., Yu, J.-Y., Li, T., Sun, C., Li, J., et al. (2023). North Atlantic oscillation controls multidecadal changes in the north tropical Atlantic-Pacific connection. *Nature Communications*, *14*(1), 862. https://doi.org/10.1038/s41467-023-36564-3

Dowdy, A. J., Qi, L., Jones, D., Ramsay, H., Fawcett, R., & Kuleshov, Y. (2012). Tropical cyclone climatology of the south Pacific Ocean and its relationship to El NIÑO–southern oscillation. *Journal of Climate*, *25*(18), 6108–6122. https://doi.org/10.1175/jcli-d-11-00647.1

Dramsch, J. S., Kuglitsch, M. M., Fernández-Torres, M.-Á., Toreti, A., Albayrak, R. A., Nava, L., et al. (2025). Explainability can foster trust in artificial intelligence in geoscience. *Nature Geoscience*, *18*(2), 112–114. https://doi.org/10.1038/s41561-025-01639-x

Emanuel, K. (2003). Tropical cyclones [Journal Article]. *Annual Review of Earth and Planetary Sciences*, *31*(31), 75–104. https://doi.org/10.1146/annurev.earth.31.100901.141259

Emanuel, K. A. (1986). An air-sea interaction theory for tropical cyclones. Part I: Steady-state maintenance. *Journal of the Atmospheric Sciences*, *43*(6), 585–605. https://doi.org/10.1175/1520-0469(1986)043⟨0585:aasitf⟩2.0.co;2

Emanuel, K. A., & Nolan, D. S. (2004). Tropical cyclone activity and the global climate system. In *26th conference on hurricanes and tropical meteorology*.

Fu, D., Chang, P., & Liu, X. (2023). Using convolutional neural network to emulate seasonal tropical cyclone activity. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2022MS003596. https://doi.org/10.1029/2022MS003596

Fudeyasu, H., Hirose, S., Yoshioka, H., Kumazawa, R., & Yamasaki, S. (2014). A global view of the landfall characteristics of tropical cyclones. *Tropical Cyclone Research and Review*, *3*(3), 178–192. https://doi.org/10.6057/2014TCRR03.04

Gilford, D. (2020). dgilford/pypi: pypi v1.3 (initial package release). *Zenodo*. https://doi.org/10.5281/zenodo.3985975

Gilford, D. M. (2021). pypi (v1.3): Tropical cyclone potential intensity calculations in python. *Geoscientific Model Development*, *14*(5), 2351–2369. https://doi.org/10.5194/gmd-14-2351-2021

Gray, W. M. (1968). Global view of the origin of tropical disturbances and storms. *Monthly Weather Review*, *96*(10), 669–700. https://doi.org/10.1175/1520-0493(1968)096⟨0669:GVOTOO⟩2.0.CO;2

Gray, W. M. (1979). *Meteorology over the tropical oceans* (pp. 155–218). Royal Meteorological Society.

Gray, W. M. (1984). Atlantic seasonal hurricane frequency. Part I: El Niño and 30 Mb quasi-biennial oscillation influences. *Monthly Weather Review*, *112*(9), 1649–1668. https://doi.org/10.1175/1520-0493(1984)112⟨1649:ashfpi⟩2.0.co;2

Gray, W. M., Landsea, C. W., Mielke, P. W., & Berry, K. J. (1993). Predicting Atlantic basin seasonal tropical cyclone activity by 1 August. *Weather and Forecasting*, *8*(1), 73–86. https://doi.org/10.1175/1520-0434(1993)008⟨0073:pabstc⟩2.0.co;2

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The Era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hsieh, T.-L., Vecchi, G. A., Yang, W., Held, I. M., & Garner, S. T. (2020). Large-scale control on the frequency of tropical cyclones and seeds: A consistent relationship across a hierarchy of global atmospheric models. *Climate Dynamics*, *55*(11), 3177–3196. https://doi.org/10.1007/s00382-020-05446-5

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science and Engineering*, *9*(3), 90–95. https://doi.org/10.1109/mcse.2007.55

Irwin, I. I. I., & Davis, R. E. (1999). The relationship between the southern oscillation index and tropical cyclone tracks in the eastern north Pacific. *Geophysical Research Letters*, *26*(15), 2251–2254. https://doi.org/10.1029/1999GL900533

Jiang, L., & Li, T. (2021). Impacts of tropical north Atlantic and equatorial Atlantic SST anomalies on ENSO. *Journal of Climate*, *34*(14), 5635–5655. https://doi.org/10.1175/jcli-d-20-0835.1

Jihoon, S., Song, C., Kim, S., & Park, S. (2019). A heuristic estimation of the genesis probability of tropical cyclones using genesis frequency and genesis potential index. *Journal of the Korean Earth Science Society*, *40*(6), 561–571. https://doi.org/10.5467/jkess.2019.40.6.561

Kang, N.-Y., Kim, D., & Elsner, J. B. (2019). The contribution of super typhoons to tropical cyclone activity in response to ENSO. *Scientific Reports*, *9*(1), 5046. https://doi.org/10.1038/s41598-019-41561-y

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization.

Knaff, J. A. (1997). Implications of summertime sea level pressure anomalies in the tropical Atlantic region. *Journal of Climate*, *10*(4), 789–804. https://doi.org/10.1175/1520-0442(1997)010⟨0789:iosslp⟩2.0.co;2

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The international best track archive for climate stewardship (IBTRACS): Unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, *91*(3), 363–376. https://doi.org/10.1175/2009bams2755.1

Kurth, T., Subramanian, S., Harrington, P., Pathak, J., Mardani, M., Hall, D., et al. (2023). Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive Fourier neural operators. In *Proceedings of the platform for advanced scientific computing conference*. Association for Computing Machinery. https://doi.org/10.1145/3592979.3593412

Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., et al. (2024). Aifs—Ecmwf's data-driven forecasting system. Retrieved from https://arxiv.org/abs/2406.01465

Lee, C.-Y., Camargo, S. J., Sobel, A. H., & Tippett, M. K. (2020). Statistical–dynamical downscaling projections of tropical cyclone activity in a warming climate: Two diverging genesis scenarios. *Journal of Climate*, *33*(11), 4815–4834. https://doi.org/10.1175/jcli-d-19-0452.1

Lee, M., Kim, T., Cha, D.-H., Min, S.-K., Park, D.-S. R., Yeh, S.-W., & Chan, J. C. L. (2021). How does pacific decadal oscillation affect tropical cyclone activity over far east Asia? *Geophysical Research Letters*, *48*(24), e2021GL096267. https://doi.org/10.1029/2021GL096267

Li, W., Li, L., & Deng, Y. (2015). Impact of the interdecadal Pacific oscillation on tropical cyclone activity in the north Atlantic and eastern north Pacific. *Scientific Reports*, *5*(1), 12358. https://doi.org/10.1038/srep12358

Li, Y., Tang, Y., Li, X., Song, X., & Wang, Q. (2023). Recent increase in the potential threat of western north Pacific tropical cyclones. *npj Climate and Atmospheric Science*, *6*(1), 53. https://doi.org/10.1038/s41612-023-00379-2

Liu, C., An, S.-I., Kim, S.-K., Stuecker, M. F., Zhang, W., Jin, F.-F., et al. (2024). Synchronous decadal climate variability in the tropical central pacific and tropical south Atlantic. *npj Climate and Atmospheric Science*, *7*(1), 253. https://doi.org/10.1038/s41612-024-00806-y

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Retrieved from https://arxiv.org/abs/1705.07874

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth berkeley symposium on mathematical statistics and probability* (Vol. 5, pp. 281–298). Statistics.

Materia, S., García, L. P., van Straaten, C., O, S., Mamalakis, A., Cavicchia, L., et al. (2024). Artificial intelligence for climate prediction of extremes: State of the art, challenges, and future perspectives. *WIREs Climate Change*, *15*(6), e914. https://doi.org/10.1002/wcc.914

Mendelsohn, R., Emanuel, K., Chonabayashi, S., & Bakkensen, L. (2012). The impact of climate change on global tropical cyclone damage. *Nature Climate Change*, *2*(3), 205–209. https://doi.org/10.1038/nclimate1357

Meng, L., & Garner, S. T. (2023). Nonlocal controls on tropical cyclogenesis: A trajectory-based genesis potential index. *Journal of the Atmospheric Sciences*, *80*(12), 2925–2946. https://doi.org/10.1175/jas-d-23-0025.1

Meng, X.-L., Rosenthal, R., & Rubin, D. (1992). 01). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*(1), 172–175. https://doi.org/10.1037/0033-2909.111.1.172

Menkes, C. E., Lengaigne, M., Marchesiello, P., Jourdain, N. C., Vincent, E. M., Lefèvre, J., et al. (2012). Comparison of tropical cyclogenesis indices on seasonal to interannual timescales. *Climate Dynamics*, *38*(1), 301–321. https://doi.org/10.1007/s00382-011-1126-x

Met Office. (2010–2015). Cartopy: A cartographic python library with a matplotlib interface [Computer Software Manual]. *Exeter, Devon*. Retrieved from https://scitools.org.uk/cartopy

NOAA Physical Sciences Laboratory. (2023). Climate indices: Monthly atmospheric and ocean time series. Retrieved from https://psl.noaa.gov/data/climateindices/list/

Park, Y.-H., Kim, B.-M., Pak, G., Yamamoto, M., Vivier, F., & Durand, I. (2018). A key process of the nonstationary relationship between ENSO and the western pacific teleconnection pattern. *Scientific Reports*, *8*(1), 9512. https://doi.org/10.1038/s41598-018-27906-z

Patricola, C. M., Saravanan, R., & Chang, P. (2014). The impact of the El Niño–southern oscillation and Atlantic meridional mode on seasonal Atlantic tropical cyclone activity. *Journal of Climate*, *27*(14), 5311–5328. https://doi.org/10.1175/jcli-d-13-00687.1

Pérez-Alarcón, A., Fernández-Alvarez, J. C., Sorí, R., Nieto, R., & Gimeno, L. (2021). The relationship of the sea surface temperature and climate variability modes with the north Atlantic tropical cyclones activity. *Revista cubana de meteorología*, *27*(3), 1–15.

Pérez-Aracil, J., Camacho-Gómez, C., Lorente-Ramos, E., Marina, C. M., Cornejo-Bueno, L. M., & Salcedo-Sanz, S. (2023). New probabilistic, dynamic multi-method ensembles for optimization based on the Cro-sl. *Mathematics*, *11*(7), 1666. https://doi.org/10.3390/math11071666

Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D., et al. (2025). Probabilistic weather forecasting with machine learning. *Nature*, *637*(8044), 84–90. https://doi.org/10.1038/s41586-024-08252-9

Qian, Q., Jia, X., & Lin, Y. (2022). Reduced tropical cyclone genesis in the future as predicted by a machine learning model. *Earth's Future*, *10*(2), e2021EF002455. https://doi.org/10.1029/2021EF002455

Robertson, A. W., Vitart, F., & Camargo, S. J. (2020). Subseasonal to seasonal prediction of weather to climate with application to tropical cyclones. *Journal of Geophysical Research: Atmospheres*, *125*(6), e2018JD029375. https://doi.org/10.1029/2018JD029375

Salcedo-Sanz, S. (2017). A review on the coral reefs optimization algorithm: New development lines and current applications. *Progress in Artificial Intelligence*, *6*(1), 1–15. https://doi.org/10.1007/s13748-016-0104-2

Salcedo-Sanz, S., Del Ser, J., Landa-Torres, I., Gil-López, S., & Portilla-Figueras, J. A. (2014). The coral reefs optimization algorithm: A novel metaheuristic for efficiently solving optimization problems. *The Scientific World Journal*, *2014*(1), 739768. https://doi.org/10.1155/2014/739768

Salcedo-Sanz, S., Pérez-Aracil, J., Ascenso, G., Del Ser, J., Casillas-Pérez, D., Kadow, C., et al. (2024). Analysis, characterization, prediction, and attribution of extreme atmospheric events with machine learning and deep learning techniques: A review. *Theoretical and Applied Climatology*, *155*(1), 1–44. https://doi.org/10.1007/s00704-023-04571-5

Sharmila, S., & Walsh, K. J. E. (2017). Impact of large-scale dynamic versus thermodynamic climate conditions on contrasting tropical cyclone genesis frequency. *Journal of Climate*, *30*(22), 8865–8883. https://doi.org/10.1175/jcli-d-16-0900.1

Sobel, A. H., Wing, A. A., Camargo, S. J., Patricola, C. M., Vecchi, G. A., Lee, C.-Y., & Tippett, M. K. (2021). Tropical cyclone frequency. *Earth's Future*, *9*(12), e2021EF002275. https://doi.org/10.1029/2021EF002275

Song, J., & Klotzbach, P. J. (2019). Relationship between the pacific-north american pattern and the frequency of tropical cyclones over the western north pacific. *Geophysical Research Letters*, *46*(11), 6118–6127. https://doi.org/10.1029/2019GL082925

Song, Y., Wang, L., Lei, X., & Wang, X. (2015). Tropical cyclone genesis potential index over the western north pacific simulated by CMIP5 models. *Advances in Atmospheric Sciences*, *32*(11), 1539–1550. https://doi.org/10.1007/s00376-015-4162-3

Sugi, M., Yamada, Y., Yoshida, K., Mizuta, R., Nakano, M., Kodama, C., & Satoh, M. (2020). Future changes in the global frequency of tropical cyclone seeds. *Inside Solaris*, *16*, 70–74. https://doi.org/10.2151/sola.2020-012

Tang, B. H., & Neelin, J. D. (2004). ENSO influence on Atlantic hurricanes via tropospheric warming. *Geophysical Research Letters*, *31*(24). https://doi.org/10.1029/2004GL021072

Tippett, M. K., Camargo, S. J., & Sobel, A. H. (2011). A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *Journal of Climate*, *24*(9), 2335–2357. https://doi.org/10.1175/2010jcli3811.1

Tong, B., Hu, G., Deng, Y., Huang, Y., & He, Y. (2025). Deep-learning based simulation of tropical cyclone genesis in northwest pacific. *Journal of Wind Engineering and Industrial Aerodynamics*, *257*, 106003. https://doi.org/10.1016/j.jweia.2024.106003

Vimont, D. J., & Kossin, J. P. (2007). The Atlantic meridional mode and hurricane activity. *Geophysical Research Letters*, *34*(7). https://doi.org/10.1029/2007GL029683

Vu, T.-A., Kieu, C., Chavas, D., & Wang, Q. (2021). A numerical study of the global formation of tropical cyclones. *Journal of Advances in Modeling Earth Systems*, *13*(1), e2020MS002207. https://doi.org/10.1029/2020MS002207

Walsh, K. J., McBride, J. L., Klotzbach, P. J., Balachandran, S., Camargo, S. J., Holland, G., et al. (2016). Tropical cyclones and climate change. *WIREs Climate Change*, *7*(1), 65–89. https://doi.org/10.1002/wcc.371

Walsh, K. J. E., Sharmila, S., Thatcher, M., Wales, S., Utembe, S., & Vaughan, A. (2020). Real world and tropical cyclone world. Part ii: Sensitivity of tropical cyclone formation to uniform and meridionally varying sea surface temperatures under aquaplanet conditions. *Journal of Climate*, *33*(4), 1473–1486. https://doi.org/10.1175/jcli-d-19-0079.1

Wang, B., & Chan, J. C. L. (2002). How strong enso events affect tropical storm activity over the Western north Pacific. *Journal of Climate*, *15*(13), 1643–1658. https://doi.org/10.1175/1520-0442(2002)015⟨1643:hseeat⟩2.0.co;2

Wang, B., & Murakami, H. (2020). Dynamic genesis potential index for diagnosing present-day and future global tropical cyclone genesis. *Environmental Research Letters*, *15*(11), 114008. https://doi.org/10.1088/1748-9326/abbb01

Wang, Z. (2012). Thermodynamic aspects of tropical cyclone formation. *Journal of the Atmospheric Sciences*, *69*(8), 2433–2451. https://doi.org/10.1175/jas-d-11-0298.1

Xie, L., Yan, T., Pietrafesa, L. J., Morrison, J. M., & Karl, T. (2005). Climatology and interannual variability of north Atlantic hurricane tracks. *Journal of Climate*, *18*(24), 5370–5381. https://doi.org/10.1175/jcli3560.1

Yang, R., Hu, J., Li, Z., Mu, J., Yu, T., Xia, J., et al. (2024). Interpretable machine learning for weather and climate prediction: A survey. Retrieved from https://arxiv.org/abs/2403.18864

Yip, Z. K., & Yau, M. K. (2012). Application of artificial neural networks on north Atlantic tropical cyclogenesis potential index in climate change. *Journal of Atmospheric and Oceanic Technology*, *29*(9), 1202–1220. https://doi.org/10.1175/jtech-d-11-00178.1

Yu, J.-Y., Hsiao, L.-P., & Chiu, P.-G. (2018). Evaluating the emanuel-nolan genesis potential index: Contrast between north Atlantic and Western north Pacific. *Terrestrial, Atmospheric and Oceanic Sciences*, *29*(2), 201–214. https://doi.org/10.3319/tao.2017.09.27.01

Zhang, W., Gao, J., Cheung, K. K. W., Chen, J.-M., & You, L. (2020). Intraseasonal variability of the genesis potential index and its relationship with tropical cyclogenesis over the Western north Pacific. *International Journal of Climatology*, *40*(1), 22–35. https://doi.org/10.1002/joc.6190

Zhao, C., & Li, T. (2019). Basin dependence of the mjo modulating tropical cyclone genesis. *Climate Dynamics*, *52*(9), 6081–6096. https://doi.org/10.1007/s00382-018-4502-y

Zhao, J., Zhan, R., Murakami, H., Wang, Y., Xie, S.-P., Zhang, L., & Guo, Y. (2023). Atmospheric modes fiddling the simulated ENSO impact on tropical cyclone genesis over the northwest Pacific. *npj Climate and Atmospheric Science*, *6*(1), 213. https://doi.org/10.1038/s41612-023-00537-6

Zhao, M., Held, I. M., & Lin, S.-J. (2012). Some counterintuitive dependencies of tropical cyclone frequency on parameters in a GCM. *Journal of the Atmospheric Sciences*, *69*(7), 2272–2283. https://doi.org/10.1175/jas-d-11-0238.1

Zong, H., & Wu, L. (2023). What affects the timing of tropical cyclone formation within a monsoon trough? *Frontiers in Earth Science*, *10*, 1046107. https://doi.org/10.3389/feart.2022.1046107

References From the Supporting Information

Gao, S., Chen, Z., & Zhang, W. (2018). Impacts of tropical north Atlantic SST on Western north Pacific landfalling tropical cyclones. *Journal of Climate*, *31*(2), 853–862. https://doi.org/10.1175/jcli-d-17-0325.1

Liu, Y., Huang, P., & Chen, G. (2019). Impacts of the combined modes of the tropical indo-pacific sea surface temperature anomalies on the tropical cyclone genesis over the Western north Pacific. *International Journal of Climatology*, *39*(4), 2108–2119. https://doi.org/10.1002/joc.5938

Wu, L., Zhang, H., Chen, J.-M., & Feng, T. (2018). Impact of two types of El NIÑO on tropical cyclones over the Western north Pacific: Sensitivity to location and intensity of Pacific warming. *Journal of Climate*, *31*(5), 1725–1742. https://doi.org/10.1175/jcli-d-17-0298.1

Wu, R., Cao, X., & Yang, Y. (2020). Interdecadal change in the relationship of the Western north Pacific tropical cyclogenesis frequency to tropical Indian and north Atlantic Ocean SST in early 1990s. *Journal of Geophysical Research: Atmospheres*, *125*(2), e2019JD031493. https://doi.org/10.1029/2019JD031493

Wu, R., Yang, Y., & Cao, X. (2019). Respective and combined impacts of regional SST anomalies on tropical cyclogenesis in different sectors of the Western north Pacific. *Journal of Geophysical Research: Atmospheres*, *124*(16), 8917–8934. https://doi.org/10.1029/2019JD030736

Zhan, R., Wang, Y., & Zhao, J. (2019). Contributions of SST anomalies in the indo-pacific ocean to the interannual variability of tropical cyclone genesis frequency over the western north Pacific. *Journal of Climate*, *32*(11), 3357–3372. https://doi.org/10.1175/jcli-d-18-0439.1