Atmospheric Research 327 (2026) 108404

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

Assessment of synthetic tropical cyclones in the North Atlantic Basin

David Romero [a], Christian M. Appendini [b,c,*], Kerry Emanuel [d], Chia-Ying Lee [e], Kees Nederhoff [f], Nadia Bloemendaal [g,h], Pablo Ruiz-Salcines [b,i], Jonathan Vigh [j], Christian Domínguez [k]

[a] Escuela Nacional de Estudios Superiores Unidad Mérida, Universidad Nacional Autónoma de México, Ucú, Yucatán, Mexico
[b] Laboratorio de Ingeniería y Procesos Costeros, Instituto de Ingeniería, Universidad Nacional Autónoma de México, Puerto de abrigo s/n, Sisal, Yucatán 97356, Mexico
[c] Aarhus Institute of Advanced Studies, Aarhus University, 1630, Høegh-Guldbergs Gade, Aarhus 8000, Denmark
[d] Lorenz Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[e] Lamont-Doherty Earth Observatory, Columbia University, USA
[f] Deltares USA, 8601 Georgia Ave, Silver Spring, MD 20910, USA
[g] Royal Netherlands Meteorological Institute, De Bilt, the Netherlands
[h] Institute for Environmental Studies, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[i] Centre for Climate Change, Geography Department, Universitat Rovira i Virgili, Tarragona, Spain
[j] U.S. National Science Foundation, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80301, USA
[k] Instituto de Ciencias de la Atmósfera y Cambio Climático, UNAM, Mexico

ARTICLE INFO

Keywords:
Tropical cyclones
Synthetic events
North Atlantic Basin
Hurricanes
Hazards

ABSTRACT

Tropical cyclones (TCs) pose significant risks due to their associated hazards, including powerful winds, inland and coastal flooding, and wind waves. However, more reliable TC records are required to ensure a robust statistical analysis for risk assessment. To overcome this limitation, researchers have developed methods to generate synthetic tropical cyclones (STCs) that provide a larger sample size of occurrences at specific locations. This study compares STC databases from different sources such as Massachusetts Institute of Technology (MIT), Columbia HAZard model (CHAZ), Synthetic Tropical cyclOne geneRation Model (STORM), and Deltares with historical TCs from the International Best Track Archive for Climate Stewardship (IBTrACS) on a basin-wide scale in the North Atlantic Basin. The aim is to assess the effectiveness of STCs in replicating crucial historical tropical cyclones parameters for risk analysis and to identify potential biases in the STC generation models. The comparison uses a hexagonal mesh to evaluate characteristics such as maximum winds, translation speed, and residence time. The study acknowledges the validation paradox arising from the limited IBTrACS data at specific locations that make it difficult to rigorously validate the accuracy of STCs in those areas and from systematic differences across the STC datasets. Despite the historical TCs database limitation, comparing STC with IBTrACS characteristics remains the only viable method for assessing biases in STC generation models. The evaluated STCs reveal spatial bias patterns, which may indicate deficiencies in the underlying hazard models. Identifying and describing these biases aim to guide the use of these events and highlight key aspects for further development in STC generation methods.

1. Introduction

Tropical cyclones (TCs) are hydrometeorological events characterized by their associated hazards, including powerful winds, coastal and inland flooding, extreme rainfall and large waves. The combined impact of these hazards results in one of the highest insured losses of all natural hazards, second only to earthquakes and tsunamis in terms of fatalities. Given their significant risks, TCs have been extensively studied across various research fields, from forecasting to long-term trend analysis and climatological assessments. However, reliable TC records have only been available since the satellite era (Vecchi and Knutson, 2011; Landsea and Franklin, 2013). Consequently, the existing data are scarce and lack sufficient TC information to produce reliable statistics for risk assessments. To overcome this data scarcity, researchers have developed

* Corresponding author at: Laboratorio de Ingeniería y Procesos Costeros, Instituto de Ingeniería, Universidad Nacional Autónoma de México, Puerto de abrigo s/n, Sisal, Yucatán 97356, Mexico.
E-mail address: cappendinia@iingen.unam.mx (C.M. Appendini).

https://doi.org/10.1016/j.atmosres.2025.108404
Received 23 January 2025; Received in revised form 14 June 2025; Accepted 3 August 2025
Available online 6 August 2025

alternative methods for creating robust statistics by generating synthetic events that can provide a larger sample size of occurrences at specific locations. Synthetic tropical cyclone datasets offer a valuable alternative to historical records, enabling the exploration of low-frequency, high-impact events and future climate scenarios. This study aims to assess the reliability of commonly used STC products by comparing them to observations, thereby addressing the growing need for robust data in climate risk modeling and adaptation planning.

Various methods can be employed to generate synthetic events, including the utilization of sea surface temperatures (Vickery et al., 2000; Hall and Jewson, 2007; Yonekura and Hall, 2011), statistical approaches (Bloemendaal et al., 2020b; Nederhoff et al., 2021), and hybrid statistical-deterministic techniques (Emanuel, 2006; Lee et al., 2018; Xi and Lin, 2022). Synthetic tropical cyclones (STCs) have been utilized to evaluate TCs on global, basin-wide, and location-specific scales, encompassing assessments of winds (e.g., Nederhoff et al., 2021), storm surges (e.g., Ruiz-Salcines et al., 2021), ocean waves (e.g., Appendini et al., 2017; Ruiz-Salcines et al., 2019) and TC costs (e.g., Meiler et al., 2022). The primary advantage of synthetic models is their ability to provide a large number of STCs, which enables robust statistical analysis. Particular advantages of statistical-dynamical downscaling models are their climate dependence and their capability to incorporate the effects of climate change.

Although using databases with numerous STCs offers benefits, the validation process often involves the removal of detailed geographic information when performing basin-wide analyses, as reported by Emanuel et al. (2006, 2008), Lee et al. (2018), and Bloemendaal et al. (2020a). On one hand, this method is deemed suitable because synthetic events are designed to overcome the constraints of short historical records, and validation of these datasets can only be accomplished by comparing them to another extensive one, specifically basin-wide historical occurrences. On the other hand, validation at a local scale can be valuable, but it brings out a validation paradox: while synthetic models are designed to overcome the data scarcity imposed by the historical tropical cyclones data by providing a more extensive set of events for statistical analyses, the scarcity of historical tropical cyclones data in specific locations makes it difficult, if not impossible, to rigorously validate the accuracy of the synthetic data in those areas. This creates a situation in which the STCs might offer a better statistical representation of potential TC activity; however, there are insufficient historical tropical cyclones data to confirm whether the synthetic models accurately reflect the observed activity. Consequently, location-specific biases may remain undetected, leading to potential inaccuracies in local risk assessments based on the STC data. Therefore, although STCs can be a valuable tool, the application of such methods at the local scale should be approached with caution, and further efforts should be made to enhance validation methods whenever possible (Strazzo et al., 2013).

In this study, we acknowledge the validation paradox but proceed from the premise that comparing STCs with historical tropical cyclones is the only viable method to assess biases in STC generation models (Lee et al., 2022). Therefore, we analyzed STC databases from different academic and applied research sources on a basin-wide scale to determine their effectiveness in replicating the historical tropical cyclones parameters crucial for risk analysis, particularly their ability to reproduce local data. Our goal is to identify and describe the biases in the synthetic datasets as derived from the different models and thus to guide their use

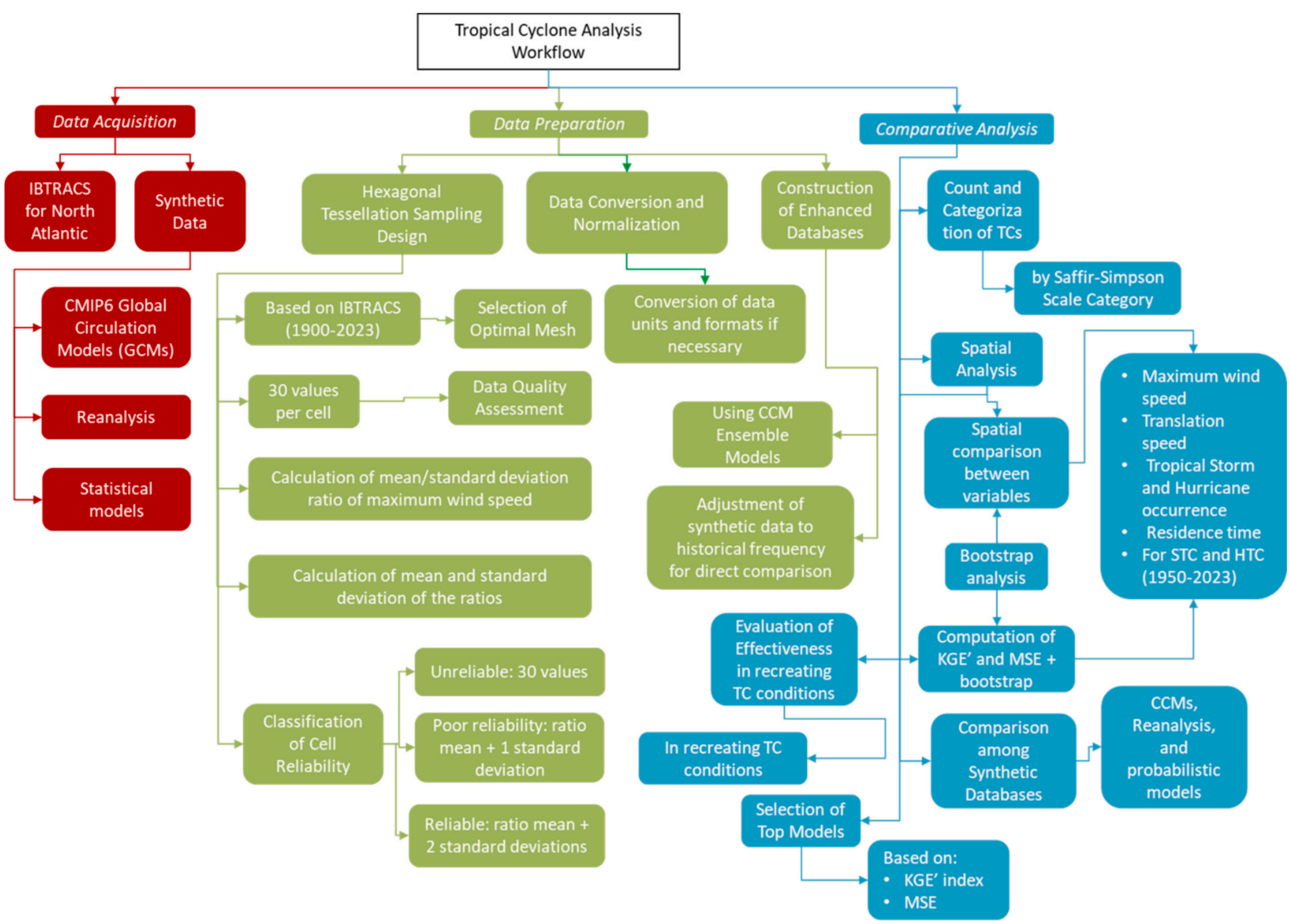

Fig. 1. Methodological flux diagram.

and highlight key aspects for further development. Section 2 describes the analyzed databases and evaluation method, followed by the results, a discussion of the relevant biases in Section 3, and the conclusions.

2. Methods and data

The methodology used in this study is summarized in Fig. 1 and described in the following subsections. This method compares historical tropical cyclones with STC in a hexagonal mesh to assess characteristics, such as maximum winds, translation speed, and residence time. The residence time is calculated as the time, in hours, that a TC trajectory remains in a spatial area. The following sections describe the datasets and analyses used in this study:

2.1. Tropical cyclone data

2.1.1. Historical tropical cyclone data

The historical tropical cyclones data used in this study were obtained from the International Best Track Archive for Climate Stewardship (IBTrACS), managed by the National Centers for Environmental Information (NCEI) of the United States (Knapp et al., 2010; Gathan et al., 2024). The database contains the following variables for each oceanic basin: hurricane name, year, month, day, hour, 1-min 10-m average sustained wind speed in knots, latitude and longitude of the location of minimum pressure (the eye of the TC), and central atmospheric pressure in mb or hPa. Estimates on storm parameters are reported every three hours, with the NHC as the source for the North Atlantic basin.

Geographical positions have a spatial resolution of 0.1° or greater, and the maximum sustained wind speed is given in knots, generally multiples of five. While the available Best Track data extend back to 1851 for the North Atlantic Basin, it is important to note that the observation systems have improved dramatically as new observing platforms came online and observing technologies have improved. Key developments affecting the Best Track data quality include the advent of routine aircraft reconnaissance in 1946, the beginning of the satellites in the mid-1960s (Neumann and Elms, 1993; Landsea and Franklin, 2013), and the GPS dropsonde (Hock and Franklin, 1999). Generally, the Best Track data have been more reliable since the 1990s when the observational platforms became more consistent (Camargo et al., 2023). The information recorded in the database is the result of the compilation and integration of observational data from various sources, including weather satellites, aerial reconnaissance, buoys, and weather stations. This study used a 3-h segment database with the National Hurricane Center information, focusing on the North Atlantic Basin. We retained all the 1900 to 2023 TCs for the sampling size determination and tropical storms or hurricanes in the study area from 1950 to 2023 (942 events) for the STC comparison.

To make the databases comparable, we excluded the track segments north of 64°N and east of 2°E and those events crossing into the Pacific basin. We retained all STC events that reached at least tropical storm force intensity (34 knots) in the basin.

2.1.2. Synthetic tropical cyclone (STC) databases

The STC databases used in this study include those derived by statistical methods like the Synthetic Tropical cyclOne geneRation Model (STORM) and Deltares, deterministic methods using reanalysis for the Massachusetts Institute of Technology (MIT) and Columbia HAZard model (CHAZ), and Coupled Model Intercomparison Project (CMIP) models (MIT). Following is a description of each STC database used in this study.

2.1.2.1. MIT STC database. Most STC databases used in this study are from the Lorenz Center at the Massachusetts Institute of Technology (referred to hereafter as MIT events). These STCs were generated using a downscaling deterministic technique to create synthetic TCs by applying the Coupled Hurricane Intensity Prediction System (CHIPS) model to events initiated through random seeding in space and time and propagated forward using a beta-and-advection (Emanuel, 2006, 2008; Emanuel et al., 2006). They were obtained for the North Atlantic Basin, with 3050 tracks downscaled from 8 global circulation models (Zhu et al., 2021; Rivera, 2023) of the sixth phase of the Coupled Model Intercomparison Project (CMIP6) (Tebaldi et al., 2021). The selected reanalyses and GCMs offer high spatial resolution, global coverage, and are widely used in recent TC climatology studies (e.g., Xi et al., 2025; Camargo et al., 2020, 2025).

The models used were CESM2 (Danabasoglu et al., 2020), CNRM-CM6–1 (Voldoire et al., 2019), EC-Earth3 (Döscher et al., 2021), HadGEM3-GC31-LL (Sellar et al., 2020), IPSL-CM6A-LR (Boucher et al., 2020), MIROC6 (Tatebe et al., 2019), MPI-ESM1–2-HR (Müller et al., 2018), and UKESM1–0-LL (Tang et al., 2019) (hereafter referred to as MIT-CESM, MIT-CNRM, MIT-ECEARTH, MIT-HADGEM, MIT-IPSL, MIT-MIROC, MIT-MPI, and MIT-UKESM, respectively). These events were generated for the period 1950–2014. This time interval corresponds to the baseline “present climate”, which serves as a reference for future work related to climate change analysis using CMIP6.

The objective in downscaling GCMs is to estimate hurricane activity consistent with the GCM’s climate conditions, not with observed activity. Actually, the aim of the GCM is not to describe the present climate but to improve the scientific knowledge of how variations in greenhouse gases, aerosols, and incoming solar radiation affect the planet’s radiation balance and climatic response (Myhre et al., 2022). Indeed, present-climate risk assessments should use STCs derived from reanalysis rather than GCMs in the current climate. Using reanalysis derived events provides more accurate assessments because reanalyses better capture the mean environmental conditions that influence storm genesis and evolution than uncorrected GCM output. While GCM-derived STCs fail to reproduce the actual climate variability due to model biases (Sobel et al., 2023), they are useful for detecting climate change effects on TCs as the trend of GCM simulations captures the climate forcing signal. On the contrary, the objective of deriving STCs from reanalysis is to obtain events that are broadly consistent with historically observed activity (Emanuel, 2021a). We included CERA20C, ERA5, MERRA2, and NCEP reanalysis-derived MIT datasets, which cover more recent periods and contain significantly more events, each providing approximately 20,000 STCs.

Storm frequency is governed by an externally prescribed seeding rate so that the mean annual storm count in the basin matches the observed IBTrACS climatology; this required calibration is intrinsic to the MIT method rather than a post-hoc bias correction. Beyond that step, no additional bias correction is applied, neither to frequency nor to wind-intensity. Although bias-correction methods such as frequency matching to reanalysis climatology (Gori et al., 2022) or parameter-level adjustments for wind speed, wave height, or storm tide (Appendini et al., 2025; Xi and Lin, 2022; Qiu et al., 2025), are common in the literature, here we analyzed the unaltered outputs to enable a direct comparison of model formulations.

2.1.2.2. CHAZ database. The Columbia HAZard model (CHAZ) database contains 686,396 STCs from 1981 to 2019 (Lee et al., 2018, 2022). In the CHAZ model, STCs are initially seeded in time and space based on the Tropical Cyclone Genesis Index (Tippett et al., 2011). These precursors are then steered by the background flow using a beta-and-advection similar to that of Emanuel (2008). Along each track, a stochastic autoregressive intensity model is applied to simulate the intensity evolution of the STCs. Only those STCs whose lifetime maximum intensity reaches at least 34 knots are recorded. Due to the design of the stochastic intensity model, CHAZ consists of 40 members, each with 25,170 tracks; while the members share trajectories, they have different wind speeds. The CHAZ model generates STCs from ERA5 reanalysis data, offering a robust sample size for statistical analysis and enhancing the depth of comparative insights.

2.1.2.3. STORM database. The Synthetic Tropical cyclOne geneRation Model (STORM; Bloemendaal et al., 2020b, 2022) database includes 109,026 STCs based on observed IBTrACS statistics from the IBTrACS dataset (between 1980 and 2017). STORM is an open-source, global-scale, full statistical model, which takes information on TC track, characteristics (intensity, radius of maximum winds, and genesis month) from IBTrACS, combined with environmental variables (monthly averaged mean sea-level pressure and sea-surface temperature) from ERA5. After assigning a genesis month, location, and first-step changes in track and intensity, the TC track and intensity changes are modeled following a series of autoregressive formulas. This way, STORM creates a new, synthetic database of 10,000 years of STCs based on the statistics found in the 38 years of input IBTrACS data. Validation of basin-wide STORM STC statistics shows that they closely resemble those found in the IBTrACS dataset (Bloemendaal et al., 2020a). In contrast to the other datasets, the wind speeds in the STORM dataset are based on 10-min sustained winds ($V10min$). Therefore, the data were converted accordingly (Harper et al., 2010) (1) as part of the preprocessing.

$$V_1min = V_{10}min/0.88 \quad (1)$$

2.1.2.4. Deltares TCWiSE database. The Deltares TCWiSE database is built using the Tropical Cyclone Wind and Impact Simulation Environment (TCWiSE; Nederhoff et al., 2021) tool, developed by Deltares. This database for the North Atlantic includes 40,000 STC tracks generated using a Monte Carlo-based statistical approach grounded in historical cyclone data from IBTrACS (1886–2019). The tool employs four stages: track initiation, evolution, wind field construction, and extreme wind speed determination. STC tracks are generated using a Markov model to simulate 3-hly changes in track location, forward speed, intensity, and heading, incorporating kernel density estimates (KDEs) derived from historical data. Unlike other databases, TCWiSE stands out for its flexibility as a tool. Users can calibrate and validate cyclone generation processes for specific regions and easily adapt the tool to include additional features. Unlike other databases, TCWiSE is a tool that can also generate detailed wind fields, making it ideal for risk analysis and understanding climate variability impacts.

2.2. Analysis

To evaluate synthetic tropical cyclone (STC) databases on a common spatial grid, we first selected IBTrACS, a high-quality historical database covering the same spatial domain. We developed our sampling design by analyzing IBTrACS data from 1900 to 2023, using hexagonal tessellations which are particularly well-suited for spatial analysis of tropical cyclone data (Elsner et al., 2012). Indeed, a hexagonal mesh is more adequate than a traditional square grid because hexagons minimize edge effects and spatial bias due to their uniform adjacency and isotropic geometry (Elsner et al., 2012). We tested various hexagonal mesh apertures ranging from 0.5° to 6° to identify the most appropriate spatial scale, extracting the maximum sustained wind speed from both STC and IBTrACS events for each grid cell. Through this analysis, we determined that 5° represented the optimal sampling size, providing adequate data coverage while maintaining sufficient spatial resolution.

Following established statistical practices for extreme value analysis (Coles et al., 2001), we implemented a minimum threshold of 30 m/s wind speed values per cell, with each tropical cyclone contributing only one value per cell to avoid spatial autocorrelation. We classified cells based on their data reliability, categorizing as “unreliable” those cells with less than 30 m/s wind speed values or located entirely north of 40° latitude due to limitations in STC downscaling methods at higher latitudes (Bloemendaal et al., 2020a). All other cells were considered candidate-reliable and were further classified based on the coefficient of variation (CV) of maximum wind speeds, calculated as the standard deviation divided by mean. The CV basically represents the spread relative to the average value, where larger values in our case will typically indicate greater wind speed variability and may suggest either genuine climatic variability or data quality issues (Wilks, 2019). For the cells meeting the criteria to be candidate-reliable, we calculated the CV in each cell j as follows:

$$CV_j = \frac{\sigma_j}{\mu_j} \quad (2)$$

where σ_j and μ_j are, respectively, the mean and standard deviation of the maximum sustained-wind speeds recorded in that cell. Treating the set $\{CV_j\}$ across all candidate-reliable cells as a sample, we calculated its mean $\underline{CV}$ and standard deviation S_{cv}. Cells were then categorised as follows:

- Reliable if $CV_j \leq \underline{CV} + 0.5S_{cv}$
- Rather reliable if $\underline{CV} + 0.5S_{cv} < CV_j \leq \underline{CV} + S_{cv}$
- Less reliable if $CV_j > \underline{CV} + S_{cv}$

This two-stage approach ensured that our spatial analysis emphasizes grid cells with both adequate sample size and typical within-cell variability, while flagging areas where unusually large dispersion may reflect either genuine climatic heterogeneity or data-quality limitations, thereby providing a robust foundation for comparing historical and synthetic tropical cyclone databases.

TC events in each cell were created in the R programming language (R Core Team, 2022) using the sf R-package (Pebesma, 2018). Subsequently, the numbers of IBTrACS TCs and STCs in each category were counted using the thresholds from the Saffir-Simpson Hurricane Wind Scale. In the next step, sf was used with *dplyr* R-package (Wickham et al., 2023) for sampling TCs per hexagon. The values of the number of tropical storms and hurricanes per cell were adjusted using the relationship between the average annual total number of STC and the actual IBTrACS year frequency (12.73) for 1950–2023. The mean translation speed and the maximum sustained wind speed were extracted for each cell. The translation speed was calculated by considering a straight line between the TC position records (Kossin, 2018), which might have underestimated the actual values. TC datasets' time step varies: IBTrACS provides data at 3-hour intervals, whereas STC datasets typically use 1–2-h intervals. While this could theoretically introduce differences in derived quantities (e.g., translation speed or residence time), we chose not to interpolate STC trajectories to match the 3-h resolution of IBTrACS. Interpolation could introduce artificial values or smooth key features of the trajectory, and given the size of our spatial grid (5° aperture), we expect any discrepancy due to time-step differences to have a negligible effect on the aggregated metrics. In addition, to avoid redundancy in counting tropical cyclone (TC) entries into grid cells, we applied a consistent rule: each TC was counted once per cell, regardless of how many times it re-enters that cell. This ensures no double-counting of storm occurrences. However, for metrics such as maximum wind speed, translation speed, and residence time, all valid trajectory segments are considered in every cell they pass through. Thus, a storm contributes to the statistics of each cell it traverses, as appropriate, but without duplication of storm counts.

A comparison between each synthetic TC database and recorded data was performed on two scales. First, we recorded the number of events for the basin in the study area as determined by the minimum sustained wind thresholds for each category on the Saffir-Simpson scale. The percentages of systems in each category were calculated considering the lifetime maximum intensity (LMI) in the basin.

Subsequently, a spatial analysis was conducted. The effectiveness of each STC generation method in recreating tropical cyclone conditions was spatially compared by computing the Mean Standard Error (MSE) between the values corresponding to the IBTrACS and those representing each model at the level of each hexagonal cell (3).

$$MSE = \sqrt{\frac{\sum_{n=1}^{n}(Vo_n - Vs_n)^2}{n}} \quad (3)$$

where Vo_n is a value from the observed data at hexagon n, and Vs_n is the value from the simulated data.

We also calculated the modified Kling-Gupta Efficiency (KGE') index (4) (Gupta et al., 2009; Kling et al., 2012). The KGE' metric, originally developed for hydrological model validation, has also proven effective in climatology (Centella-Artola et al., 2020) and in flood-impact studies. For instance, Konduru et al. (2023) used KGE' to assess the accuracy of hydrological simulations and rainfall forecasts during the 2015 Chennai flood, demonstrating the metric's utility in evaluating extreme weather-related processes.

$$KGE' = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (4)$$

where:

r is the Pearson correlation coefficient between IBTrACS and STCs.

α is a term representing the differences of the standardized variability between two datasets (5),

$$\alpha = \frac{\sigma_s}{\mu_s} \times \frac{\mu_o}{\sigma_o} \quad (5)$$

μ_s is the mean of the simulated values.

μ_o is the mean of the observed values.

σ_s is the standard deviation of the simulated values.

σ_o is the standard deviation of the observed values.

β is a ratio of the STC and IBTrACS mean values (6)

$$\beta = \frac{\mu_s}{\mu_o} \quad (6)$$

In other words, KGE' comprehensively considers the discrepancies in the mean, variability, and spatial pattern between STC and IBTrACS. KGE' was applied to the maximum wind speed, translation speed, and frequency and residence time of tropical storms and hurricanes, which are important variables for characterizing TC hazards (Kantha, 2006; Sobel et al., 2021). A low MSE and a KGE' index closer to 1 identify a better fit of the model to the measured data.

We also calculated the hexagonal mean values from the GCM-derived STC and the ensemble of those databases. The use of ensembles in climate models allows for the provision of a measure of uncertainty and compensates for model biases for improved fitting to measured values (Camargo and Wing, 2016). This database was subsequently evaluated using the MSE and KGE' index calculations. For the CHAZ data we computed the KGE' and MSE of each member and ensemble, averaging the statistics and the spatial values for mapping. In order to obtain the MSE and KGE' 95 % confidence intervals, we applied 1000-iterations bootstraps with replacement (Efron, 1979; Efron and Tibshirani, 1994), these bootstraps were done with aleatory sampling of the hexagonal cells metrics.

Moreover, to ensure comparability across datasets with different track counts and evaluate the sensitivity of the analysis to the larger number of STC tracks compared to IBTrACS, we performed a bootstrap analysis with 1000 resamples of size N, where $N = 942$ corresponds to the number of tracks in IBTrACS (dataset with the minimum number of tracks). The selections of the sampled TCs were done without replacement, and for each iteration, we computed (1) percentage of tropical storms and hurricanes categories, (2) tropical cyclones spatially aggregated metrics, (3) Kling–Gupta Efficiency (KGE'), and (4) mean squared error (MSE) between synthetic and historical tracks. The median and 95 % confidence intervals of these metrics were then derived over 1000 iterations, providing robust, sample-size–controlled performance estimates.

Finally, to detect the spatial bias linked to STC generation methods, we created maps to elucidate the spatial variation of differences between IBTrACS and STC for the dataset with a higher fit to actual data.

Taking advantage of the fact that this study includes two datasets of STCs based on ERA5, generated through different methodologies, we constructed seeding density maps for both CHAZ-ERA5 and MIT-ERA5 and computed spatial statistical significance using a Poisson test, applied to 5° hexagonal mesh cells and considering the 5th–95th quantiles of the historical genesis count (Supplementary Fig. S1). This complementary analysis may help elucidate differences in the outcomes of the evaluations.

3. Results and discussion

The section compares observational tropical cyclone data and synthetic datasets derived from statistical models, reanalysis, and GCMs. It evaluates the performance of these datasets across various metrics, focusing on maximum sustained wind speeds, storm intensity distributions, and area-integrated metrics like MSE and KGE'. It also discusses regional differences in model performance and highlights the complexities of tropical cyclone modeling.

3.1. Comparison of maximum sustained wind speeds and intensity distribution

As a first approach, we compared observational data from IBTrACS with synthetic datasets from statistical models (Deltares and STORM), synthetic datasets from statistical-dynamical models based on reanalysis data (CHAZ downscaled from ERA5, and the MIT runs downscaled from NCEP, MERRA2, ERA5, and CERA20C), and MIT synthetic datasets downscaled from GCMs. The information spans from 1950 to 2023, with individual databases covering various time frames within this range (Fig. 2, Table S1). We decided to use this timeframe based on a sensitivity analysis of assessing each STC database period of IBTrACS to the selected period, finding that the KGE' differences were not significant (not shown). Maintaining the selected period for all databases ensures all STC are compared to the same database.

The historical IBTrACS data provides the baseline for comparing tropical cyclone intensities, showing a maximum wind speed of 84.88 $m \cdot s^{-1}$, while synthetic databases range from 77.61 to 117.38 $m \cdot s^{-1}$. According to IBTrACS, approximately half of all storms maintain tropical storm intensity, with decreasing frequencies for higher categories. This observed pattern serves as a reference point for evaluating synthetic databases.

The statistically derived models STORM and Deltares demonstrate different approaches to representing intensity distributions. Since they are derived directly from IBTrACS data, their maximum values remain close to the IBTrACS observations, with Deltares providing results that are more consistent with IBTrACS. The differences between these datasets highlight how different resampling methodologies and assumptions can produce varying statistics even when sampling from the same historical data. The potential impact of different resampling methodologies results in synthetic TCs with different statistics despite being sampled from the same IBTrACS.

Reanalysis-derived products show distinctive characteristics in their representation of tropical cyclones. These datasets generally maintain more realistic proportions of intense hurricanes than GCM-derived products, though they produce higher maximum intensities than historical data. The reanalysis products show relatively consistent maximum intensities, clustering around 105–117 ms^{-1}. STCs databases from reanalysis stand out with the highest maximum wind speed, significantly higher than the historical observations from IBTrACS and other STCs. MIT-ERA and CHAZ-ERA demonstrate the highest values, 117.38 and 116.76 $m \cdot s^{-1}$, respectively, with a 95 % confidence interval of 3.9. Most synthetic databases show lower percentages of tropical storms and higher percentages of intense hurricanes. However, CHAZ data shows distributions more similar to IBTrACS, with approximately 50 % of tropical storms aligning more closely with the historical

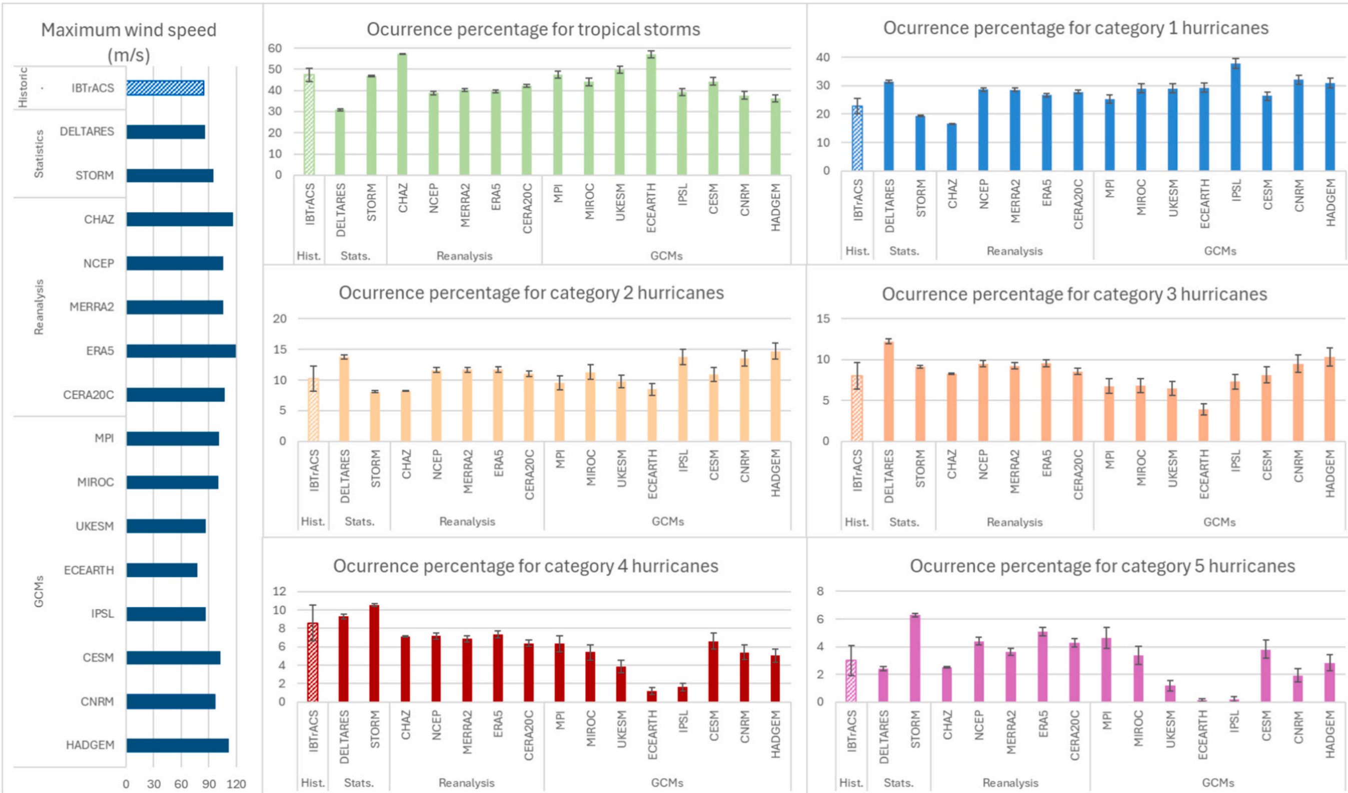

Fig. 2. Maximum sustained wind speed and proportion of TCs per Saffir-Simpson scale category across the North Atlantic basin, showing the 95 % confidence interval based on a bootstrap analysis of 1000 resamples.

observations. CHAZ, downscaled from ERA5, exhibits unique characteristics with very high maximum intensity due to the stochastic design of the intensity component of the model. This design allows it to get the highest possible intensity of each track at each time, increasing the likelihood of obtaining particularly high wind speed values. The model, however, shows lower proportions of intense hurricanes. Regarding MIT-reanalysis, the great number of events per database relative to GCMs also increases the likelihood of obtaining high wind speed values. The differences may be related to observational sample errors, imperfect reanalysis data, and imperfect physical assumptions in the downscaling models.

GCM-derived results present the widest variation in tropical cyclone representation. GCMs show wide variation in their representation of TC intensities; when downscaled from some GCMs (like MIT-ECEARTH and MIT-IPSL), MIT models struggle to produce very intense hurricanes and generally overestimate the proportion of tropical storms while underestimating intense hurricanes. Among the GCM-derived MIT STCs (MIT-CESM, MIT-CNRM, MIT-ECEARTH, MIT-HADGEM, MIT-IPSL, MIT-MIROC, MIT-MPI, and MIT-UKESM), there is substantial variation in maximum intensities, ranging from 77.61 $m{\cdot}s^{-1}$ (MIT-ECEARTH) to 111.67 $m{\cdot}s^{-1}$ (HADGEM). Most GCM-derived events demonstrated significantly lower percentages of tropical-storm-force occurrences, with MIT-ECEARTH and MIT-UKESM being exceptions. Additionally, MIT-UKESM, MIT-ECEARTH, MIT-IPSL events exhibited substantially lower percentages of Category 5 events, while MIT-CESM and MIT-MPI show an overestimation. The over- and underestimation in GCM-derived MIT STCs are related to GCMs biases (Camargo et al., 2025), as they all use the MIT model that can generate STCs with intensity distributions similar to IBTrACS. For instance, some GCMs over(under)estimate the wind shear in the Main Development Region (Camargo, 2013).

The variation in GCMs maximum wind speeds suggests differences in the TC-relevant large-scale conditions in the GCMs used for downscaling in the MIT model and underscores the uncertainty of modeling tropical cyclone intensities in GCMs. The discrepancies among historical records and synthetic databases suggest that different methodologies and data sources can lead to significantly different representations of extreme storm events. STC events from reanalysis and some GCMs frequently report extremely high wind speeds, which is why 99 %-ile is used for risk studies (Appendini et al., 2017). GCMs have a large range in maximum wind speed as they are free-running models (Zarzycki, 2022), meaning that these models are not constrained by external data to guide the evolution of the TC-relevant large-scale fields as they do in reanalysis. Thus, they are free to run based on their own internal physics. The differences in the basin-wide maximum intensity among the four models are consistent with Meiler et al. (2022), who state that the probability of the costliest events is sensitive to the model choice. These findings emphasize the importance of using multiple data sources and approaches when studying tropical cyclone characteristics. However, multiple-model approaches will not address uncertainties associated with model biases, such as the cold tongue biases in the tropical Pacific (e.g., Seager et al., 2019, 2022; Sobel et al., 2023) and uncertainties due to scientific limitation, such as the lack of genesis theory (Lee et al., 2020, 2023; Sobel et al., 2021). The differences between CHAZ and MIT events derived from ERA5 show the differences between two TC downscaling models despite being downscaled from the same reanalysis data. Supplementary Fig. S1 shows genesis density for CHAZ-ERA5 and MIT-ERA5. CHAZ presents significant hotspots north of Panama and near Europe (outside the 5th–95th Poisson bounds of IBTrACS), whereas MIT's genesis is more evenly distributed, with only an excess north of Cuba–Hispaniola and in the Lesser Antilles. These differences stem from their seeding frameworks, where CHAZ draws seeds in proportion to the Tropical-Cyclone Genesis Index (TCGI) and accepts those that reach 34 kt within five days, whereas MIT seeds uniformly over ocean and retains only disturbances that exceed 40 kt. The Panama hotspot in CHAZ may also relate to how CHAZ handles TCs across/over small land areas. Thus, even with the same reanalysis forcing, deterministic downscaling schemes can produce markedly different storm populations, leaving extreme-event estimates strongly model-dependent. It is also relevant to notice the disparities with GCM-derived events is particularly relevant, underscoring the limitations of assessing potential changes in a warming climate.

The contrast between deterministic and statistical frameworks is crucial when evaluating basin-wide maxima. All MIT datasets are deterministic; because their intensities are generated dynamically, they are free to exceed the historical peak, especially when thousands of storms are simulated. While CHAZ is a statistical-dynamical downscaling model, the wind intensity is derived from an autoregressive model with a stochastic term. Conversely, STORM and Deltares are based on the historical record and therefore remain constrained by the historical record. It is noteworthy that two GCM-derived datasets (MIT-UKESM and MIT-IPSL) match the historical maximum, while the MIT-ECEARTH yields an even lower value. In the case of deterministic events underestimating the maximum wind speeds, particularly in large datasets, the downscaling technique is limited from the parent GCM fields, such as excessive vertical wind shear in the Main Development Region (Camargo, 2013), rather than from the downscaling procedure itself. Accordingly, a systematic upward shift in maximum winds within deterministic, reanalysis-based catalogues should not be viewed as poor adjustment but as the natural consequence of (i) larger effective sample size and (ii) freedom from the cap imposed by the short observational record, not as poor model performance.

3.2. Evaluation of model performance on area-integrated metrics

The maps in Fig. 3 and Fig. 4 present the spatial statistics from the IBTrACS data we used for the model assessment. Each hexagon

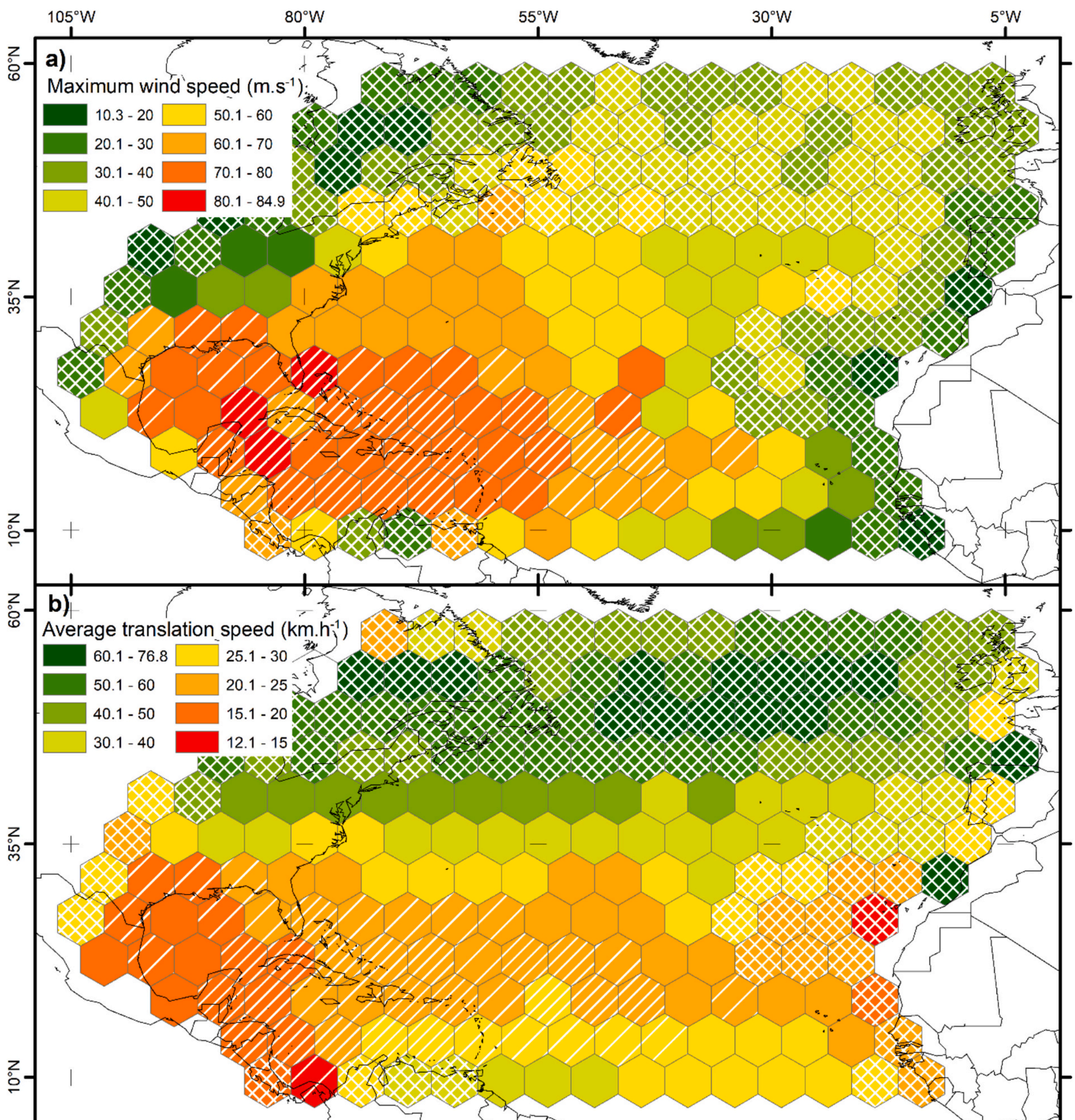

Fig. 3. Spatial historical values of a) maximum wind speed and b) average translation velocity. No hatch: reliable cell; single white hatch: rather reliable cell; double white hatch: low reliable cell; crossed white hatch: not reliable.

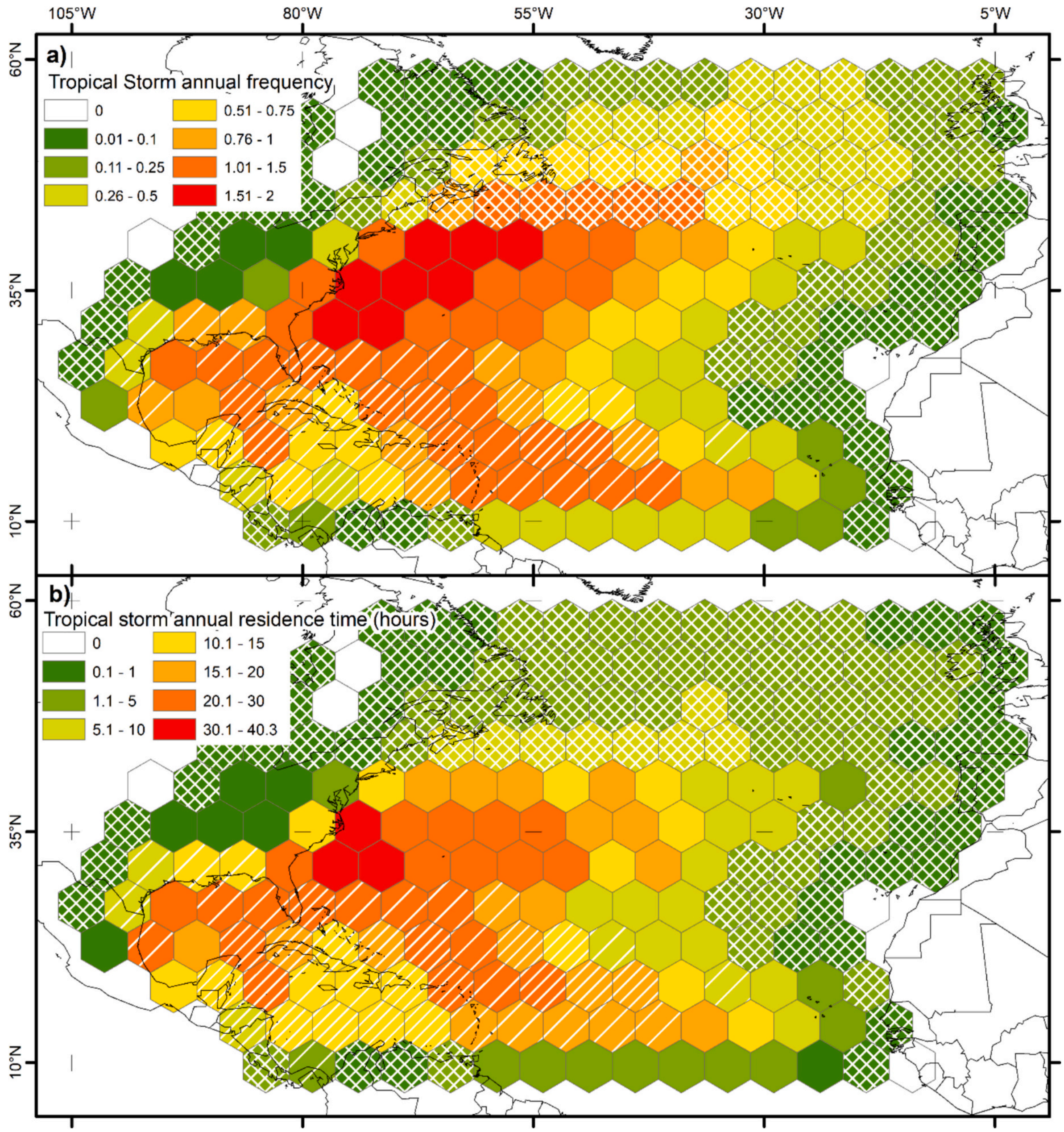

Fig. 4. Spatial historical values of a) tropical storm mean annual frequency and b) tropical storm mean annual residence time. No hatch: reliable cell; single white hatch: rather reliable cell; double white hatch: low reliable cell; crossed white hatch: not reliable.

represents a spatially aggregated metric (e.g., annual storm frequency) computed over a uniform grid. Please note that the figures we show are the most relevant for the discussion in the text, nonetheless, we are showing the maps for all variables and models in the supplementary information: https://observatoriocostero.lipc.unam.mx/supmat_maps.php. The data presented in these maps (Figs. 3 and 4) –maximum wind speed per cell, average translational speed, annual frequency of occurrence of storm wind or greater, residence time of trajectory with storm wind or greater– constitute the historical reference against which the STC databases were compared for the evaluation of their adjustment using MSE and KGE', or for the identification of regional biases (Section 3.3).

The MSE value ranges are specific for each variable (Fig. 5, Table S2), whereas KGE's values are more standardized (Fig. 5, Table S3), making them easier to compare for different models. The KGE' offers a more interpretable performance metric, allowing an overall estimation and comparison between variables. In contrast, MSE provides error values that are less challenging to interpret because they are expressed in the variable unit. The bootstrap analysis to derive the confidence intervals demonstrates that the spatial analysis and its metrics are robust. Specifically, the confidence intervals associated with good values of KGE' or MSE are relatively narrow.

3.2.1. MSE analysis

Fig 5 (Table S2) shows the performance of various models for representing the spatial patterns of tropical cyclone characteristics compared to IBTrACS data based on MSE. Please note that this metric is highly sensitive to outliers in the data.

In the MSE analysis, Deltares (7.23 $m \cdot s^{-1}$) and NCEP (3.78 km/h) performed best in terms of maximum wind speed and translation velocity, respectively. In contrast, the CHAZ (26.17 $m \cdot s^{-1}$, Confidence Interval CI = + − 0.31) and STORM (7.31 km/h) showed the largest

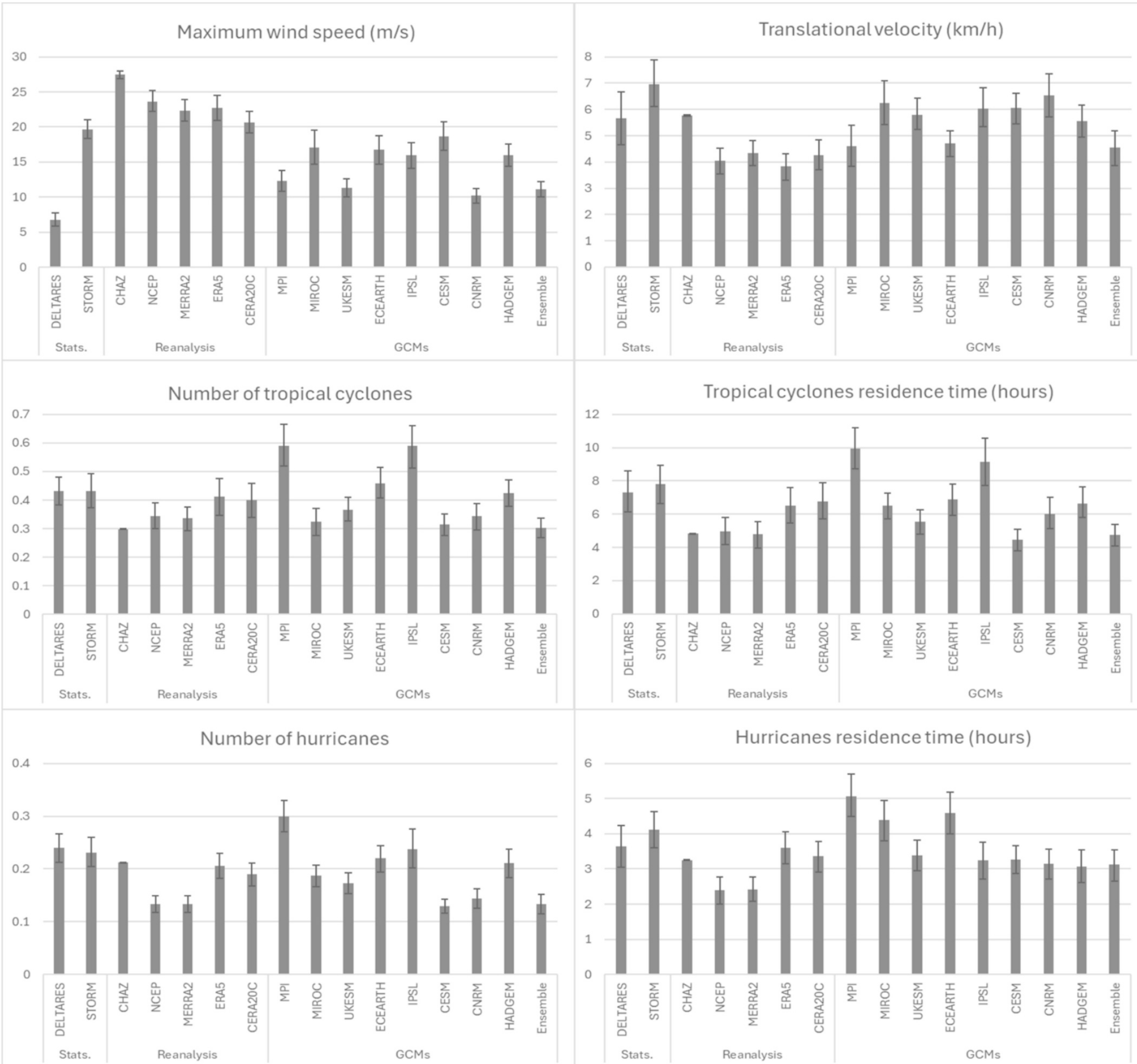

Fig. 5. MSE between the STC and IBTrACS analyzed variables, showing the 95 % confidence interval based on a bootstrap analysis of 1000 resamples.

deviations. The annual number of tropical storms (nTC) and hurricanes (nH) showed relatively low variation, with the best-performing databases, MIT-CESM (nTC = 0.3) and MERRA2 (nH = 0.12), which closely matched historical data, whereas MIT-IPSL (nTC = 0.65) and MIT-MPI (nH = 0.3) were notable outliers. Similarly, residence times for both tropical storms (resTC) and hurricanes (resH) showed some consistency, with MERRA2 (resTC = 4.72, resH = 2.57 h) having the lowest errors and MIT-IPSL (resTC = 10.48) and MIT-MPI (resH = 5.59 h) having the highest, indicating poor performance. MIT-MPI stands out as the most divergent from the observed data in hurricane frequency and residence time.

3.2.2. *KGE' analysis*

Fig. 6 (Table S3) shows the performance of various models for representing the spatial patterns of tropical cyclone characteristics compared to IBTrACS data based on KGE'. Note that a KGE' value of 1 means there is perfect agreement between simulations and observations and value of 0 indicates that the model's simulation performance is no better than simply using the mean of the observed data as a prediction.

The results from the STORM and Deltares statistical models exhibit notable differences in their performance. Deltares performed better in fitting the maximum wind speed, TC number, and residence time. This disparity highlights the necessity of carefully choosing models based on the specific TC parameters being studied. As observed by Camargo and Wing (2016), although statistical models can provide valuable insights into TC behavior, their effectiveness may vary considerably depending on the chosen parameters and analytical approaches.

STC downscaled databases from MIT reanalysis models, including CERA20c, ERA5, MERRA2, and NCEP, exhibited robust overall effectiveness in reproducing most variables associated with tropical cyclones. Their ability to capture the maximum wind speed and translation speed is particularly noteworthy, with KGE' values often exceeding 0.8, particularly for MERRA2. This is consistent with the findings of Thompson et al. (2024), who highlighted that reanalysis datasets have

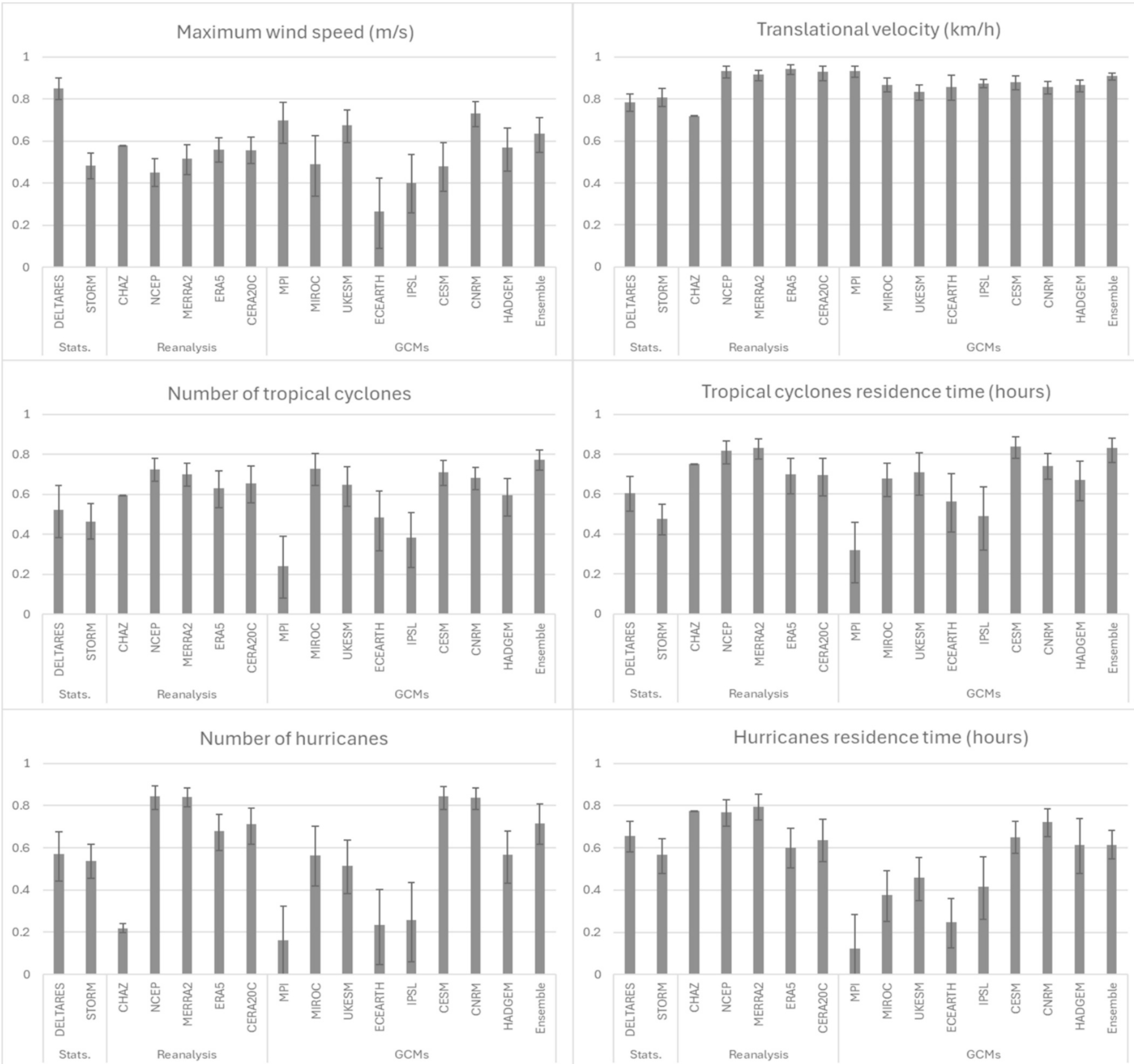

Fig. 6. KGE' index between the STC and IBTrACS analyzed variables, showing the 95 % confidence interval based on a bootstrap analysis of 1000 resamples.

become increasingly important for studying tropical cyclone climatology and variability. However, it is crucial to recognize that while synthetic events are generated from reanalysis data, they are distinct from the actual storms represented in the reanalysis that Thompson et al. (2024) refers to. They also noted that these datasets offer a better view of tropical cyclone activity than the observational datasets alone. This suggests that the precision of events derived from reanalysis is linked to how accurately the reanalysis captures the general environmental conditions conducive to tropical cyclone formation. However, CHAZ's maximum wind speed performance (KGE' = 0.25 and CI = + − 0.02) ensemble has a lower score compared to those derived from MIT-reanalysis STCs due to the presence of elevated maximum velocities in zones with habitually low categories hurricanes or tropical storms, adversely affecting the correlation term r and the variability α of the KGE. The maximum wind speed >100 m.s^{-1} shown in Fig. 1 also impacts the term and the MSE (Fig. 5).

Events derived from GCMs demonstrate diverse performance outcomes. They typically excel in depicting translation speed, but their accuracy in capturing maximum wind speed, TC, or hurricane number often falls short compared with events derived from reanalysis. This discrepancy underscores the ongoing challenges in modeling tropical cyclone intensity, as highlighted by Sugi et al. (2017) when identifying a latitudinal bias. Combining multiple GCMs in an ensemble approach (Scafetta, 2023) shows potential for addressing individual model shortcomings, particularly in terms of storm and hurricane frequencies and residence times. This method enables the use of GCMs in ensembles to evaluate future climatic scenarios; indeed, the GCM ensemble outperforms individual models for translation velocity, TC number, and residence time. Notably, GCM STC datasets, unlike reanalysis or statistical models, do not aim to describe the present but to establish a baseline for evaluating climate change.

The analysis revealed that no single model or model type consistently outperformed the others for all tropical cyclone variables. These findings confirm that the dynamics of tropical cyclones and their interactions with the broader environment (Lin et al., 2020; Meiler et al., 2022; Singh and Roxy, 2022) make it difficult to develop a single model that fully captures every aspect of these storms (Malakar et al., 2020), with significant differences in the results between models and methods (Knutson et al., 2020; Meiler et al., 2023).

The general high performance in translation speed for most models is encouraging, as it suggests a good understanding of the large-scale atmospheric patterns that govern tropical cyclone movements. However, the variability in performance for the maximum wind speed in reanalysis and GCM-derived STC indicates that further research is required in the downscaling techniques to represent cyclone intensity accurately. This is consistent with the authors who observed that despite notable advances in modeling tropical cyclones, accurately representing storm intensity continues to be a significant challenge for the present and concerning climate change projections (Roberts et al., 2020; Sobel et al., 2023). However, broad environmental conditions are utilized when deriving synthetic events from GCMs, so the findings reflect issues related to large-scale fields rather than the model's ability to generate tropical cyclones through its resolution and physics.

Comparing Fig. 5 and Fig. 6, the trend in model rankings remains similar, suggesting that models performing well under one metric tend to perform relatively well under another, albeit with different performance scales. Notable differences between metrics exist for MIT-ECEARTH and CHAZ, which have low MSE but poor KGE' for the maximum wind speed and TC residence time, respectively. Conversely, MERRA2 and MIT-MIROC have high MSE but good KGE' for the maximum wind speed and translational velocity, respectively.

Regarding the percentages of cyclones by category, the bootstrap analysis with replacement confirms a relative stability in the STC data, with narrow confidence intervals (Fig. 2). For the same analogous analysis using the bootstrap with a reduced number of events and without replacement we obtained similar results (Fig. S2). While the

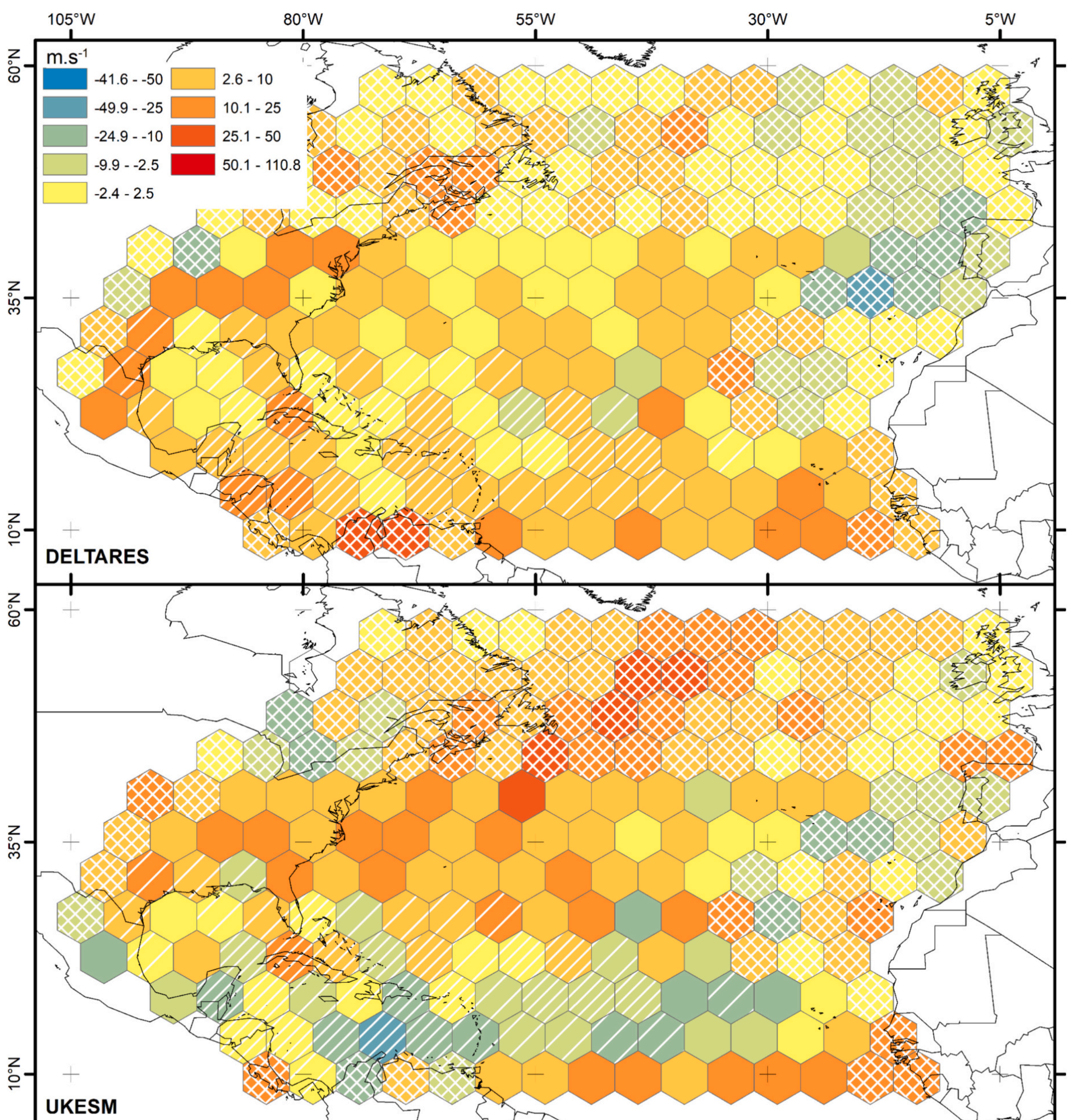

Fig. 7. Spatial differences between the best-scored STC dataset and IBTrACS for maximum wind speed. No hatch: reliable cell; single white hatch: rather reliable cell; double white hatch: low reliable cell; crossed white hatch: not reliable.

MSE and KGE' results from the bootstrap analysis with replacement have been already described, the analysis of MSE and KGE' using bootstrap without replacement (Figs. S3 and S4) reveal that model rankings shift when all datasets are compared using equal-sized samples (942), a relatively small number. Notably, MIT-GCMs exhibit high MSE values and very low, sometimes even negative, KGE' scores, indicative of poor model performance and limited agreement with observed data. The relative stability observed in reanalysis-based STCs compared to those derived from GCMs does not undermine the study's findings; rather, it highlights the importance of having a large number of STC events, particularly for GCM-derived data, which show greater variability due to systematic biases in GCMs. This result further supports the use of ensemble approaches to help address such variability.

3.3. Evaluation of model performance on regional scale

The maps with the spatial differences for the best evaluated STC datasets, based on MSE and KGE' results (Figs. 5 and 6) follow (Fig. 7 to Fig. 10). These spatial differences represent the subtraction between the metrics of the studied variables at each hexagon level for IBTrACS and the same metrics for each STC database. The spatial difference maps for all datasets are shown in the supplementary information found at https://observatoriocostero.lipc.unam.mx/supmat_maps.php.

3.3.1. Maximum sustained wind speed

While Fig. 3a shows the maximum sustained wind speeds for IBTrACS in each hexagonal cell; its differences with Deltares and MIT-UKESM GCM derived events are shown in Fig. 7, as these are the best evaluated models. Maps for the other datasets are presented as

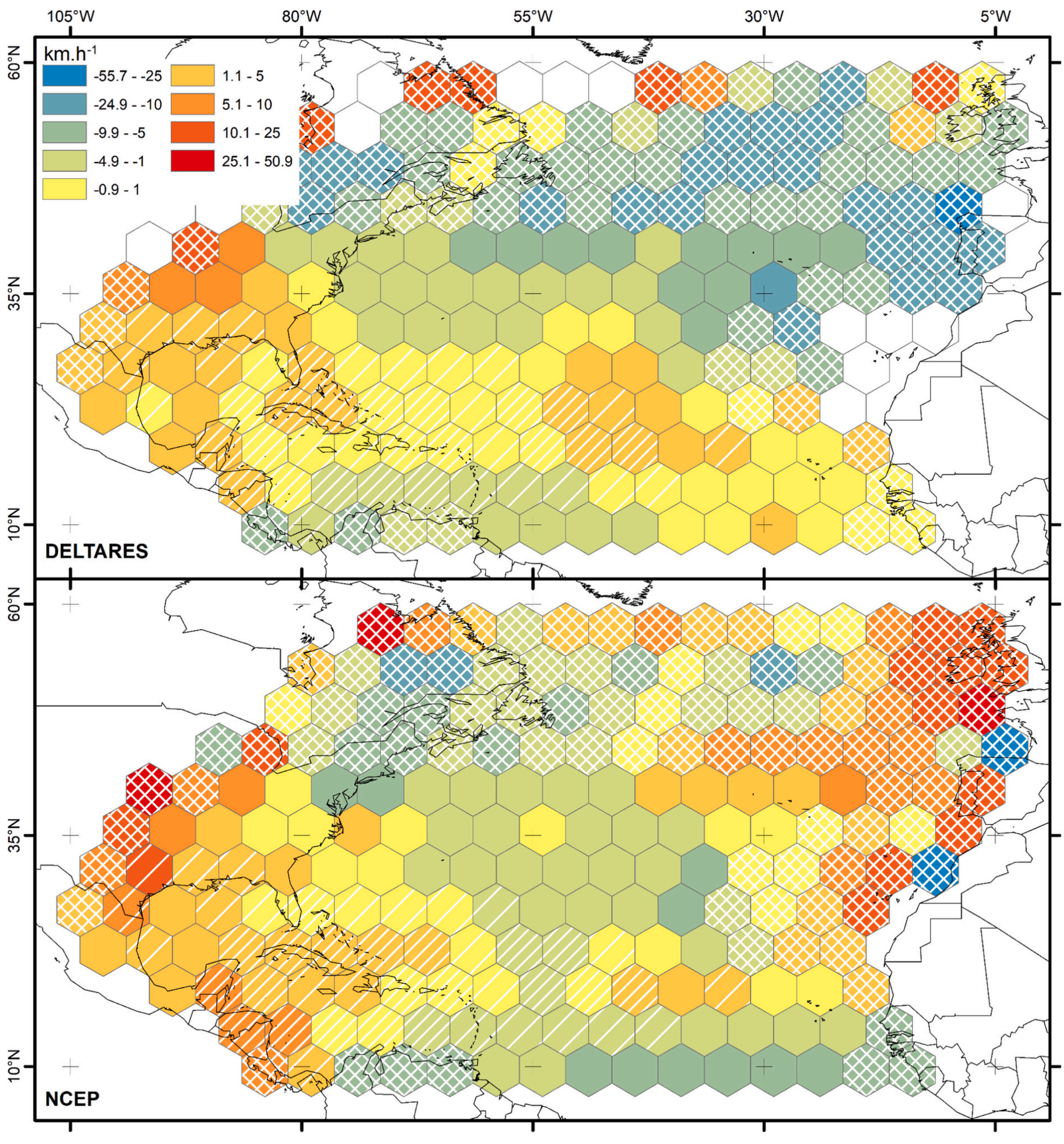

Fig. 8. Spatial differences between the best-scored STC dataset and IBTrACS data for translation speed. No hatch: reliable cell; single white hatch: rather reliable cell; double white hatch: low reliable cell; crossed white hatch: not reliable.

supplementary materials. Generally, wind speeds in Deltares events are slightly overestimated over most oceanic areas and have a high overestimation over land, up to $+25$ m·s^{-1} in hexagons with some reliability. For the MIT-UKESM events, there was a wider regional bias, with an area having underestimated velocities (<-10 m·s^{-1}) in the central tropical Atlantic and a larger with overestimations ($> +10$ m·s^{-1}) in the temperate zone. It should be noted that because of the higher number of events, the probability of obtaining higher velocities for STC than for IBTrACS increases; therefore, underestimations of this variable reveal difficulties in describing actual conditions. Nonetheless, the validation paradox indicates that a higher probability of stronger winds is accurate, but insufficient historical data exists to validate such information.

3.3.2. Translational velocities

The STC databases that best describe the spatial variation of the mean translational velocities are those derived from Deltares and NCEP reanalysis. Both datasets show similar regional biases (Fig. 8). Overestimations of translation velocity are present for the Caribbean, the Gulf of Mexico, and the continental USA. Both models show underestimations in the northeastern United States and eastern equatorial Atlantic. Regional differences can be found around the Azores, where Deltares underestimates the translation velocity, and NCEP overestimates relative to IBTrACS (Fig. 3b).

3.3.3. Tropical storm occurrences

The differences in tropical storms calibrated annual frequency (Fig. 9) relative to IBTrACS (Fig. 4a) also exhibited strong regional biases, with underestimations along the eastern coast of the United States, with values below -0.5 events/year for MERRA2. Minor underestimations were also observed for MIT-CESM. The main difference

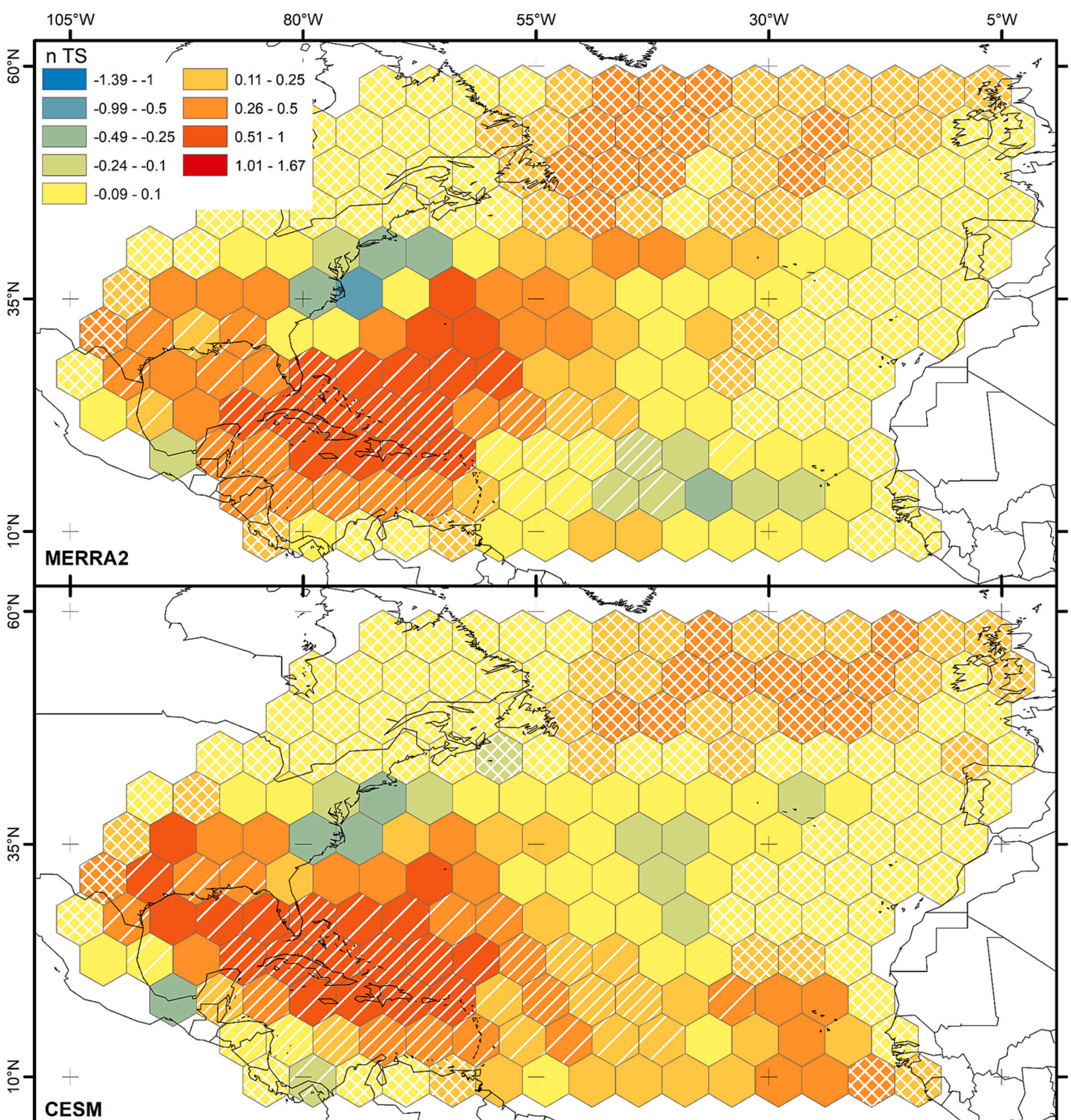

Fig. 9. Spatial differences between the best-scoring STC dataset and IBTrACS data for annual tropical storm occurrence. No hatch: reliable cell; single white hatch: rather reliable cell; double white hatch: low reliable cell; crossed white hatch: not reliable.

between the results from MERRA2 and MIT-CESM is that MERRA2 underestimates in the southeastern region of the North Atlantic and overestimates in the northern region. At the same time, this situation is reversed for MIT-CESM. For the reanalysis, the overestimation region was considerably broader, and generally located over the Gulf of Mexico, the Caribbean, and parts of the western and central Atlantic, with values exceeding +0.75 in the west Caribbean Sea for MERRA2 and in the zone offshore Texas for MIT-CESM. The presence of regional bias is consistent with the results of Strazzo et al. (2013). However, the regions differed because they used CMIP5 with GFDL and COAPS GCMs. Also, we can note that the regions with greater overestimations are reported for hexagonal cells with poor reliability. However, it is important to be cautious in considering this area as low reliability. In this sector, well-developed TC trajectories occur but there is also the formation of TCs. This contributes to a high natural variability in historical maximum wind speeds, which does not necessarily imply poor data quality. Actually, southern Florida and Cuba are regions where consistent and reliable reports have been documented.

3.3.4. Tropical storm residence time

STC based on reanalysis data exhibited superior performance in estimating annual tropical storm residence times compared with GCMs and statistically derived STC datasets, yielding the highest KGE or lowest MSE values. Also, in this case, the GCM ensemble outperformed every single model. Indeed, the spatial patterns of the differences are very similar among the presented datasets (MERRA2 and GCMs Ensemble) (Fig. 10), showing significant overestimations over the Antilles and Bermuda, as well as underestimations in the Central Atlantic and near the U.S. coast north of Florida.

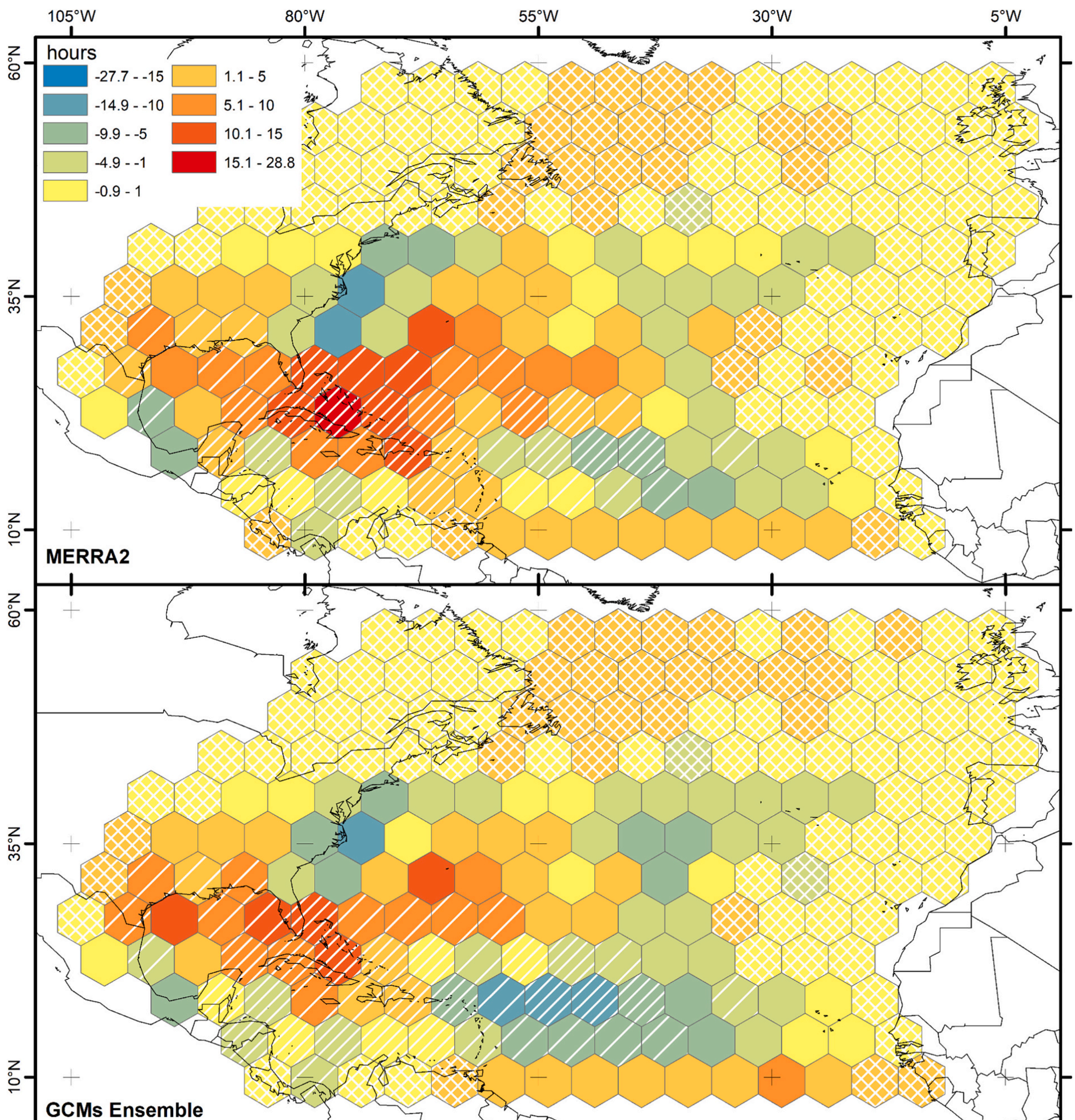

Fig. 10. Spatial differences between the best-scoring STC dataset and IBTrACS data for tropical storm residence time. No hatch: reliable cell; single white hatch: rather reliable cell; double white hatch: low reliable cell; crossed white hatch: not reliable.

3.4. General discussion

Discrepancies between STC datasets and IBTrACS arise from various factors, reflecting the complexities of modeling TCs and the inherent limitations of both datasets. STC models rely on assumptions and simplifications, such as uniform environmental conditions or generalized relationships between climate variables and cyclone behavior. These assumptions can lead to inaccuracies when actual atmospheric and oceanic conditions deviate from the models' assumptions. Simplified representations of physical processes like cloud formation, convection, and wind-field structures in the models further contribute to deviations from historical records.

STC datasets are typically generated under a broader range of hypothetical climatic conditions compared to the observed IBTrACS dataset, which reflects actual conditions. Variability in critical factors such as sea surface temperature, atmospheric pressure, and wind patterns at the time of cyclone formation significantly affect cyclone behavior. Even slight differences in initial conditions can lead to significant variations in the cyclone intensity, track, and duration, resulting in substantial divergences between STC and IBTrACS datasets.

Another key factor influencing discrepancies is the reliance of STC models on GCM outputs to simulate large-scale climatic conditions (Knutson et al., 2020). However, GCMs often exhibit biases, particularly in tropical regions. For example, overestimations or underestimations of sea surface temperatures, wind shear, and humidity levels, critical elements for TC formation and intensification, can propagate into STC datasets, introducing systematic errors when compared to IBTrACS. Additionally, GCM-derived STC events often display a wider range of intensities than historical records due to the extended temporal and hypothetical scope of the models, which increases the likelihood of generating extreme events.

The disparity in spatial and temporal resolutions between STC and IBTrACS datasets also contributes to observed differences. IBTrACS records, particularly from the pre-satellite era, often lack the resolution needed to fully capture cyclone lifecycles and intensities. In contrast, STC models typically produce high-resolution data, potentially capturing more detailed cyclone characteristics. However, this increased resolution can also introduce discrepancies, as finer details in STC data may need to align with coarser historical records. Furthermore, the nonlinear processes governing storm intensification, such as feedback mechanisms between atmospheric and oceanic systems, pose challenges for STC models, particularly for extreme events with wind speeds exceeding 100 ms^{-1} (Fig. 2, Table S1). While these extreme events may appear overrepresented in synthetic datasets, the validation paradox complicates determining their accuracy due to the limitations of historical observational data and the premise that STCs are designed to cover more extended periods or more scenarios. We could argue that STCs include events that have not occurred historically but are theoretically possible under current or future climate conditions (Lin and Emanuel, 2016). Including these hypothetical events can cause STC datasets to show higher frequencies or intensities of higher category events than those recorded in the historical data.

Human-mediated processes also influence IBTrACS data, which aggregates information from diverse sources such as ship reports, aircraft reconnaissance, satellite observations, and land-based weather stations (Emanuel, 2021b). Each observation type has limitations and potential for error. For example, early historical records may lack precision owing to less advanced technology, leading to the underestimation of storm intensity or incomplete track records (Vecchi and Knutson, 2008). In contrast, STC models use consistent methodologies, which might lead to more consistent but potentially less accurate representations of specific cyclone characteristics compared to more heterogeneous IBTrACS data. We might also mention that historical records, even if they were free of error, may no longer be a perfect guide to the present because of climate change that has already occurred (Thompson and Kuo, 2012).

The findings from the assessment of STC databases have several implications for policy decisions related to climate adaptation and disaster risk reduction. Understanding the biases and performances of these synthetic models is critical for shaping effective strategies in these areas. Studies have shown that certain models may overestimate or underestimate storm intensity in specific regions (Hodges et al., 2017). This information is critical for tailoring climate adaptation strategies to meet the regional needs. For instance, areas where models tend to underestimate storm intensity might require more conservative building codes. At the same time, regions with overestimation may focus on improving the precision of local climate models. Accurate modeling of TCs can inform the design of resilient infrastructures. By understanding where synthetic models align with historical data and where they diverge, policymakers can prioritize investments in areas most vulnerable to model inaccuracies, ensuring that infrastructure is built to withstand realistic worst-case scenarios.

The biases identified in different synthetic models can guide risk assessment processes concerning disaster risk reduction. For example, if a model consistently overestimates wind speeds in a particular area, it could lead to inflated risk assessments unless corrected. Policymakers can adjust their risk management strategies to account for these biases and improve resource allocation for disaster preparedness and response.

Insurance companies also rely on accurate risk assessments to set premiums and reserves (Raible et al., 2012). The findings of this study on model performance could influence how insurers calculate the risks associated with TCs, potentially leading to more equitable and accurate insurance pricing. Additionally, governments can use this for economic planning, especially in sectors such as agriculture and real estate, which are highly vulnerable to storm impact.

The validation paradox has implications for confidence in local-scale risk projections, as demonstrated by the high variability observed in our bootstrap analysis with $N = 942$ tracks. When synthetic datasets are resampled to match the limited size of historical records, model rankings shift substantially, and GCM-derived datasets exhibit particularly high MSE values and very low, sometimes negative, KGE' scores. This variability underscores a challenge: while STC models are designed to overcome data scarcity by generating thousands of events, the validation process itself is constrained by the same limited historical record that necessitated synthetic data in the first place. Moreover, the different seeding frameworks employed by various models introduce significant spatial biases that further complicate local validation. For instance, CHAZ draws seeds in proportion to the Tropical Cyclone Genesis Index and shows significant genesis hotspots north of Panama, whereas MIT seeds randomly over the ocean, resulting in more evenly distributed genesis patterns but with excess activity in the Antilles. These differences in spatial seeding strategies mean that even models using identical reanalysis forcing can produce markedly different storm populations in specific regions, making local-scale validation highly dependent on the particular seeding approach adopted. Consequently, local risk assessments based on STC data may carry significant uncertainty that cannot be adequately quantified using traditional validation approaches, as the spatial representation of storm activity is inherently model-dependent regardless of the quality of the underlying climate data.

4. Conclusions

The evaluation of STC databases provides valuable information on their effectiveness in replicating the critical IBTrACS parameters for risk analysis. Our research, which includes various STC downscaled from models such as GCMs, reanalysis datasets, and those from statistical models, highlights both the advantages and limitations of these synthetic methods.

Our findings indicate notable differences in the maximum wind speeds and storm intensity distributions among the different models, revealing ongoing challenges in accurately representing tropical cyclone intensity. Considering the maximum wind speed, the statistically derived STC, particularly Deltares, results in more accurate estimations than dynamically derived events, which is expected as they are derived directly from historical events. Statistical models have shown potential in specific areas but also exhibit explicit biases, particularly in the spatial representation of storm characteristics. Downscaling of reanalysis models, particularly MERRA2 and NCEP, have shown better performance in capturing translation speeds, TC frequency, and residence times than other STC datasets. However, inconsistencies in storm intensity distributions remain. STC downscaled of GCMs have demonstrated high variability in their performance, excelling in some aspects such as maximum wind speed (MIT-UKESM), translation speed, and tropical storm residence time (GCMs ensemble) but struggling with others such as hurricane residence time, reflecting the complexity of tropical cyclone modeling. However, the use of the GCMs ensemble is not well suited for maximum wind speed owing to the high extreme speeds in the database. Nevertheless, it yielded good results for TC frequency and residence time, demonstrating the usefulness of ensembles, as they help reduce the biases inherent to each model (Aijaz et al., 2019). It is relevant to note that even a hypothetically perfect downscaling technique applied to an imperfect climate model will give biased results, and a downscaling technique cannot compensate for what the GCM lacks. A bias correction of the GCM fields input to the downscaling could be implemented to compensate for their inaccuracies. Nonetheless, when downscaling reanalysis, most TC climatology errors are due to the downscaling technique, which still requires research and development. No single model or model type consistently outperformed the others for all tropical cyclone variables, highlighting the necessity of using a multimodel approach for thorough risk assessments.

These findings have important implications for climate science and hazard assessments. Although STC databases can extend limited historical records, their use should be approached cautiously, especially in areas or for variables where substantial model inaccuracies exist. The variability in model performance for different tropical cyclone characteristics emphasizes the need to choose appropriate models based on specific parameters of interest in risk assessments. The ongoing difficulties in accurately modeling tropical cyclone intensity, particularly for extreme events, point to a critical area for future research and model improvement. The success of ensemble approaches in addressing individual model weaknesses suggests that future risk assessments should integrate multiple model types to achieve more robust results.

In conclusion, this study highlights the complexity of tropical cyclone modeling and the need to refine our understanding and representation of these powerful weather systems. As climate change continues to impact tropical cyclone behavior, accurate risk assessment and climate projection modeling have become increasingly vital.

We recommend the following steps: refining STC generation methods with a focus on improving cyclone intensity representation when downscaling from GCMs; developing innovative approaches to combine the strengths of different model types; investigating sources of spatial biases in STC models to improve their reliability for localized risk assessments.

This study is intended as a diagnostic assessment of current STC track datasets and does not produce operational risk projections or insurance-grade hazard maps. Instead, by systematically evaluating model skill and identifying regions of high uncertainty, our results highlight where improvements in synthetic track generation (e.g., seeding strategy, enhanced downscaling, bias correction, or parameter tuning) are most needed. It is clear from our analysis that no single dataset is consistently more appropriate than others, as their accuracy differs between variables and spatial locations. As such, professionals performing activities such as coastal infrastructure planning, insurance modeling, and other applications should be aware of these limitations and use the databases accordingly, with ensemble approaches employed when possible. We anticipate that climate risk modelers, insurers, and policymakers can use these diagnostics to prioritize model development and data refinement before translating STC outputs into actionable risk estimates. By providing a critical evaluation of widely used STC datasets and identifying their strengths and weaknesses in replicating key storm characteristics, these insights are essential for researchers and stakeholders who rely on STC data for hazard assessment, insurance modeling, and climate adaptation planning.

CRediT authorship contribution statement

David Romero: Writing – original draft, Validation, Resources, Investigation, Formal analysis, Conceptualization, Writing – review & editing, Visualization, Software, Methodology, Funding acquisition, Data curation. **Christian M. Appendini:** Writing – review & editing, Visualization, Supervision, Investigation, Conceptualization, Writing – original draft, Validation, Methodology, Formal analysis. **Kerry Emanuel:** Validation, Data curation, Writing – review & editing, Methodology, Conceptualization. **Chia-Ying Lee:** Writing – review & editing, Methodology, Conceptualization, Validation, Data curation. **Kees Nederhoff:** Validation, Data curation, Writing – review & editing, Methodology, Conceptualization. **Nadia Bloemendaal:** Validation, Data curation, Writing – review & editing, Methodology. **Pablo Ruiz-Salcines:** Formal analysis, Writing – review & editing, Data curation. **Jonathan Vigh:** Writing – review & editing. **Christian Domínguez:** Writing – review & editing.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used *Grammarly* and *Paperpal* to improve readability of the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

UNAM-DGAPA PAPIIT Projects IA101823 and TA100420 provided funding for this project. CMA was supported by UNAM-DGAPA-PASPA and Aarhus University Research Foundation fellowships. This material is partially based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977. We would like to thank Gonzalo Martin Uriel Ruiz for IT support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosres.2025.108404.

Data availability

Data will be made available on request.

References

Aijaz, S., Kepert, J.D., Ye, H., Huang, Z., Hawksford, A., 2019. Bias correction of tropical cyclone parameters in the ECMWF ensemble prediction system in Australia. Mon. Weather Rev. 147 (11), 4261–4285. https://doi.org/10.1175/MWR-D-18-0377.1.

Appendini, C.M., Pedrozo-Acuña, A., Meza-Padilla, R., Torres-Freyermuth, A., Cerezo-Mota, R., López-González, J., Ruiz-Salcines, P., 2017. On the role of climate change on wind waves generated by tropical cyclones in the Gulf of Mexico. Coast. Eng. J. 59 (2), 1740001-1–1740001-32. https://doi.org/10.1142/S0578563417400010.

Appendini, C.M., Ruiz-Salcines, P., Duran, R., Marsooli, R., Al Azad, A.A., Emanuel, K., 2025. Redefining design wave conditions in the Gulf of Mexico under a changing climate. Ocean Eng. 334, 121685. https://doi.org/10.1016/j.oceaneng.2025.121685.

Bloemendaal, N., de Moel, H., Muis, S., Haigh, I.D., Aerts, J.C.J.H., 2020a. Estimation of global tropical cyclone wind speed probabilities using the STORM dataset. Sci Data 7 (1), 377. https://doi.org/10.1038/s41597-020-00720-x.

Bloemendaal, N., Haigh, I.D., De Moel, H., Muis, S., Haarsma, R.J., Aerts, J.C.J.H., 2020b. Generation of a global synthetic tropical cyclone hazard dataset using STORM. Sci Data 7 (1), 40. https://doi.org/10.1038/s41597-020-0381-2.

Bloemendaal, N., De Moel, H., Martinez, A.B., Muis, S., Haigh, I.D., Van Der Wiel, K., Haarsma, R.J., Ward, P.J., Roberts, M.J., Dullaart, J.C.M., Aerts, J.C.J.H., 2022. A globally consistent local-scale assessment of future tropical cyclone risk. Sci. Adv. 8 (17), eabm8438. https://doi.org/10.1126/sciadv.abm8438.

Boucher, O., Servonnat, J., Albright, A.L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Lionel, Guez E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A.K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., Vuichard, N., 2020. Presentation and evaluation of the IPSL-CM6A-LR climate model. J. Adv. Model. Earth Syst. 12 (7), e2019MS002010. https://doi.org/10.1029/2019MS002010.

Camargo, S.J., 2013. Global and regional aspects of tropical cyclone activity in the CMIP5 models. J. Clim. 26 (24), 9880–9902. https://doi.org/10.1175/JCLI-D-12-00549.1.

Camargo, S.J., Wing, A.A., 2016. Tropical cyclones in climate models. WIREs Clim. Change 7 (2), 211–237. https://doi.org/10.1002/wcc.373.

Camargo, S.J., Giulivi, C.F., Sobel, A.H., Wing, A.A., Kim, D., Moon, Y., Strong, J.D.O., Del Genio, A.D., Kelley, M., Murakami, H., Reed, K.A., Scoccimarro, E., Vecchi, G.A., Wehner, M.F., Zarzycki, C., Zhao, M., 2020. Characteristics of model tropical cyclone climatology and the large-scale environment. J. Climate 33 (11), 4463–4487. https://doi.org/10.1175/JCLI-D-19-0500.1.

Camargo, S.J., Murakami, H., Bloemendaal, N., Chand, S., Deshpande, M.S., Dominguez-Sarmiento, C., González-Alemán, J.J., Knutson, T.R., Lin, I.-I., Moon, I.-J., Patricola, C.M., Reed, K.A., Roberts, M.J., Scoccimarro, E., Tam, C.Y., Wallace, E.J., Wu, L., Yamada, Y., Zhang, W., Zhao, H., 2023. An update on the influence of natural climate variability and anthropogenic climate change on tropical cyclones. Trop. Cyclone Res. Rev. S2225603223000437. https://doi.org/10.1016/j.tcrr.2023.10.001.

Camargo, S.J., Tippett, M.K., Sobel, A.H., Lee, C.-Y., Fosu, B., Hodges, K.I., 2025. Tropical cyclones and associated environmental fields in CMIP6 models. J. Clim. 38 (15), 3877–3902. https://doi.org/10.1175/JCLI-D-24-0629.1.

Centella-Artola, A., Bezanilla-Morlot, A., Taylor, M.A., Herrera, D.A., Martinez-Castro, D., Gouirand, I., Sierra-Lorenzo, M., Vichot-Llano, A., Stephenson, T., Fonseca, C., Campbell, J., Alpizar, M., 2020. Evaluation of sixteen gridded precipitation datasets over the caribbean region using gauge observations. Atmosphere 11 (12), 1334. https://doi.org/10.3390/atmos11121334.

Coles, S., Bawa, J., Trenner, L., Dorazio, P., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D.A., DuVivier, A.K., Edwards, J., Emmons, L.K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M.M., Large, W.G., Lauritzen, P.H., Lawrence, D.M., Lenaerts, J.T.M., Lindsay, K., Lipscomb, W.H., Mills, M.J., Neale, R., Oleson, K.W., Otto-Bliesner, B., Phillips, A.S., Sacks, W., Tilmes, S., Van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J.E., Kinnison, D., Kushner, P. J., Larson, V.E., Long, M.C., Mickelson, S., Moore, J.K., Nienhouse, E., Polvani, L., Rasch, P.J., Strand, W.G., 2020. The community earth system model version 2 (CESM2). J. Adv. Model. Earth Syst. 12 (2), e2019MS001916. https://doi.org/10.1029/2019MS001916.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, S., Caron, L.-P., 2021. The EC-earth3 Earth system model for the climate model intercomparison project 6. In: Geoscientific Model Development Discussions. Göttingen, Germany, Vol. 2021, pp. 1–90.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7 (1), 1–26.

Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap, 1st ed. Chapman and Hall/CRC. https://doi.org/10.1201/9780429246593.

Elsner, J.B., Hodges, R.E., Jagger, T.H., 2012. Spatial grids for hurricane climate research. Clim. Dyn. 39 (1–2), 21–36. https://doi.org/10.1007/s00382-011-1066-5.

Emanuel, K., 2006. Climate and tropical cyclone activity: a new model downscaling approach. J. Clim. 19 (19), 4797–4802. https://doi.org/10.1175/JCLI3908.1.

Emanuel, K., 2008. Results from Downscaling IPCC AR4 Simulations, pp. 1–21. https://doi.org/10.1175/BAMS-89-3-xxx.

Emanuel, K., 2021a. Atlantic tropical cyclones downscaled from climate reanalyses show increasing activity over past 150 years. Nat. Commun. 12 (1), 7027. https://doi.org/10.1038/s41467-021-27364-8.

Emanuel, K., 2021b. Response of global tropical cyclone activity to increasing CO2: results from downscaling CMIP6 models. J. Clim. 34 (1), 57–70. https://doi.org/10.1175/JCLI-D-20-0367.1.

Emanuel, K., Ravela, S., Vivant, E., Risi, C., 2006. A statistical deterministic approach to hurricane risk assessment. Bull. Am. Meteorol. Soc. 87 (3), 299–314. https://doi.org/10.1175/BAMS-87-3-299.

Gathan, J., Knapp, K.R., Schreck, C.J.I., Diamond, H.J., Kossin, J.P., Kruk, M.C., 2024. International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4.1. NOAA National Centers for Environmental Information.

Gori, A., Lin, N., Xi, D., Emanuel, K., 2022. Tropical cyclone climatology change greatly exacerbates US extreme rainfall–surge hazard. Nat. Clim. Chang. 12 (2), 171–178. https://doi.org/10.1038/s41558-021-01272-7.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hall, T.M., Jewson, S., 2007. Statistical modelling of North Atlantic tropical cyclone tracks. Tellus A Dyn. Meteorol. Oceanogr. 59 (4), 486. https://doi.org/10.1111/j.1600-0870.2007.00240.x.

Harper, B., Kepert, J., Ginger, J., 2010. Guidelines for Converting between Various Wind Averaging Periods in Tropical Cyclone Conditions. Citeseer.

Hock, T.F., Franklin, J.L., 1999. The NCAR GPS dropwindsonde. Bull. Am. Meteorol. Soc. 80 (3), 407–420. https://doi.org/10.1175/1520-0477(1999)080<0407:TNGD>2.0.CO;2.

Hodges, K., Cobb, A., Vidale, P.L., 2017. How well are tropical cyclones represented in reanalysis datasets? J. Clim. 30 (14), 5243–5264. https://doi.org/10.1175/JCLI-D-16-0557.1.

Kantha, L., 2006. Time to replace the Saffir-Simpson hurricane scale? EOS Trans. Am. Geophys. Union 87 (1), 3–6. https://doi.org/10.1029/2006EO010003.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J. Hydrol. 424–425, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011.

Knapp KR, Kruk MC, Levinson DH, Diamond HJ, Neumann CJ. 2010. The international best track archive for climate stewardship (IBTrACS): unifying tropical cyclone data Bulletin of the American Meteorological Society. American Meteorological Society: Boston MA, USA, vol. 91(3): 363–376. doi: https://doi.org/10.1175/2009BAMS2755.1.

Knutson, T., Camargo, S.J., Chan, J.C.L., Emanuel, K., Ho, C.-H., Kossin, J., Mohapatra, M., Satoh, M., Sugi, M., Walsh, K., Wu, L., 2020. Tropical cyclones and climate change assessment: part II: projected response to anthropogenic warming. Bull. Am. Meteorol. Soc. 101 (3), E303–E322. https://doi.org/10.1175/BAMS-D-18-0194.1.

Konduru, R.T., Mrudula, G., Singh, V., Srivastava, A.K., Singh, A.K., 2023. Unravelling the causes of 2015 winter monsoon extreme rainfall and floods over Chennai: influence of atmospheric variability and urbanization on the hydrological cycle. Urban Clim. 47, 101395. https://doi.org/10.1016/j.uclim.2022.101395.

Kossin, J.P., 2018. A global slowdown of tropical-cyclone translation speed. Nature 558 (7708), 104–107. https://doi.org/10.1038/s41586-018-0158-3.

Landsea, C.W., Franklin, J.L., 2013. Atlantic hurricane database uncertainty and presentation of a new database format. Mon. Weather Rev. 141 (10), 3576–3592. https://doi.org/10.1175/MWR-D-12-00254.1.

Lee, C.-Y., Tippett, M.K., Sobel, A.H., Camargo, S.J., 2018. An environmentally forced tropical cyclone hazard model. J. Adv. Model. Earth Syst. 10 (1), 223–241. https://doi.org/10.1002/2017MS001186.

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. J. Clim. 33, 4815–4834. https://doi.org/10.1175/JCLI-D-19-0452.1.

Lee, C.-Y., Sobel, A.H., Camargo, S.J., Tippett, M.K., Yang, Q., 2022. New York state hurricane hazard: history and future projections. J. Appl. Meteorol. Climatol. 61 (6), 613–629. https://doi.org/10.1175/jamc-d-21-0173.1.

Lee, C.-Y., Sobel, A.H., Tippett, M.K., Camargo, S.J., Wüest, M., Wehner, M., Murakami, H., 2023. Climate change signal in Atlantic tropical cyclones today and near future. Earth's Future 11, e2023EF003539. https://doi.org/10.1029/2023EF003539.

Lin, N., Emanuel, K., 2016. Grey swan tropical cyclones. Nat. Clim. Chang. 6 (1), 106–111. https://doi.org/10.1038/nclimate2777.

Lin, I.-I., Camargo, S.J., Patricola, C.M., Boucharel, J., Chand, S., Klotzbach, P., Chan, J. C.L., Wang, B., Chang, P., Li, T., Jin, F.-F., 2020. ENSO and tropical cyclones. In: El Niño Southern Oscillation in a Changing Climate. American Geophysical Union (AGU), pp. 377–408.

Malakar, P., Kesarkar, A.P., Bhate, J.N., Singh, V., Deshamukhya, A., 2020. Comparison of reanalysis data sets to comprehend the evolution of tropical cyclones over North Indian Ocean. Earth Space Sci. 7 (2), e2019EA000978. https://doi.org/10.1029/2019EA000978.

Meiler, S., Vogt, T., Bloemendaal, N., Ciullo, A., Lee, C.-Y., Camargo, S.J., Emanuel, K., Bresch, D.N., 2022. Intercomparison of regional loss estimates from global synthetic tropical cyclone models. Nat. Commun. 13 (1), 6156. https://doi.org/10.1038/s41467-022-33918-1.

Meiler, S., Ciullo, A., Kropf, C.M., Emanuel, K., Bresch, D.N., 2023. Uncertainties and sensitivities in the quantification of future tropical cyclone risk. Commun. Earth Environ. 4 (1), 371. https://doi.org/10.1038/s43247-023-00998-w.

Müller, W.A., Jungclaus, J.H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornblueh, L., Li, H., Modali, K., Notz, D., Pohlmann, H., Roeckner, E., Stemmler, I., Tian, F., Marotzke, J., 2018. A higher-resolution version of the max planck institute earth system model (MPI-ESM1.2-HR). J. Adv. Model. Earth Syst. 10 (7), 1383–1413. https://doi.org/10.1029/2017MS001217.

Myhre, G., Samset, B., Forster, P.M., Hodnebrog, Ø., Sandstad, M., Mohr, C.W., Sillmann, J., Stjern, C.W., Andrews, T., Boucher, O., Faluvegi, G., Iversen, T., Lamarque, J.-F., Kasoar, M., Kirkevåg, A., Kramer, R., Liu, L., Mülmenstädt, J., Olivié, D., Quaas, J., Richardson, T.B., Shawki, D., Shindell, D., Smith, C., Stier, P.,

Tang, T., Takemura, T., Voulgarakis, A., Watson-Parris, D., 2022. Scientific data from precipitation driver response model intercomparison project. Sci Data 9 (1), 123. https://doi.org/10.1038/s41597-022-01194-9.
Nederhoff, K., Hoek, J., Leijnse, T., Van Ormondt, M., Caires, S., Giardino, A., 2021. Simulating synthetic tropical cyclone tracks for statistically reliable wind and pressure estimations. Nat. Hazards Earth Syst. Sci. 21 (3), 861–878. https://doi.org/10.5194/nhess-21-861-2021.
Neumann, C.J., Elms, J.D., 1993. Tropical Cyclones of the North Atlantic Ocean, 1871–1992. National Climatic Data Center.
Pebesma, E., 2018. Simple features for R: standardized support for spatial vector data. R J. 10 (1), 439–446. https://doi.org/10.32614/RJ-2018-009.
Qiu, J., Ravela, S., Emanuel, K., 2025. From decades to years: rising seas and cyclones amplify Bangladesh's storm-tide hazards in a warming climate. One Earth 8 (4), 101273. https://doi.org/10.1016/J.ONEEAR.2025.101273.
R Core Team, 2022. R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.
Raible, C.C., Kleppek, S., Wüest, M., Bresch, D.N., Kitoh, A., Murakami, H., Stocker, T.F., 2012. Atlantic hurricanes and associated insurance loss potentials in future climate scenarios: limitations of high-resolution AGCM simulations. Tellus A Dyn. Meteorol. Oceanogr. 64 (1), 15672. https://doi.org/10.3402/tellusa.v64i0.15672.
Rivera, P., 2023. Evaluation of historical simulations of CMIP6 models for temperature and precipitation in guatemala. Earth Syst. Environ. 7 (1), 43–65. https://doi.org/10.1007/s41748-022-00333-x.
Roberts, M.J., Camp, J., Seddon, J., Vidale, P.L., Hodges, K., Vannière, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E., Caron, L., Chauvin, F., Terray, L., Valcke, S., Moine, M., Putrasahan, D., Roberts, C.D., Senan, R., Zarzycki, C., Ullrich, P., Yamada, Y., Mizuta, R., Kodama, C., Fu, D., Zhang, Q., Danabasoglu, G., Rosenbloom, N., Wang, H., Wu, L., 2020. projected future changes in tropical cyclones using the CMIP6 HighResMIP multimodel ensemble. Geophys. Res. Lett. 47 (14), e2020GL088662. https://doi.org/10.1029/2020GL088662.
Ruiz-Salcines, P., Salles, P., Robles-Díaz, L., Díaz-Hernández, G., Torres-Freyermuth, A., Appendini, C.M., 2019. On the use of parametric wind models for wind wave modeling under tropical cyclones. Water 11 (10), 2044. https://doi.org/10.3390/w11102044.
Ruiz-Salcines, P., Appendini, C.M., Salles, P., Rey, W., Vigh, J.L., 2021. On the use of synthetic tropical cyclones and hypothetical events for storm surge assessment under climate change. Nat. Hazards 105 (1), 431–459. https://doi.org/10.1007/s11069-020-04318-9.
Scafetta, N., 2023. CMIP6 GCM ensemble members versus global surface temperatures. Clim. Dyn. 60 (9–10), 3091–3120. https://doi.org/10.1007/s00382-022-06493-w.
Seager, R., Cane, M., Henderson, N., Lee, D.-E., Abernathey, R., Zhang, H., 2019. Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. Nat. Clim. Chang. 9 (7), 517–522. https://doi.org/10.1038/s41558-019-0505-x.
Seager, R., Henderson, N., Cane, M., 2022. Persistent discrepancies between observed and modeled trends in the Tropical Pacific Ocean. J. Clim. 35 (14), 4571–4584. https://doi.org/10.1175/JCLI-D-21-0648.1.
Sellar, A.A., Walton, J., Jones, C.G., Wood, R., Abraham, N.L., Andrejczuk, M., Andrews, M.B., Andrews, T., Archibald, A.T., De Mora, L., Dyson, H., Elkington, M., Ellis, R., Florek, P., Good, P., Gohar, L., Haddad, S., Hardiman, S.C., Hogan, E., Iwi, A., Jones, C.D., Johnson, B., Kelley, D.I., Kettleborough, J., Knight, J.R., Köhler, M.O., Kuhlbrodt, T., Liddicoat, S., Linova-Pavlova, I., Mizielinski, M.S., Morgenstern, O., Mulcahy, J., Neininger, E., O'Connor, F.M., Petrie, R., Ridley, J., Rioual, J., Roberts, M., Robertson, E., Rumbold, S., Seddon, J., Shepherd, H., Shim, S., Stephens, A., Teixiera, J.C., Tang, Y., Williams, J., Wiltshire, A., Griffiths, P. T., 2020. Implementation of U.K. earth system models for CMIP6. J. Adv. Model. Earth Syst. 12 (4), e2019MS001946. https://doi.org/10.1029/2019MS001946.
Singh, V.K., Roxy, M.K., 2022. A review of ocean-atmosphere interactions during tropical cyclones in the North Indian Ocean. Earth Sci. Rev. 226, 103967. https://doi.org/10.1016/j.earscirev.2022.103967.
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.
Sobel, A.H., Lee, C.-Y., Bowen, S.G., Camargo, S.J., Cane, M.A., Clement, A., Fosu, B., Hart, M., Reed, K.A., Seager, R., Tippett, M.K., 2023. Near-term tropical cyclone risk and coupled Earth system model biases. Proc. Natl. Acad. Sci. 120 (33), e2209631120. https://doi.org/10.1073/pnas.2209631120.
Strazzo, S., Elsner, J.B., LaRow, T., Halperin, D.J., Zhao, M., 2013. Observed versus GCM-generated local tropical cyclone frequency: comparisons using a spatial lattice. J. Clim. 26 (21), 8257–8268. https://doi.org/10.1175/JCLI-D-12-00808.1.
Sugi, M., Murakami, H., Yoshida, K., 2017. Projection of future changes in the frequency of intense tropical cyclones. Clim. Dyn. 49 (1–2), 619–632. https://doi.org/10.1007/s00382-016-3361-7.
Tang, Y., Rumbold, S., Ellis, R., Kelley, D., Mulcahy, J., Sellar, A., Walton, J., Jones, C., 2019. MOHC UKESM1. 0-LL Model Output Prepared for CMIP6 CMIP Historical. World Data Center for Climate (WDCC) at DKRZ.
Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O'ishi, R., Yamazaki, D., Suzuki, T., Kurogi, M., Kataoka, T., Watanabe, M., Kimoto, M., 2019. Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. Geosci. Model Dev. 12 (7), 2727–2765. https://doi.org/10.5194/gmd-12-2727-2019.
Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., Van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K.B., Hurtt, G., Kriegler, E., Lamarque, J.-F., Meehl, G., Moss, R., Bauer, S.E., Boucher, O., Brovkin, V., Byun, Y.-H., Dix, M., Gualdi, S., Guo, H., John, J.G., Kharin, S., Kim, Y., Koshiro, T., Ma, L., Olivié, D., Panickal, S., Qiao, F., Rong, X., Rosenbloom, N., Schupfner, M., Séférian, R., Sellar, A., Semmler, T., Shi, X., Song, Z., Steger, C., Stouffer, R., Swart, N., Tachiiri, K., Tang, Q., Tatebe, H., Voldoire, A., Volodin, E., Wyser, K., Xin, X., Yang, S., Yu, Y., Ziehn, T., 2021. Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6. Earth Syst. Dynam. 12 (1), 253–293. https://doi.org/10.5194/esd-12-253-2021.
Thompson, L.G., Kuo, G., 2012. Climate change: the evidence and our options. In: World Futures Review, Vol. 4. SAGE Publications Inc, pp. 114–122. https://doi.org/10.1177/194675671200400215.
Thompson, D.T., Keim, B.D., Brown, V.M., 2024. Construction of a tropical cyclone size dataset using reanalysis data. Int. J. Climatol. 44 (9), 3028–3053. https://doi.org/10.1002/joc.8511.
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. J. Clim. 24 (9), 2335–2357. https://doi.org/10.1175/2010JCLI3811.1.
Vecchi, G.A., Knutson, T.R., 2008. On estimates of Historical North Atlantic Tropical Cyclone activity*. J. Clim. 21 (14), 3580–3600. https://doi.org/10.1175/2008JCLI2178.1.
Vecchi, G.A., Knutson, T.R., 2011. Estimating annual numbers of Atlantic Hurricanes missing from the HURDAT Database (1878–1965) using ship track density. J. Clim. 24 (6), 1736–1746. https://doi.org/10.1175/2010JCLI3810.1.
Vickery, P.J., Skerlj, P.F., Steckley, A.C., Twisdale, L.A., 2000. Hurricane wind field model for use in hurricane simulations. J. Struct. Eng. 126 (10), 1203–1221. https://doi.org/10.1061/(ASCE)0733-9445(2000)126:10(1203).
Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., Nabat, P., Roehrig, R., Salas Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., Waldman, R., 2019. Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. J. Adv. Model. Earth Syst. 11 (7), 2177–2213. https://doi.org/10.1029/2019MS001683.
Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., 2023. Dplyr: A Grammar of Data Manipulation.
Wilks DS. 2019. Statistical Methods in the Atmospheric Sciences. Elsevier: Amsterdam, Netherlands ; Cambridge, MA.
Xi, D., Lin, N., 2022. Understanding Uncertainties in Tropical Cyclone Rainfall Hazard Modeling using Synthetic Storms. J. Hydrometeorol. 23 (6), 925–946. https://doi.org/10.1175/JHM-D-21-0208.1.
Xi, D., Murakami, H., Lin, N., et al., 2025. Shifts of future tropical cyclone genesis date in north atlantic and north pacific basins: an ensemble modeling investigation. npj Clim Atmos Sci 8, 182. https://doi.org/10.1038/s41612-025-01077-x.
Yonekura, E., Hall, T.M., 2011. A statistical model of tropical cyclone tracks in the Western North Pacific with ENSO-dependent cyclogenesis. J. Appl. Meteorol. Climatol. 50 (8), 1725–1739. https://doi.org/10.1175/2011JAMC2617.1.
Zarzycki, C.M., 2022. Sowing storms: how model timestep can control tropical cyclone frequency in a GCM. J. Adv. Model. Earth Syst. 14 (3), e2021MS002791. https://doi.org/10.1029/2021MS002791.
Zhu, L., Emanuel, K., Quiring, S.M., 2021. Elevated risk of tropical cyclone precipitation and pluvial flood in Houston under global warming. Environ. Res. Lett. 16 (9). https://doi.org/10.1088/1748-9326/ac1e3d.