Natural Hazards (2025) 121:11957–11974
https://doi.org/10.1007/s11069-025-07268-2

ORIGINAL PAPER

Graph-theoretical investigation of trajectory dynamics and size characteristics in tropical cyclones

Yixiang Wang[1] · **Jiayao Wang**[2] · **Yu Chang**[3] · **Kang Cai**[2,4,5] · **Sunwei Li**[3] · **You Dong**[2]

Received: 16 June 2024 / Accepted: 29 March 2025 / Published online: 6 May 2025

Abstract

The intensification of climate changes has led to increased tropical cyclone (TC) intensities and subsequent damage, emphasizing the critical need for accurate trajectory prediction to mitigate their impact. In this study, a graph-theory-based approach was employed for the identification of TC trajectory. Using reanalysis data, each targeted TC can be constructed as a graph during its TC lifetime. Four graph metrics are computed from each graph constructed using different data sources, including mean sea level pressure, wind speed, and total precipitation. Among the graphs constructed, those representing mean sea level pressure (MSLP) and wind speed at 10 m (WD10) graphs show superior advantages in identifying TC trajectory. Furthermore, the metric PageRank of MSLP graph even reveals a notable ability to estimate TC size. Comparisons with a similar graph-theoretical approach demonstrate that our method exhibits superior performance in capturing complex TC dynamics. We anticipate to integrating the graph-theory-based approach into machine learning models to enhance the accuracy of predicting TC trajectories and intensities in future studies.

Keywords Tropical cyclone · Machine learning · Forecast model · Trajectory

✉ Jiayao Wang
jiayao.wang@connect.ust.hk

✉ Kang Cai
kangc_hk@163.com

1 Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

2 Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

3 Institute for Ocean Engineering, Tsinghua Shenzhen International Graduate School, Shenzhen, Guangdong, China

4 Institute of Structural Engineering, College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China

5 School of Civil Engineering and Architecture, Guangxi University, Nanning, China

1 Introduction

Tropical cyclones (TCs), also known as hurricanes or typhoons, are intense atmospheric dynamic activities characterized by strong winds, heavy rainfall, and storm surges. These powerful and destructive weather phenomena have significant impacts around the world. On 12 November 1970, a TC named Bhola caused the death of more than 250,000 people when striking Bangladesh and India's West Bengal, which became one of the deadliest TCs after 1900 (Hossain et al. 2008). Among the costliest TCs to strike the United States between 1900–2010 include Hurricane Katrina of 2005, which caused at least $108 billions of property damage and resulted in approximately 1200 deaths (Blake et al. 2011). In eastern Asia, typhoons have caused considerable economic losses in the last 3 decades in Korea and Japan with total losses exceeding USD $60 billion (Choi et al. 2016). Assessments of the expected response of TCs to anthropogenic warming indicate that climate change will likely lead to changes in the intensity and frequency of TCs (Knutson et al. 2020; Middelanis et al. 2022; Wang et al. 2022b), the consequence of which will lead to far-reaching economic, social, and environmental impacts on a global scale (Bakkensen and Mendelsohn 2016; Cao et al. 2023; Sharpe and Davison 2022; Wang et al. 2025).

Many meteorologists and warning centers have dedicated efforts to the research of TCs, leading to advancements in observational techniques, enhanced understanding of atmospheric physics, atmospheric environment interactions, air-sea interface dynamics, ocean response, and forecasting methodologies (Emanuel 2018; Wang et al. 2022a, b). In the light of the changes in TC behavior caused by climate change and the complex interactions between atmospheric and oceanic processes, as well as the influence of human activities on the climate system, accurate prediction of TC trajectory has become increasingly vital, empowering both citizens and governments to enhance disaster preparedness, evacuation planning, and resource allocation. However, when TCs form, they are influenced by various factors, including the meteorological environment, thermodynamics, and dynamics of the TC system. Following the landfall, the path of a TC is further influenced by complex factors including oceanic conditions, coastal topography, and inland terrain (Wang et al. 2023a; Yu et al. 2012). These complexities pose significant challenges to TC path prediction. Consequently, given the impact of TCs on society and the complexity of their prediction, exploring and implementing new techniques for TC path prediction holds great significance.

TC trajectory prediction encompasses three distinct perspectives, i.e., the analysis of large quantities of historical TC records, considering a variety of contextual information along with positional data, and forecast time (Farmanifard et al. 2023). Traditional track prediction methods are mainly based on numerical methods (Goerss 2000; Kotal and Bhowmik 2011). However, studies have highlighted several limitations associated with these approaches. Numerical weather prediction models are notably complex, requiring substantial time and computational resources, which can be both expensive and impractical (Chang et al. 2024; Hall and Jewson 2007; Vyavahare and Khanuja 2021). Such traditional methods are also constrained by their intrinsic characteristics. Such approaches often depend excessively on historical data patterns, overlooking the intricate interplay of meteorological variables and TC pathways across both attribute and time dimensions, which prevent them from analyzing implicit data features (Lian et al. 2020). Moreover, the rapid expansion of meteorological satellites, ocean observation stations, and ground stations has resulted in a continuously growing volume of data. While this increased data availability offers valuable insights, it also complicates TC trajectory forecasting. The large volume of high-dimensional and heterogeneous data requires advanced techniques for data assimilation, model calibration, and feature extraction. Traditional methods often struggle to effectively integrate and process such complex data, as they were not designed to handle the intricate relationships between multiple meteorological variables over different spatial and temporal scales. Consequently, while more data can potentially improve forecast accuracy, it also necessitates more sophisticated analytical methods to manage and interpret the information effectively.

The application of machine learning algorithms, which are recognized for their proficiency in deciphering intricate patterns within vast datasets for the prediction of TC trajectory, is becoming increasingly prevalent. A variety of machine learning techniques—ranging from artificial neural networks, support vector machines, and random forests—have been explored by researchers to enhance prediction accuracy (Lian et al. 2020; Moradi Kordmahalleh et al. 2016; Rüttgers et al. 2019). For example, Vyavahare and Khanuja (2021) employed the Convolutional Neural Network (CNN) technique to forecast cyclone paths, leveraging visual imagery obtained from satellites and radars. The accuracy of AI models was evaluated by comparing simulated trajectories with observed ones across different sea regions. In recent study, Bi et al. (2023) introduced the 3D neural network Pangu-Weather, achieving superior accuracy in medium-range global weather forecasting, particularly in predicting TC trajectories. Wang et al. (2023b) utilized recurrent neural networks (RNN), long short-term memory (LSTM), gated recurrent unit (GRU), and CNN to model complex relationships. Their model, incorporating reanalysis data of environmental factors, outperformed traditional methods in short-term track forecasting. Several scientific reviews have been published in the last decade to summarize the recent progress and advances in forecasting techniques for TCs. These techniques include statistical methods, dynamical and numerical methods, and more recently, machine learning-based methods (Chen et al. 2020; Elsberry 2014; Heming et al. 2019; Klotzbach et al. 2019; Leroux et al. 2018; Roy and Kovordányi 2012). In summary, the researches recognize the importance of leveraging advanced forecasting techniques, including machine learning-based methods, to deliver accurate cyclone forecasts in a timely manner, thus mitigating the impact on human lives and reducing economic loss.

This study presents a dynamical method using the concept of graph theory to identify the trajectory of a TC. This method utilizes the reanalysis data to construct a graph for a specific TC, from which some typical metrics could be calculated and used for the trajectory identification of a TC. We thoroughly discuss the properties of a TC graph and its metrics to find out that the metric of clustering coefficient from the mean sea level pressure (MSLP) graph is the most effective indicator for capturing the TC's trajectory, while the node degree from the graph based on wind speed at 10 m height graph is the second-best metric among all.

In the following parts of this paper, Sect. 2 gives the detailed methodology used for graph construction. Section 3 investigates the metrics of different graphs constructed by different data including the mean sea level pressure, wind speed at different altitudes, and total precipitation. In Sect. 4, the feasibility of identifying TC size from a metric of an MSLP graph is demonstrated, followed by a discussion about future study and a conclusion of this paper. Finally, concluding remarks are given in Sect. 5.

2 Methodology

2.1 Reanalysis data

This research applied the fifth-generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis data, ERA5, which is renowned for its high spatial and temporal resolution. The ERA5 dataset has a spatial resolution of 0.25 degrees for both latitude and longitude, with a temporal resolution of one hour. The analysis focuses on TCs in the East and Southeast Asia regions, and for this purpose, the data were selected from the geographical window extending from 105 °E to 160 °E and 5 °N to 40 °N. The period of study spans from June to November annually, which aligns with the TC season in the aforementioned area (Chow et al. 2018; Ho and Ying 2001; Wang et al. 2024). Furthermore, the current study incorporated the best track dataset for TCs, along with the TC size dataset retrieved from the China Meteorological Administration (CMA) Tropical Cyclone Data Center for the Western North Pacific Basin,[1] to augment our graph analysis and validate our findings.

MSLP is a widely recognized metric for accurately identifying the TC center path, as evidenced by its use in numerous previous studies (Gupta et al. 2021; Klotzbach et al. 2019). However, extra attention should be paid to situation where TCs cross over large islands or undergo extra-tropical transition if MSLP is the only data used for TC tracking. Heming (2017) suggests that the inclusion of additional data, such as the 850RV (relative vorticity) field, may enhance the precision of TC tracking. Therefore, this study selected the 10 m u and v wind components (WD10), MSLP, and total precipitation (PRCP) from the ERA5 hourly data on a single level, and additionally the study chose u, v, and w wind components at various pressure level at 250 hPa (WD25), 500 hPa (WD50) and 750 hPa (WD75) which are also from the ERA5 hourly dataset. The current research will firstly construct a graph for each TC using the corresponding MSLP data via Pearson correlation coefficient (PCC). One objective of the current study is to investigate whether the TC center path can be observed from the graph based on PCC of MSLP data. Furthermore, the wind speed at a height of 10 m height and three selected pressure levels will be compared with MSLP for the purpose of assessing the viability of using wind speed as an alternative dataset for the identification of TC center path from a graph-based analytical perspective.

2.2 Graph construction

The graph of each TC is constructed by the Pearson correlation coefficient. Firstly, the maximum and minimum longitude and latitude of the TC center path through one TC's lifetime are used as the reference boundaries for the spatial size of data used for graph construction. The boundaries are set to extend 5 degrees beyond these extremities, resulting in longitudinal limits of $\left[longitude_{min} - 5°, longitude_{max} + 5°\right]$ and latitudinal limits of $\left[latitude_{min} - 5°, latitude_{max} + 5°\right]$. Within this defined rectangular region, each geographic location is considered a node. Secondly, for each TC, all hourly data within the TC's lifetime for each node would be used for PCC calculation. This temporal framework is precisely determined using the best TC track dataset. Lastly, an edge is established between two nodes if their correlation coefficient exceeds a predetermined

[1] https://tcdata.typhoon.org.cn/en/index.html.

threshold, which varies according to the specific dataset being used for graph construction. The threshold and its application for different datasets will be discussed in subsequent sections. It is important to note that the graphs we generated in this study are undirected. This approach enables the systematic representation of the intricate relationships between various data points over the course of a TC's activity, thereby facilitating a more profound understanding of its characteristics.

2.3 Graph metrics

This study will utilize four node metrics to extract meaningful information from the graphs that represent our subject of interest. These metrics include the node degree, which measures the number of connections a node has; the clustering coefficient, which indicates the likelihood of a node's neighbors being connected; the mean geographical distance, providing an average measure of the spatial spread of a node's connections; and PageRank, which assesses the relative importance of a node within the graph. Each of these metrics will be instrumental in unraveling the intricate patterns and characteristics embedded in the data.

The metric degree k_i of a node i in a graph is the number of all neighbors connected to the node i.

$$k_i = \sum_{j=1}^{n} A_{ij} \tag{1}$$

where n is the total number of nodes; A_{ij} is the element at the i-th row and j-th column in the adjacency matrix A, if the edge of nodes i and j exists, $A_{ij} = 1$, otherwise, $A_{ij} = 0$. The clustering coefficient C_i is the ratio between the number of edges that truly exist among all neighbors of node i and the twice of total number when all neighbors of node i are fully connected (i.e., a complete graph) (Albert and Barabási 2002).

$$C_i = \frac{\sum_{j,q=1}^{n} A_{ij}A_{jq}A_{qi}}{k_i(k_i - 1)} \tag{2}$$

The clustering coefficient is a measure of the extent to which the neighbors of node i connect with each other. The mean geographical distance (MGD) $\mathcal{L}_i$ measures the averaged geographical distance between node i and all of its neighbors (Albert and Barabási 2002),

$$\mathcal{L}_i = \frac{\sum_{j=1}^{n} \mathcal{L}_{ij}A_{ij}}{k_i} \tag{3}$$

where the geographical distance $\mathcal{L}_{ij}$ is the great-circle distance between nodes i and j which is calculated by the Haversine equation. At last, we calculate the PageRank of a node i in a graph. It was originally developed by Google for the internet network. It is capable of quantifying the importance of a page (node) within a graph, which depends on the number and importance of the neighbor pages (nodes) that are linked to it (Albert and Barabási 2002).

The PageRank score of one node i is the sum of the PageRank score of its connected neighbors (Li and He 2018).

$$PR_i = \sum_{j=1}^{n} A_{ij} \frac{PR_j}{k_j} \tag{4}$$

3 Results

3.1 Graphs based on mean sea level pressure

Firstly, we select MSLP as the original data for graph construction. An edge is established between two nodes in the adjacency matrix if their PCC is equal to or greater than 0.9. Four metrics of the MSLP graph of the 13th TC Hato in 2017 in the Northwest Pacific Ocean are calculated and presented in Fig. 1. The trajectory of Hato is also integrated within the contours for reference. It should be emphasized that while the metric results presented in Fig. 1 do not include land-based data, the construction of the graph initially incorporates data from both over land and water to ensure comprehensive metric calculations.

The contour of the node degree, as illustrated in Fig. 1a, shows an intriguing correlation: the trajectory of the TC Hato corresponds to areas with lower node degree values. Given the edge definition from the previous section, a lower node degree indicates that a node with a diminished number of strong correlations with other nodes. This pattern suggests that nodes adjacent to the TC's trajectory form strong connections with only a limited number of nearby nodes. This characteristic could potentially be utilized to approximate the trajectory of a TC or even to make predictions based on historical data. However, node degree does not provide an accurate representation of the physical distance between the connected nodes and the center nodes. To address this, we use the MGD as the measure to

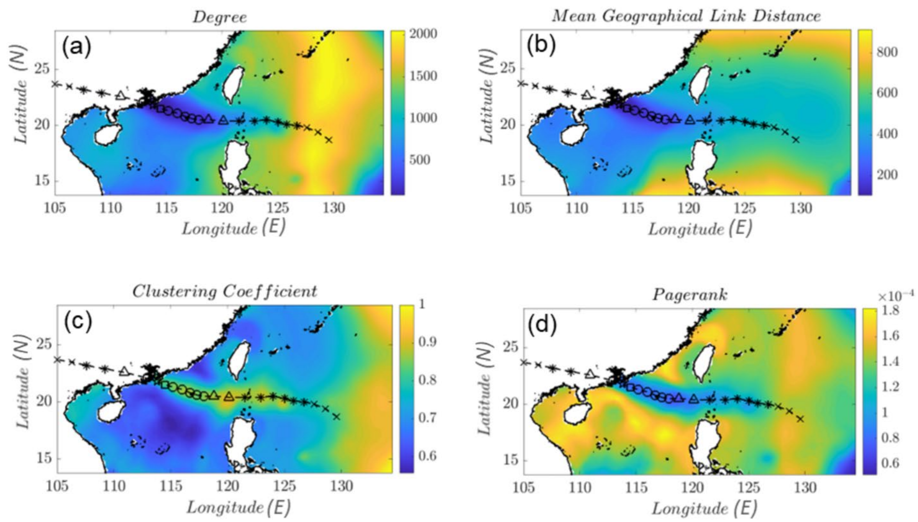

Fig. 1 Metrics of MSLP graph of TC 1713-Hato: **a** degree of nodes; **b** mean geographical distance; **c** clustering coefficient; **d** PageRank. Marker lines are TC trajectory with different marker shapes representing different TC intensities. TC intensity: ×—tropical depression, *—tropical storm, Δ—severe tropical storm, O—tropical cyclone, □—strong tropical cyclone, ☆—super tropical cyclone

quantify the spatial relationship between nodes. As shown in Fig. 1b, the TC's trajectory is in line with the relatively small MGD values, indicating that the neighbors connected to the center nodes are very close to the TC center physically. This phenomenon further reveals that only nodes or locations near the TC center are strongly correlated (PCC ≥ 0.9) to the TC center, while those far away from the center are irrelevant. These findings underscore the potential of MGD as a metric for enhancing the precision of TC center identification and trajectory prediction.

The clustering coefficient of TC 1713-Hato is depicted in Fig. 1c. The clustering coefficient is more distinguishable than the node degree, as it exhibits a much higher value along the TC's trajectory than the surrounding nodes. Moreover, it captures trends in the movement of the TC even in the early stages of their development. Considering the definition of the clustering coefficient, it becomes apparent that nodes with higher clustering coefficients along the trajectory of a Tropical Cyclone (TC) exhibit a stronger correlation among their physically proximate neighbors. This characteristic establishes the clustering coefficient as a superior method for identifying the trajectory of a TC. However, due to the contour's emphasis solely on the nodes at the TC center, determining the location of the neighboring nodes becomes challenging. Hence, an alternative metric called PageRank is explored, and its outcomes are depicted in Fig. 1d. It is evident that the outlines along the TC's trajectory are more distinguishable using this metric. Through a meticulous examination of the aforementioned three metrics and considering that the PageRank score of one node is the weighted average of PageRank score of those nodes connected to it (Eq. 4), it can be inferred that nodes with relatively low PageRank scores and low degree values on both sides of the TC's trajectory are connected to each other, resulting in a higher clustering coefficient along the TC's trajectory. Therefore, we can say these nodes are primarily significant to each other, since the PageRank score normally is explained as the node importance within a graph (Li and He 2018).

In order to assess the robustness of this method for identification of TC trajectory, we calculate the clustering coefficient for a range of TCs generated in the northwest Pacific Ocean. For the sake of simplicity, the results from four typical TCs are provided in Fig. 2. We investigate the influence of TC lifetime, TC location, and islands in the ocean on the identification of TC trajectory. Firstly, it is evident that even the TC with a lifetime as short as 3.5 days allows for the observation of its trajectory from the contour of the clustering coefficient (Fig. 2a). Regardless of whether TC during its lifetime is much closer to the Asian continent or meandering in the Pacific Ocean, the clustering coefficient is always high enough to capture the TC's trajectory (Fig. 2b, c). Finally, even in the case of tropical cyclones crossing numerous islands, such as those observed in the Philippines (Fig. 2d), the trajectory is consistent with high clustering coefficients. These evidences suggest that the clustering coefficient is sufficiently robust to identify the TC's trajectory.

Figure 2 illustrates a noteworthy phenomenon wherein nodes along the trajectory of a Tropical Cyclone (TC) consistently display significantly high clustering coefficients, approaching a value of 1.0. Upon scrutinizing the results obtained from all TCs in 2017, it becomes apparent that two potential rules can be formulated. The first rule suggests that the proximity of the TC center to land or the prolonged duration of the TC's presence in a confined region correlates with higher clustering coefficient values. The second rule is particularly remarkable. As depicted in Fig. 1, nodes in close proximity to the TC center exhibit limited interdependence. When a TC remains within a small region for an extended period, the nodes near the TC center accumulate longer MSLP data, which subsequently yields higher PCC values. Consequently, the clustering coefficient values also increase. Furthermore, since MSLP represents the pressure exerted on mean sea levels, it is

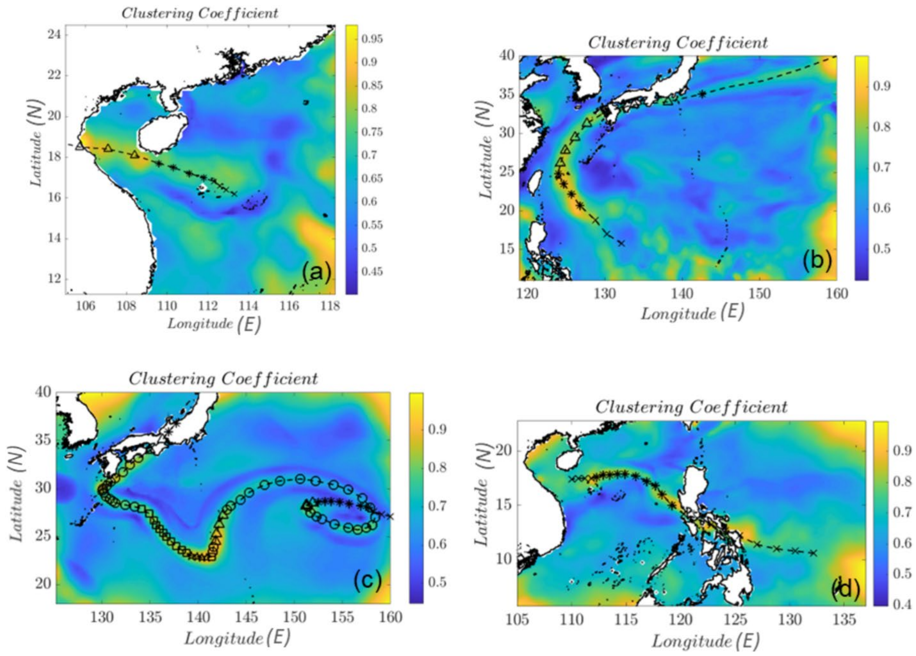

Fig. 2 Effect of TC lifetime, TC location, and islands in the ocean on identification of TC trajectory from clustering coefficient. **a** 1704-Talas (life time 3.5 days); **b** 1703-Nanmadol (6 days); **c** 1705-Noru (21 days); **d** 1724-Haikui (5.5 days)

reasonable to infer that when the TC center is closer to land, the nodes near the TC center on the ocean exhibit stronger mutual correlations through MSLP data, while displaying weaker correlations with nodes located on land. This inference might implicate that the MSLP data may only be effective for TC trajectory estimation over oceanic regions but is likely unreliable when applied to land areas after the TC makes landfall which is not investigated in this article.

3.2 Graphs based on wind speed

In this section, the wind speed at 10 m height is used for graph construction. Similar to MSLP graph, a connection is established between two nodes if their PCC is greater than or equal to 0.9. Consequently, four metrics of the WD10 graph could be calculated and are presented in Fig. 3. It is apparent to find a discernable trajectory of the 2nd TC Merbok in 2017 from the contours of node degree, mean geographical distance, clustering coefficient, and PageRank. This means that the 10 m wind speed data are practicable to identify the TC's trajectory. In addition, the results of mean geographical distance (Fig. 3b) are still highly corelated to those of node degree (Fig. 3a)) in the WD10 graph, reflecting that nodes at TC center are only connected to the nodes nearby. Similar to Fig. 1d, the relatively low PageRank score in Fig. 3d also clearly represents a higher clustering coefficient along the TC's trajectory.

However, notable differences exist between the metrics derived from WD10 data (Fig. 3) and those from MSLP data (Fig. 1). A key observation is that the regions

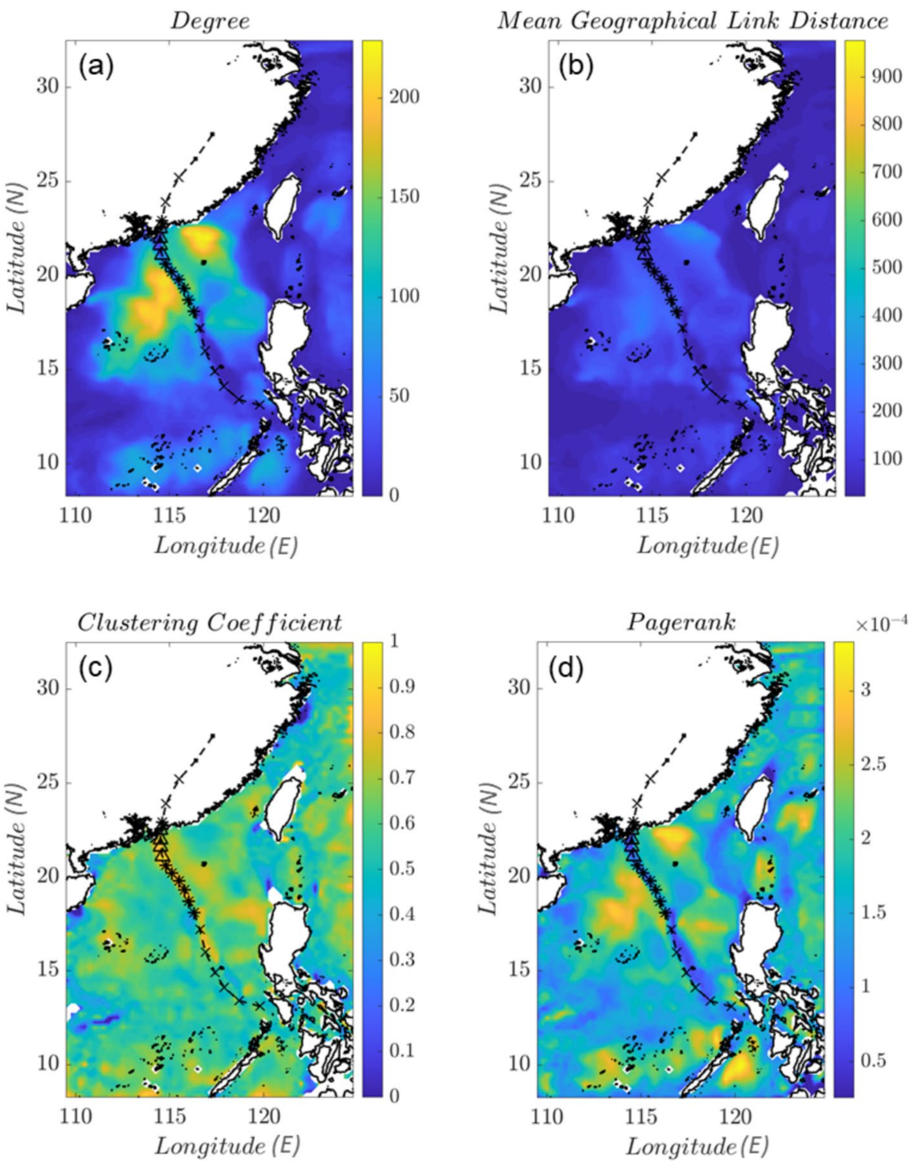

Fig. 3 Metrics of WD10 graph of TC 1702-Merbok. **a** Degree of nodes; **b** mean geographical link distance; **c** clustering coefficient; **d** PageRank

indicating the TC's trajectory in the WD10 graph are much narrower than those in the MSLP graph. This suggests that using WD10 data might lead to more precise identification of the TC's trajectory. It's important to point out, however, that the TC's trajectory is relatively challenging to discern from the clustering coefficient contour in the WD10 graph (Fig. 3c), particularly if the trajectory is not known beforehand. The nature of the clustering coefficient, which implies a dense interconnection among numerous neighboring nodes, may obscure the clarity of the contour, thereby complicating the identification of the TC's trajectory using this metric.

Besides, as shown in Fig. 3a, although the node degree remains regionally minimal along the TC's trajectory, there are large areas on both sides of trajectory with relatively high degree values, which are not evident in the contour of the node degree of the MSLP graph (Fig. 1a). This means that these nodes with a higher degree of connectivity to a vast number of other nodes, a phenomenon that is consistently observed in a scale-free graph. To further investigate this structure, we calculated the probability distribution function (PDF) of node degrees for the WD10 graphs of all TCs in 2017 with the results presented in Fig. 4. The evidence demonstrates that the average PDF adheres to a power law distribution across a range of one order of magnitude in node degree, thereby affirming the scale-free nature of the WD10 graph. Within this range, an area of high degrees can be identified, which can be characterized as "hubs". These hubs serve as pivotal relay stations facilitating information exchange among nodes dispersed across various locations. Remarkably, these hubs are consistently positioned at distances ranging from 0 to 2 degrees, both longitudinally and latitudinally, relative to the trajectory of the TC.

The objective of this study is to examine whether graphs constructed using wind speed data from different pressure levels could reflect similar metrics. The node degree of three graphs on 750 hPa, 500 hPa, and 250 hPa pressure levels for TC 1702-Merbok are presented in Fig. 5. The degree distribution of the 750hPa graph is strikingly similar to that of its WD10 counterpart, whereas the other two graphs exhibit markedly disparate patterns. Considering a TC has a complex 3D structure (Wang and Wu 2004; Wang 2012), it is reasonable to obtain different metrics from graphs defined by different sources. Figure 6 provides a proof of the vertical vorticity component at three pressure levels at 9 a.m. on 11 June, 2017. The vortex structure is completely destroyed at the 250hPa pressure level, which results in the inability to identify the TC's trajectory from its degree distribution at this altitude. It is important to note that the representation of the surface at this pressure level does not perfectly conform to the earth's curvature, complicating the data analysis. Therefore, it is preferable to construct graphs using wind speed data at 10 m height, despite the adequacy of metrics derived from the 750 hPa graph. The lower atmospheric data at 10 m provides a more reliable and accurate depiction of the TC's dynamics on a

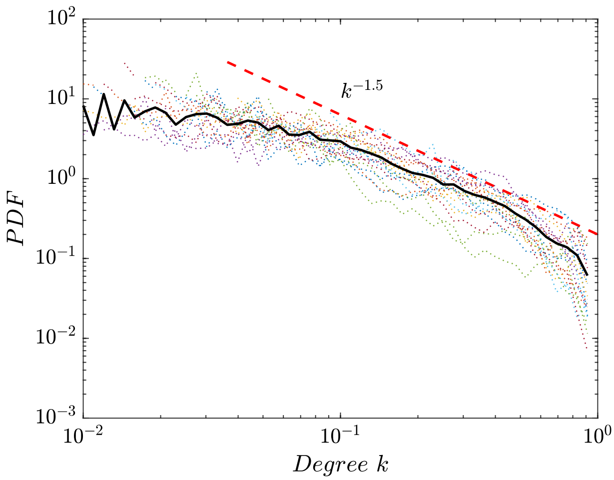

Fig. 4 PDF of node degree of WD10 graph for all TCs in 2017. The dotted lines represent the results of each TC in 2017 and black line is the average result of all TCs in 2017. The red dash line has an exponent of − 1.5

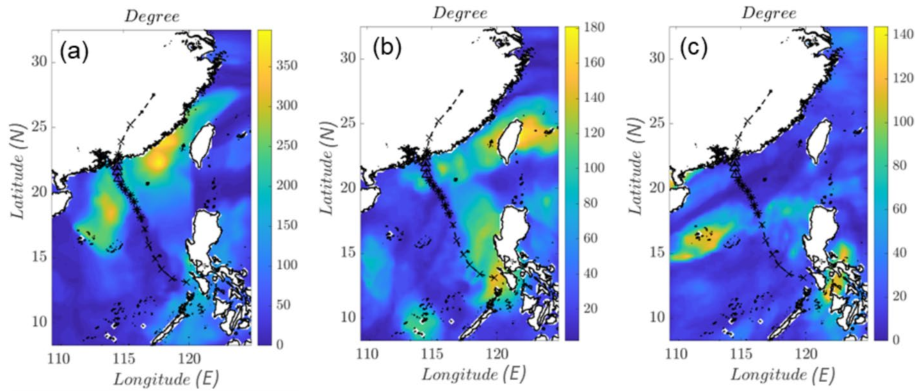

Fig. 5 Node degree of three graphs defined by wind speed on three pressure levels. Pressure level: **a** 750 hPa; **b** 500 hPa; **c** 250 hPa

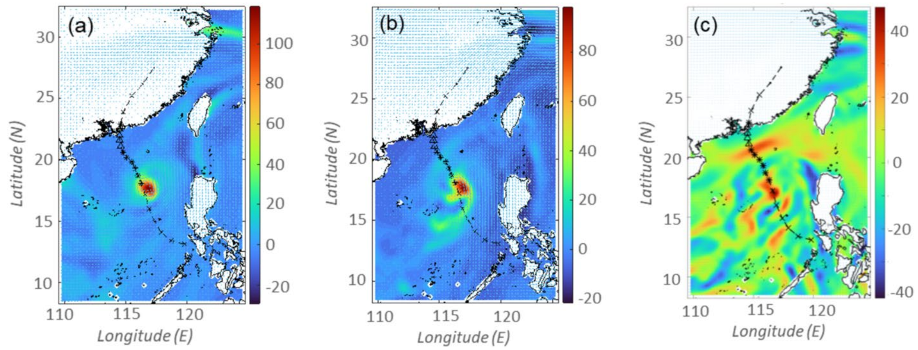

Fig. 6 Vertical vorticity ω_z on three pressure levels at 9am on 11 June 2017. Pressure level: **a** 750 hPa; **b** 500 hPa; **c** 250 hPa

surface closer to actual ground conditions, thereby ensuring better alignment with the TC's observed trajectory.

3.3 Graphs based on total precipitation

In this section, we conduct an analysis of the graph constructed using total precipitation data. The establishment of edges between nodes is contingent upon the PCC of total precipitation surpassing 0.8. Figure 7 presents the total precipitation and the distribution of node degree. Regrettably, the PRCP graph does not serve as a reliable indicator for identifying the trajectory of the TC. Consequently, the examination of other metrics is omitted in this section, as they also fail to provide clear insights into the TC's path. Nevertheless, the contour depicting total precipitation in Fig. 7a reveals a consistent association between heavy rainfall and the TC throughout its entire lifespan, with a predominant occurrence in the vicinity of the TC center. Furthermore, the regions experiencing precipitation exhibit a significant correlation with areas characterized by relatively high node degrees, as demonstrated in Fig. 7b. Similar to the WD10 and MSLP graphs, this observation can be

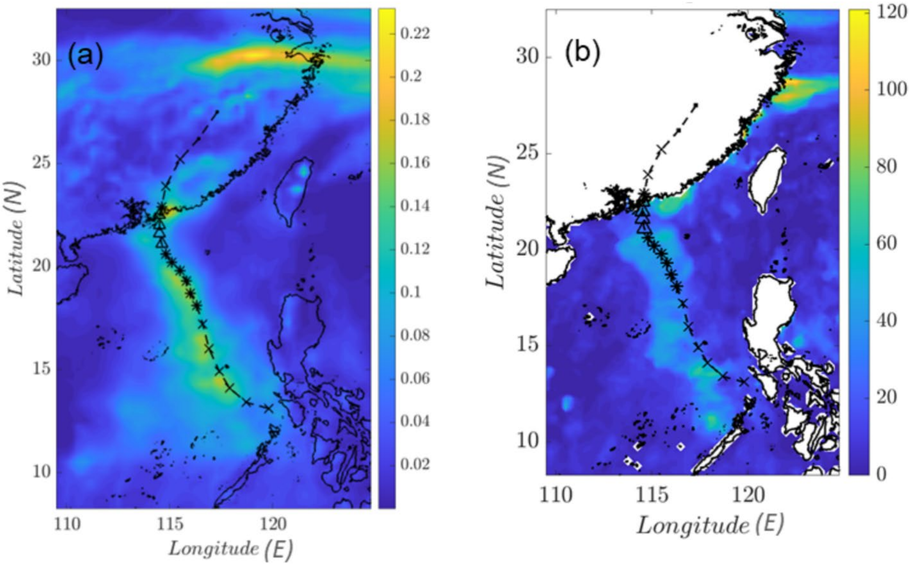

Fig. 7 **a** Total precipitation (m) during the whole life time of TC 1702-Merbok; **b** Node degree of PRCP graph of TC 1702-Merbok

attributed to the strong correlation between rainfall events near the regions encompassing the TC center.

4 Discussions

4.1 TC size identification from graphs

In this section, we explore the potential of using specific metrics from TC graphs to determine the size of a TC. Specifically, we examine the clustering coefficient and PageRank metrics within the MSLP graph of TC 1621-Sarika, which are presented in Fig. 8. This analysis utilizes data from TCs in 2016, as the CMA Tropical Cyclone Data Center (Lu et al. 2017) only provides the retrieved TC size dataset spanning from 1980 to 2016. The objective is to assess the potential of these metrics to effectively reflect the physical dimensions of a TC, with the aim of developing a new approach to estimating TC size based on graph-based analysis.

It is interesting to find that the TC size can be clearly identified from the PageRank contour in Fig. 8b. The radius of maximum wind (RMW) of the TC represented by the dotted circles is bounded by two yellow narrow stripes. Recall that PageRank represents the importance of a node in a graph, of which higher scores means higher importance. The higher importance locations coincide with the TC's RMW, emphasizing that the RMW is a very important position inside the TC. This finding also supports that it is of great significance to record the RMW for every TC. This is not only because it is the location with the highest wind speed, but also it would be meaningful for studying TC's structure geometrically and dynamically (Chavas et al. 2015; Holland 1980; Knaff et al. 2016; Li et al. 2022;

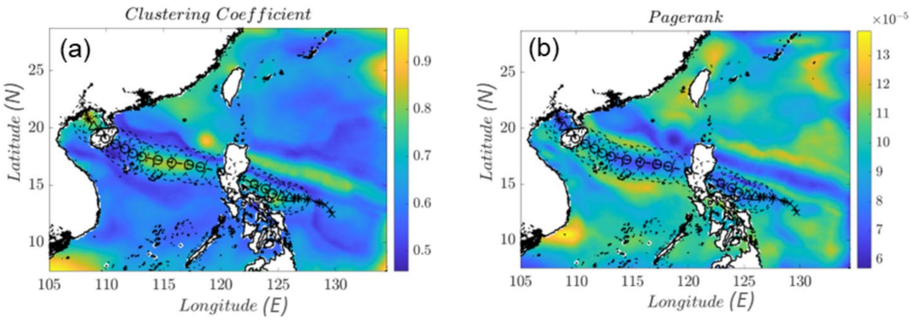

Fig. 8 Metrics of MSLP graph of TC 1621-Sarika: **a** Clustering coefficient; **b** PageRank. The dotted circles represent the radius of maximum wind speed of the TC

Nederhoff et al. 2019; Wood et al. 2013). In addition, a larger RMW is always connected to a higher PageRank score of MSLP (Figs. 8b, 9b), indicating that the larger TC structure has higher impact on MSLP within the whole lifetime of this TC. This is reasonable since a larger TC with lower MSLP in the center could absorb more energy from the ocean and influence the weather within a wider region (Sun et al. 2020).

The clustering coefficient, while serving as a useful indicator for tracking the trajectory of tropical cyclones, does not provide explicit information regarding the location of the TC's RMW, as illustrated in Fig. 8a. Only if when the size of the TC is known a priori, it becomes evident that two distinct stripes with lower scores of clustering coefficient approximately align with the RMW. Figure 8 demonstrates that these two metrics can be seen as complementary parts with each other. High PageRank regions correspond to low clustering coefficient regions, and vice versa. In Sect. 3 we already discussed the correlation between these two metrics.

Figure 9 compares the PageRank between WD10 and MSLP graphs. It is surprising to find that the PageRank of WD10 graph is unable to identify the RMW, although WD10 graph is constructed by wind speed at 10 m height. A reasonable inference could be that this mismatch means that the RMW is decorrelated with wind speed at 10 m height and strongly correlated with MSLP. Thus, we infer the sea level pressure plays a vital role for TC's structure.

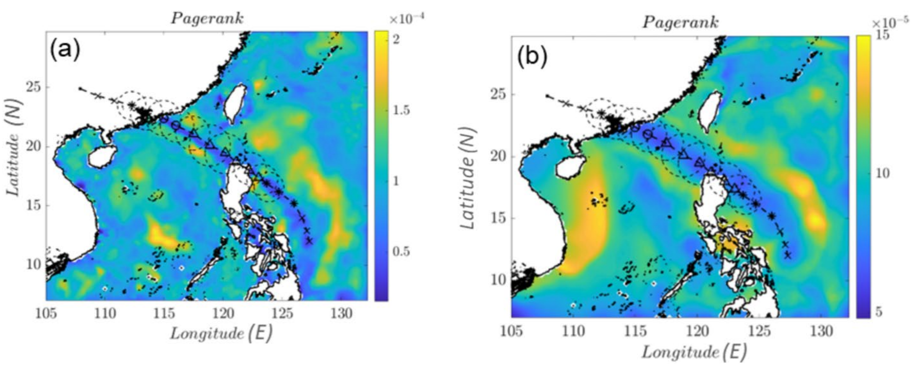

Fig. 9 PageRank of TC 1604-Nida from **a** WD10 and **b** MSLP graphs

4.2 Improvements

We compare our work with another similar study (Gupta et al. 2021), which also utilizes the graph theory to capture TC's trajectory. In their study, Gupta et al. (2021) constructed the evolving (continuous) networks using the ERA5 reanalysis data of 3-hourly MSLP only and a 10-day time window. The main distinction between our work and that of theirs lies in the fixed 10-day time window by the latter, which is unable to capture the trajectory of a TC with a very short lifetime, such as TC 1704-Talas shown in Fig. 10. The 6.5-day MSLP data, which are not included in the TC's lifetime, are not as strongly correlated with 3.5-day MSLP data within the TC's lifetime. Moreover, the feasibility of constructing a TC graph using alternative data, such as wind speed, total precipitation, and sea surface temperature was not investigated. Our work provides clear discrepancy between MSLP and WD10 graphs, despite the fact that both graphs are able to identify the TC's trajectory from their metrics. The third difference is the method employed to calculate the correlation coefficient. The TC graphs constructed in this study employ the Pearson correlation coefficient, whereas the Kendall's rank correlation coefficient was used in their study. Fortunately, two methods of correlation coefficient could both produce valid graphs for correct identification of TC's trajectory. Finally, we note that although we use 1-hourly data for graph construction, their study illustrated that 3-hourly data are sufficient for trajectory identification.

4.3 Future perspectives

The primary achievement of this study is the successful identification of TC trajectories through the analysis of various metrics derived from TC graphs, which are constructed using diverse types of data. The robustness of the proposed TC graph methodology is demonstrated by its capacity to accurately identify TC trajectories despite variations in the cyclone's lifetime and geographical location. Although our current research focuses on the analysis of graph metrics for previously occurred TCs, the foundational concept of constructing TC graphs holds substantial promise for future predictive applications. We envision extending this work to a data-driven neural network model that could leverage a graph neural network (GNN) for training and predictive purposes. By integrating

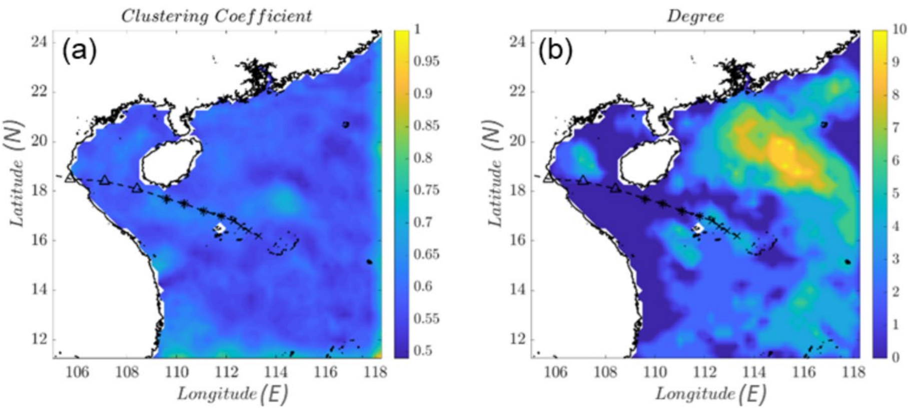

Fig. 10 Graph metrics of TC 1704-Talas. The graph is constructed by the Kendall's rank correlation coefficient. **a** Clustering coefficient of MSLP graph; **b** degree of WD10 graph

attention mechanisms and generative models, we expect that the model will be capable of extending its predictive capability longer enough for weather forecast purpose. Moreover, the application of techniques from explainable AI (xAI) has the potential to reveal the underlying physical mechanisms that govern the dynamics of TC systems. This approach not only enhances our predictive capabilities but also deepens our understanding of the behavior and driving forces of TCs.

5 Concluding remarks

In this paper, graphs of TCs are constructed using mean sea level pressure, wind speed at 10 m height, wind speed at three pressure levels, and total precipitation. The primary objective of this research is to investigate the feasibility of using these TC graphs to identify both the trajectory and the size of TCs, which are the main objectives of this study.

From the results of the metrics of MSLP and WD10 graphs, it is evident the trajectory and size of the TCs marked by RMW can be effectively identified. Among the various metrics analyzed, the clustering coefficient from the MSLP graph emerges as the most effective indicator for capturing the TC's trajectory, while the node degree from the WD10 graph is the second-best among all the metrics. The PDF of node degree also reveals the presence of a scale-free graph, indicating some regions near the TC center act as 'hubs' for information exchange within the TC. However, the results reveal that wind speeds at higher pressure levels are less suitable for TC graph construction due to the destruction of the mesoscale vortex structure of the TC at these altitudes. It is noteworthy that our study also demonstrates that the PageRank metric of a MSLP graph effectively captures the size of the TC, particularly reflecting the significance of the locations at the RMW. Overall, these insights not only enhance our understanding of TC dynamics but also contribute to the development of more accurate forecasting methods by identifying key graphical metrics that correlate with the physical characteristics of TCs.

Given the encouraging outcomes obtained in this study utilizing metrics derived from Tropical Cyclone graphs to ascertain TC trajectories, we propose the development of a data-driven Graph Neural Network (GNN) model aimed at predicting TC paths. This proposed model would leverage the fundamental principles established in our ongoing investigation of TC graphs. The initiation of this research endeavor is scheduled in the near future, with the objective of enhancing the precision and dependability of TC forecasting through the application of advanced machine learning techniques.

Author contributions Yixiang Wang: Data curation; Validation; Visualization; Writing—original draft; Jiayao Wang: Conceptualization; Formal analysis; Software; Writing—review & editing; Yu Chang: Methodology; Writing—review & editing; Kang Cai: Visualization; Writing—review & editing; Sunwei Li: Funding acquisition; Methodology; Writing—review & editing; You Dong: Writing—review & editing.

Funding Open access funding provided by The Hong Kong Polytechnic University. This work was supported by The Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. T22-501/23-R), and the grants from Guangdong Basic and Applied Basic Research Foundation (Project No. 2022B1515130006). The authors acknowledge the support from the Tsinghua Shenzhen International Graduate School—Shenzhen Pengrui Young Faculty Program of Shenzhen Pengrui Foundation (SZPR2023003).

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors have no conflicts to disclose.

References

Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47. https://doi.org/10.1103/RevModPhys.74.47

Bakkensen LA, Mendelsohn RO (2016) Risk and adaptation: evidence from global hurricane damages and fatalities. J Assoc Environ Resour Econ 3(3):555–587

Bi K, Xie L, Zhang H, Chen X, Gu X, Tian Q (2023) Accurate medium-range global weather forecasting with 3D neural networks. Nature 619(7970):533–538. https://doi.org/10.1038/s41586-023-06185-3

Blake ES, Landsea C, Gibney EJ (2011) The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts)

Cao S, Wang J, Tse TKT (2023) Life-cycle cost analysis and life-cycle assessment of the second-generation benchmark building subject to typhoon wind loads in Hong Kong. Struct Design Tall Spec Build 32(11–12):e2014. https://doi.org/10.1002/tal.2014

Chang Y, Wang J, Li S, Chan P (2024) A comprehensive review on the modeling of tropical cyclone boundary layer wind field. Phys Fluids 36(3):035165

Chavas DR, Lin N, Emanuel K (2015) A model for the complete radial structure of the tropical cyclone wind field Part I: comparison with observed structure. J Atmos Sci 72(9):3647–3662

Chen R, Zhang W, Wang X (2020) Machine learning in tropical cyclone forecast modeling: a review. Atmosphere 11(7):676

Choi W, Ho C-H, Jin C-S, Kim J, Feng S, Park D-SR, Schemm J-KE (2016) Seasonal forecasting of intense tropical cyclones over the North Atlantic and the western North Pacific basins. Clim Dyn 47:3063–3075. https://doi.org/10.1007/s00382-016-3013-y

Chow EC, Li RC, Zhou W (2018) Influence of tropical cyclones on Hong Kong air quality. Adv Atmos Sci 35:1177–1188. https://doi.org/10.1007/s00376-018-7225-4

Elsberry RL (2014) Advances in research and forecasting of tropical cyclones from 1963–2013. Asia-Pac J Atmos Sci 50:3–16. https://doi.org/10.1007/s13143-014-0001-1

Emanuel K (2018) 100 years of progress in tropical cyclone research. Meteorol Monogr 59:15.11-15.68. https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0016.1

Farmanifard S, AsgharAlesheikh A, Sharif M (2023) A context-aware hybrid deep learning model for the prediction of tropical cyclone trajectories. Expert Syst Appl 231:120701. https://doi.org/10.1016/j.eswa.2023.120701

Goerss JS (2000) Tropical cyclone track forecasts using an ensemble of dynamical models. Mon Weather Rev 128(4):1187–1193. https://doi.org/10.1175/1520-0493(2000)128%3c1187:TCTFUA%3e2.0.CO;2

Gupta S, Boers N, Pappenberger F, Kurths J (2021) Complex network approach for detecting tropical cyclones. Clim Dyn 57(11):3355–3364. https://doi.org/10.1007/s00382-021-05871-0

Hall TM, Jewson S (2007) Statistical modelling of North Atlantic tropical cyclone tracks. Tellus A Dyna Meteorol Oceanogr 59(4):486–498. https://doi.org/10.1111/j.1600-0870.2007.00240.x

Heming J (2017) Tropical cyclone tracking and verification techniques for Met Office numerical weather prediction models. Meteorol Appl 24(1):1–8

Heming JT, Prates F, Bender MA, Bowyer R, Cangialosi J, Caroff P, Coleman T, Doyle JD, Dube A, Faure G (2019) Review of recent progress in tropical cyclone track forecasting and expression of uncertainties. Trop Cyclone Res Rev 8(4):181–218. https://doi.org/10.1016/j.tcrr.2020.01.001

Ho W-S, Ying S (2001) An epidemiological study of 1063 hospitalized burn patients in a tertiary burns centre in Hong Kong. Burns 27(2):119–123. https://doi.org/10.1016/S0305-4179(00)00095-4

Holland GJ (1980) An analytic model of the wind and pressure profiles in hurricanes. Mon Wea Rev 108:1212–1218

Hossain M, Islam M, Sakai T, Ishida M (2008) Impact of tropical cyclones on rural infrastructures in Bangladesh. Agricultural Engineering International: CIGR Journal

Klotzbach P, Blake E, Camp J, Caron L-P, Chan JC, Kang N-Y, Kuleshov Y, Lee S-M, Murakami H, Saunders M (2019) Seasonal tropical cyclone forecasting. Trop Cyclone Res Rev 8(3):134–149. https://doi.org/10.1016/j.tcrr.2019.10.003

Knaff JA, Slocum CJ, Musgrave KD, Sampson CR, Strahl BR (2016) Using routinely available information to estimate tropical cyclone wind structure. Mon Weather Rev 144(4):1233–1247. https://doi.org/10.1175/MWR-D-15-0267.1

Knutson T, Camargo SJ, Chan JC, Emanuel K, Ho C-H, Kossin J, Mohapatra M, Satoh M, Sugi M, Walsh K (2020) Tropical cyclones and climate change assessment: part II: projected response to anthropogenic warming. Bull Am Meteor Soc 101(3):E303–E322

Kotal S, Bhowmik SR (2011) A multimodel ensemble (MME) technique for cyclone track prediction over the North Indian Sea. G Eofizika 28(2):275–291

Leroux M-D, Wood K, Elsberry RL, Cayanan EO, Hendricks E, Kucas M, Otto P, Rogers R, Sampson B, Yu Z (2018) Recent advances in research and forecasting of tropical cyclone track, intensity, and structure at landfall. Trop Cyclone Res Rev 7(2):85–105. https://doi.org/10.6057/2018TCRR02.02

Li K, He Y (2018) Evaluating nodes importance in complex network based on PageRank algorithm. Adv Mater Mach Electron II 1955(1):040122

Li Y, Wang Y, Tan ZM (2022) Why does the initial wind profile inside the radius of maximum wind matter to tropical cyclone development? J Geophys Res Atmos 127(16):e2022JD037039. https://doi.org/10.1029/2022JD037039

Lian J, Dong P, Zhang Y, Pan J, Liu K (2020) A novel data-driven tropical cyclone track prediction model based on CNN and GRU with multi-dimensional feature selection. Ieee Access 8:97114–97128. https://doi.org/10.1109/ACCESS.2020.2992083

Lu X, Yu H, Yang X, Li X (2017) Estimating tropical cyclone size in the Northwestern Pacific from geostationary satellite infrared images. Remote Sensing 9(7):728

Middelanis R, Willner SN, Otto C, Levermann A (2022) Economic losses from hurricanes cannot be nationally offset under unabated warming. Environ Res Lett 17(10):104013. https://doi.org/10.1088/1748-9326/ac90d8

Moradi Kordmahalleh M, Gorji Sefidmazgi M, Homaifar A (2016). A sparse recurrent neural network for trajectory prediction of atlantic hurricanes. In: Proceedings of the genetic and evolutionary computation conference 2016

Nederhoff K, Giardino A, van Ormondt M, Vatvani D (2019) Estimates of tropical cyclone geometry parameters based on best-track data. Nat Hazard 19(11):2359–2370. https://doi.org/10.5194/nhess-19-2359-2019

Roy C, Kovordányi R (2012) Tropical cyclone track forecasting techniques—a review. Atmos Res 104:40–69. https://doi.org/10.1016/j.atmosres.2011.09.012

Rüttgers M, Lee S, Jeon S, You D (2019) Prediction of a typhoon track using a generative adversarial network and satellite images. Sci Rep 9(1):6057. https://doi.org/10.1038/s41598-019-42339-y

Sharpe I, Davison CM (2022) A scoping review of climate change, climate-related disasters, and mental disorders among children in low-and middle-income countries. Int J Environ Res Public Health 19(5):2896. https://doi.org/10.3390/ijerph19052896

Sun J, Wang G, Xiong X, Hui Z, Hu X, Ling Z, Yu L, Yang G, Guo Y, Ju X, Chen L (2020) Impact of warm mesoscale eddy on tropical cyclone intensity. Acta Oceanol Sinica 39(8):1–13. https://doi.org/10.1007/s13131-020-1617-x

Vyavahare S, Khanuja H (2021) Fidelity of AI model in representing the different types of cyclone track estimation

Wang J, Cao S, Tim K, Li S, Fung JC, Li Y (2025) A novel life-cycle analysis framework to assess the performances of tall buildings considering the climate change. Eng Struct 323:119258

Wang J, Cao S, Zhang R, Li S, Tse TK (2024) Uncertainty of typhoon extreme wind speeds in Hong Kong integrating the effects of climate change. Phys Fluids 36(8):087126

Wang J, Tse KT, Li S, Fung CH (2023a) A model of the sea-land transition of the mean wind profile in the tropical cyclone boundary layer considering climate changes. Int J Disaster Risk Sci. https://doi.org/10.1007/s13753-023-00488-9

Wang J, Tse KT, Li S, Fung JC (2022a) Prediction of the typhoon wind field in Hong Kong: integrating the effects of climate change using the Shared Socioeconomic Pathways. Clim Dyn 59(7–8):2311–2329

Wang J, Tse KT, Li SW (2022b) Integrating the effects of climate change using representative concentration pathways into typhoon wind field in Hong Kong. In: 8th European African conference on wind engineering, Bucharest, Romania

Wang L, Wan B, Zhou S, Sun H, Gao Z (2023b) Forecasting tropical cyclone tracks in the northwestern Pacific based on a deep-learning model. Geosci Model Dev 16(8):2167–2179. https://doi.org/10.5194/gmd-16-2167-2023

Wang Y-Q, Wu C-C (2004) Current understanding of tropical cyclone structure and intensity changes—a review. Meteorol Atmos Phys 87(4):257–278

Wang Y (2012) Recent research progress on tropical cyclone structure and intensity. Trop Cyclone Res Rev 1(2):254–275. https://doi.org/10.6057/2012TCRR02.05

Wood VT, White LW, Willoughby HE, Jorgensen DP (2013) A new parametric tropical cyclone tangential wind profile model. Mon Weather Rev 141(6):1884–1909. https://doi.org/10.1175/MWR-D-12-00115.1

Yu J, Tang J, Dai Y, Yu B (2012) Analyses in errors and their causes of Chinese typhoon track operational forecasts. Meteorol Mon 2012(38):695–700. https://doi.org/10.7519/j.issn.1000-0526.2012.06.007