Tropical Cyclone Damage Assessment using a Projection Pursuit Dynamic Cluster Model

Chaoyong Tu
SYSU: Sun Yat-Sen University
Shumin Chen (chenshm35@mail.sysu.edu.cn)
Sun Yat-Sen University https://orcid.org/0000-0003-0687-8212
Zhongkuo Zhao
CMA: China Meteorological Administration
Weibiao Li
SYSU: Sun Yat-Sen University
Changjian Ni
CUIT: Chengdu University of Information Technology

Research Article

Keywords: Tropical cyclone, Damage, Disaster assessment, Projection pursuit dynamic cluster

Posted Date: January 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1127140/v1

Tropical Cyclone Damage Assessment using a Projection Pursuit Dynamic Cluster Model

Chaoyong Tu[1,2,3], Shumin Chen[1,2*], Zhongkuo Zhao[3], Weibiao Li[1,2*], Changjian Ni[4]

[1]School of Atmospheric Sciences, Sun Yat-Sen University, Guangzhou 510275, China

[2]Southern Laboratory of Ocean Science and Engineering, Zhuhai 519082, China

[3]Guangzhou Institute of Tropical and Marine Meteorology, China Meteorological Administration (CMA), Guangzhou 510080, China

[4]Plateau Atmospheres and Environment Key Laboratory of Sichuan Province, College of Atmospheric Science, Chengdu University of Information Technology, Chengdu 610225, China

***Corresponding authors:**

Shumin CHEN; +86-137-2670-3047; es04csm@mail2.sysu.edu.cn

Weibiao LI; +86-139-0225-8019; eeslwb@mail.sysu.edu.cn

Email addresses of other authors:

tuchy@mail2.sysu.edu.cn (C. T.); zhaozk@gd121.cn (Z. Z.);

eeslwb@mail.sysu.edu.cn (W. L.); ncj1970@163.com (C. N.);

Acknowledgments

We thank Tina Tin, PhD, from Liwen Bianji (Edanz) (www.liwenbianji.cn/), for editing the English text of a draft of this manuscript.

Abstract

Using data from 62 tropical cyclones (TCs) that landed in Guangdong Province in China between 2000 and 2019, we calculated six indices—minimum central pressure, maximum wind speed, maximum rainstorm ratio, cumulative surface rainfall, cyclone track length and lifetime—and

constructed a projection pursuit dynamic cluster (PPDC) model to assess TC damage risk. Although a single index may provide correct information on the intensity of certain types of damage, a comprehensive damage risk assessment cannot be obtained from individual indices alone. The PPDC model is a stable tool for TC damage risk assessment, especially in terms of economic loss, agricultural disaster area and disaster-affected population. Model validation improved the correlation of each of the indices. Output from the PPDC model for disaster-affected population and agricultural disaster-affected area also improved after model validation. We examined the limitations of the single indices using data from three TCs. Output from the PPDC model can closely reflect the intensity of the damage caused by the cyclones. Projection pursuit dynamic clustering is a new and objective method for typhoon damage risk assessment, and provides the scientific basis to support disaster prevention and mitigation.

Keywords

Tropical cyclone, Damage, Disaster assessment, Projection pursuit dynamic cluster

1. Introduction

Landfalling tropical cyclones (TCs) are usually accompanied by severe weather phenomena such as strong winds, heavy rains and storm surges, which cause heavy economic loss and large numbers of casualties in the affected countries and regions every year (Defu et al. 2009; 2014; Liu et al. 2020). Prevention and mitigation of disasters associated with TCs have always been a major concern. Therefore, research on typhoon intensity (Mei and Xie 2016), TC monitoring (Zhang et al. 2014), intensity assessment (Song et al. 2020; Xing et al. 2011), and disaster prevention capacity is of great importance to society because it supports industrial and agricultural operations.

Many damagerisk assessments focus on TC intensity, which is quantified by minimum sea level pressure or maximum surface wind speed and is strongly correlated with TC destructiveness. Meng et al. (2002; 2019) provided an overview of research on TCs that landed in China, tropical atmospheric dynamics, observations, and numerical predictions; they reported that large amounts of data on precipitation and TC track, intensity, and structure have been collected, and TC forecast

techniques have also been developed. Guan et al. (2018) pointed out that the percentage and annual average intensity of TCs that landed in East Asia and Southeast Asia have increased significantly between 1974 and 2013. Yuan et al. (2021) developed a long and short-term memory (LSTM) TC intensity prediction model on the basis of artificial neural networks; the model can provide meaningful results for the prediction of TC intensity within 120 hours. Liou et al. (2018) found that the winter cold front affects TC tracks and increases TC intensities. Guo et al. (2020) considered the combined effects of wind, surge and waves and used the Generalized Pareto Distribution (GPD) model to predict the risk of coastal hazards caused by TCs. Zong and Chen (1999) highlighted that the most direct and effective means to prevent TC disasters include government-led engineering projects, education programs and insurance policies that aim to reduce damage from TC winds, floods and waterlogging. Choun and Kim (2019) used Monte Carlo simulations to estimate TC wind speed and probable maximum wind speeds. They constructed logic trees using the pressure difference at the center of simulated TC wind fields, pressure profile parameter, and radius to maximum wind; they reported that the logic tree reduced the uncertainty associated with the wind intensity model and produced simulated wind speeds with an acceptable level of accuracy. Using translation velocity, storm heading and relative intensity of TC tracks in the Northwest Pacific basin as key parameters, Li and Hong (2016) developed a probabilistic model, and reported that simulated TC tracks could adequately reproduce historical best track data.

There are also TC intensity and damagerisk assessment models and systems that take into account other factors in addition to TC intensity. Fang et al. (2021) used the radius to maximum wind and a parameter of the shape of the pressure profile to reconstruct historical TC disasters, and introduced directional roughness length and a topographic speed-up factor. To build a wind hazard footprint dataset to promote risk assessment and disaster mitigation during TC events, Higa et al. (2021) integrated meteorological domain knowledge and developed a deep learning technique to estimate the intensity class of TC with high accuracy from a single satellite image. Hong et al. (2020) constructed a TC fast intensity model using the intensity and moisture dynamic equations and large-scale environmental features to predict the evolution of TC intensity; they estimated TC wind hazard along the southeastern coast of China, reported that model results were in close agreement with surface observations, and concluded that the model was suitable for engineering applications. Projection pursuit (Friedman and Tukey 1974; Glover and Hopke 1994) was first

proposed in the 1970s; it is a set of data analytic techniques used to identify structures in large multivariate datasets, and has been widely used in hydrology and meteorology. Wang and Ni (2008) combined dynamic cluster with projection pursuit, and developed a projection pursuit dynamic cluster (PPDC) model; their results show that the PPDC model is a powerful multi-factor cluster analysis tool, which can be applied to the regional division of water resources in China. This model has a number of advantages: it does not require the input of artificial parameters, it is stable and easy to operate, and dynamic cluster produces objective and clear results.

Damagerisk assessments provide the scientific basis for disaster prevention. In this study, we developed a landfalling TC classification index system using PPDC, TC intensity, track, duration and precipitation data from 45 TCs from 2000–2014. The PPDC model was validated using data from 17 TCs from 2015–2019. We introduce our data and methods in Section 2. Results are presented in Section 3. Section 4 provides a summary and discussion of the findings.

2. Data and Methods

2.1 Data

Integrated Multi-Satellite Retrievals for Global Precipitation Measurement (IMERG) products provide precipitation measurements with high spatial (0.1° x 0.1°) and temporal (half-hourly) resolutions. They are a Level-3 dataset and there are three categories of output; Early Run and Late Run consist of near-real-time monitoring products with delays of 6 and 18 hours, respectively; Final Run is a late-stage, real-time research product with a 4-month delay. In this study, we used Final Run products (Sun et al. 2018), which are freely available through the website of NASA (http://pmm.nasa.gov/data-access/downloads/gpm). Final Run products are adjusted using monthly surface precipitation gauge estimates from weather stations (Chen et al. 2020). The rainstorm ratio is defined as the ratio between daily rainfall that is greater than or equal to 50 mm and daily rainfall that is greater than or equal to 10 mm. The maximum rainstorm ratio is defined as the maximum rainstorm ratio during a TC. During TCs, light rainfall is relatively rare. Therefore, we excluded daily rainfall that is less than 10 mm from our analysis.

Best track data were obtained from the China Meteorological Administration (CMA) TC Data Center (http://tcdata.typhoon.org.cn) (Ying et al. 2014) and the Regional Specialized

Meteorological Center (RSMC) Tokyo-Typhoon Center (http://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/besttrack.html). Tropical cyclone positions, minimum central pressure, and maximum wind speed at surface level at 6-hourly intervals from July to October between 1979 and 2018 were available from both centers. Landfall locations in China were obtained from the CMA TC Data Center, and those in Japan were obtained from Digital Typhoon (http://www.digital-typhoon.org). Basic TC information such as time of TC birth and death, TC duration (hereafter referred to as lifetime), and minimum pressure is also available through the Digital Typhoon network.

2.2 The PPDC model

To retain the maximum amount of information and characteristics contained in the data, we analyzed the data from different angles to find the optimal projection direction. Projecting high-dimensional data on to a low-dimensional space allows for conventional and intuitive image analyses. Let the j^{th} index of the i^{th} sample be $x_{ij}^0 (i = 1,\ldots,n;\ \ j = 1,\ldots,m;$ where n is sample size, and m is the number of indicators selected), and the PPDC model was constructed as follows:

Step 1: Data standardization

The data were standardized to minimize differences in magnitude. For a large evaluation index, Eq. 1 was used:

$$x_{ij} = \frac{x_{ij}^0 - x_{j\min}^0}{x_{j\max}^0 - x_{j\min}^0}. \tag{1}$$

For a small index, Eq. 2 was used:

$$x_{ij} = \frac{x_{j\max}^0 - x_{ij}^0}{x_{j\max}^0 - x_{j\min}^0}, \tag{2}$$

where $x_{j\min}^0$ and $x_{j\max}^0$ are the sample minimum and maximum values of the j^{th} index, respectively.

Step 2: Linear projection

High-dimensional data were projected on to a linear space. Suppose $\vec{a}$ is an *m*-dimensional projection direction vector; its components are $a_1, a_2, \ldots, a_m$, which correspond to the weights of 1–*m* indicators. Suppose $\sum_{j=1}^{m} a_j = 1$, then the x_{ij} projection eigenvalue z_i can be expressed as:

$$z_i = \sum_{j=1}^{m} a_j x_{ij} \quad (i = 1,2 \ldots, n). \tag{3}$$

Step 3: Construction of projection indicators

To identify the optimal projection direction, high-dimensional data projections in low-dimensional space were clustered. Let $\Omega = \{z_1, z_2, \Lambda, z_n\}$ be the set of projected eigenvalue sequences of all samples. The set was dynamically clustered into p ($p \le n$) classes as follows:

(1) Randomly select *p* points and use them as *p* convergent nuclei, which are denoted as $L^0 = (A_1^0, A_2^0, \mathrm{L}\ , A_P^0)$.

(2) Divide the points in Ω by L^0; the points are classified into *p* categories. The results are denoted as $\Theta^0 = (\Theta_1^0, \Theta_1^0, \mathrm{L}\ , \Theta_p^0)$,

where $\Theta_i^0 = \{z \in \Omega \,|\, d(A_i^0 - z) \le d(A_j^0 - z), \forall j = 1,2,\mathrm{L}\ , p, j \ne i\}$, and $d(A_i^0 - z)$ is the absolute distance between point A_i^0 and any point in the set Ω.

(3) Starting from Θ, calculate the new polynucleus L^1, $L^1 = (A_1^1, A_2^1, \mathrm{L}\ , A_P^1)$, where $A_i^1 = \frac{1}{n_i} \sum_{z_i \in \Theta_i^0}^{p} z_i$, and there are n_i points in the class Θ_i^0.

(4) Repeat the steps above to obtain a classification result sequence $\vee^k = (L^k, \Theta^k)$, where

$k = 1, 2, \mathrm{L}$, $D(A_i^k, \Theta_i^k) = \sum_{z_i \in \Theta_i^k} | z_i - A_i^k |$, and $U_k = \sum_{i=1}^{p} D(A_i^k, \Theta_i^k)$. If $\frac{|U_{k+1} - U_k|}{U_{k+1}} \le \varepsilon$, the allowable error range is sufficiently small, and the algorithm is terminated. This algorithm is convergent.

The set of projected eigenvalues of all samples belonging to the h^{th} category is represented by $\Theta_h (h = 1, 2, \mathrm{L}\ , p)$; the absolute value of the distance between any two projected eigenvalues is defined as $d(z_i, z_j)$; $dd(a) = \sum_{h=1}^{p} D_h(a)$ represents the proximity of the samples within the class; $D_h(a) = \sum_{z_i, z_j \in \Theta_n} d(z_i, z_j)$ is the degree of aggregation within the class. A small $dd(a)$ indicates a high degree of aggregation of the samples within the class.

The degree of projection dispersion is defined as $ss(a) = \sum_{z_i, z_j \in \Omega} d(z_i, z_j)$. It represents the degree of dispersion of the projected eigenvalues of all sample sequences. Large projected eigenvalues of all sample sequences result in large *ss(a)*, and are associated with a high degree of dispersion.

The projection index $QQ(a)$ of the PPDC model is defined as:

$$QQ(a) = ss(a) - dd(a). \tag{4}$$

The projection index increases with increasing degree of projection dispersion, and with decreasing degree of clustering within the cluster. Maximum $QQ(a)$ is a result of the optimization between maximizing inter-class sample distances and minimizing intra-class samples distances.

Step 4: Model optimization

We applied the immune evolutionary algorithm (Chen and Huang 2005; Guo ZL et al. 2006) to identify the optimal projection direction vector $\overrightarrow{a^*}$ corresponding to maximum $QQ(a)$:

$$Min\, QQ(\vec{a}) \tag{5}$$

$$s.t. \sum_{j=1}^{m} a_j = 1, \tag{6}$$

where s.t. indicates constrained optimization.

2.3 Damage risk assessment indices

We focused on six indices as indicators of TC destructiveness: minimum central pressure, maximum wind speed, maximum rainstorm ratio, cumulative surface rainfall, and TC track length and lifetime. Minimum central pressure (Knaff and Zehr 2007; Hoarau et al. 2017) is the lowest pressure at the TC center; a strong TC has a low minimum central pressure. Maximum wind speed is the maximum average wind speed near the TC center; a strong TC has a high maximum wind speed (Zhang et al. 2014). The maximum rainstorm ratio (Su et al. 2012) is the ratio between the maximum daily precipitation that is greater than or equal to 50 mm and the daily precipitation that is greater than or equal to 10 mm during a TC. The cumulative surface rainfall (Jiang et al. 2008; Zhou and Matyas 2021) refers to the cumulative rainfall generated by the TC during the lifetime of the TC; a TC with high intensity has a high cumulative rainfall. Track length (Nakamura et al. 2015) refers to the distance covered by the TC during its lifetime; a longer track length is associated with greater impact. Lifetime refers to the length of time between the birth and death of the TC; a TC with a higher intensity has a longer lifetime and poses a higher damage risk.

3 Results

3.1 PPDC model combines advantages of individual indices

In this study, we used data from 62 TCs from 2000–2019. The PPDC model was developed using data from 45 TCs from 2000–2014, and it was validated using data from 17 TCs from 2015–2019. Our landfalling TC classification index system consists of six indices. These include minimum central pressure, maximum wind speed, maximum rainstorm ratio, cumulative surface rainfall, and TC track length and lifetime. Examples of these data are shown in Tables S1 and S2. Using the PPDC model, the data were divided into two clusters using the following settings: m=6, n=45, p=2.The projection direction vector was $\vec{a}^{*}$ = (0.1667, 0.0933, 0.1667, 0.1733, 0.2000, 0.2417).

Although a single index may provide correct information on the intensity of certain types of TC damage, a comprehensive damage risk assessment cannot be obtained from individual indices alone. Figure 1a–e shows the correlation coefficients between five parameters representing five types of TC damage (affected population, agricultural disaster area, number of collapsed buildings, number of deaths and direct economic loss) and the six indices. The correlation coefficient between minimum central pressure and affected population is 0.388, and that between minimum central pressure and agricultural disaster area is 0.4478. For affected population (Fig. 1a) and agricultural disaster area (Fig. 1b), correlation coefficients with maximum wind speed (0.349 and 0.411, respectively) are smaller than those with cumulative rainfall (0.462 and 0.598, respectively) and those with track length (0.411 and 0.510, respectively). For the number of collapsed buildings, correlation coefficients with minimum central pressure and maximum wind speed are 0.423 and 0.403, respectively, which are higher than the average value of the six correlation coefficients in Fig. 1c by 0.148 and 0.128, respectively. For the number of deaths, the correlation coefficients with minimum central pressure and maximum wind speed are 0.301 and 0.303, respectively, which are higher than the average value of the six correlation coefficients in Fig. 1d by 0.137 and 0.139, respectively. For affected population, agricultural disaster area, and direct economic loss, cumulative rainfall has the highest correlation coefficient of the six indices; for affected population, the correlation coefficient with cumulative rainfall is 0.462, while the average value of the six correlation coefficients is 0.333; for agricultural disaster area, the correlation coefficient with cumulative rainfall is 0.598, while the average value of the six correlation coefficients is 0.374; for direct economic loss, the correlation coefficient with cumulative rainfall is 0.561, while the average value of the six correlation coefficients is 0.385. However, the correlation coefficient between cumulative rainfall and number of collapsed buildings (0.182) and that between cumulative rainfall and number of deaths (0.040) are low. For the number of deaths, maximum rainstorm ratio has the highest correlation coefficient of the six indices (0.324). However, the correlation coefficients between maximum rainstorm ratio and the other four parameters are low. The mean of the five correlation coefficients between maximum rainstorm ratio and the parameters is 0.116, and it drops to 0.063 when the correlation coefficient with the number of deaths is excluded from the calculation of the mean. For lifetime, the correlation coefficients with affected population, number of collapsed buildings, and direct economic loss are relatively high (0.302, 0.266, and 0.215, respectively), but are not among the highest of the six indices, and the correlation coefficients with agricultural

disaster area (0.158) and number of deaths (0.014) are low.

We calculated the average of the correlation coefficients for each of the five parameters in Fig. 1. Direct economic loss has the highest average correlation coefficient (0.454), which is 0.081 higher than the average value of the correlation coefficients of the other four parameters. Agricultural disaster area has the second highest average correlation coefficient (0.437), which is 0.074 higher than the average value of the correlation coefficients of the other four parameters. The average correlation coefficient of affected population is 0.410, which is considerably higher than the average value of the correlation coefficients of the other four parameters (0.321). The average correlation coefficients of the number of collapsed buildings (0.313) and number of deaths (0.181) are lower, but remain higher than the average value of the correlation coefficients of the other corresponding four parameters. The PPDC considers all factors comprehensively, and TC hazard can be characterized by parameters such as affected population, agricultural disaster area, number of collapsed buildings, number of deaths and direct economic loss. The PPDC is a useful tool in the assessment of TC damage risk, and provides the scientific basis to support decisions for disaster prevention and mitigation.

We validated the model using data from 17 TCs from 2015–2019 (Fig. 2). Model output (Fig. 2) and observations (Fig. 1) exhibit similar trends, and there are some deviations. The mean of all the correlation coefficients in Fig. 1 is 0.297, and that in Fig. 2 is 0.330. For affected population (Fig. 2a), correlation coefficients with minimum central pressure (0.621), maximum wind speed (0.655), cumulative rainfall (0.550), and track length (0.568) are statistically significant, and are considerably higher than those in Fig. 1a. For agricultural disaster area (Fig. 2b), correlation coefficients with minimum central pressure (0.605), maximum wind speed (0.635), cumulative rainfall (0.405), and track length (0.442) are also high; coefficients with central minimum pressure and with maximum wind speed are higher than those in Fig. 1b; coefficients with cumulative rainfall and with track length are lower than those in Fig. 1b. For direct economic loss (Fig. 2c), correlation coefficients with minimum central pressure (0.413), maximum wind speed (0.516), cumulative rainfall (0.361), and track length (0.393) are relatively high, and are slightly smaller than those in Fig. 1c. For maximum rainstorm ratio, the trend in Fig. 1 is different from that in Fig. 2; correlation coefficient with the number of collapsed buildings is relatively high in Fig. 2, but is not the highest of the six indices; in Fig. 2, the average value of the correlation coefficients between

maximum rainstorm ratio and the other four parameters is 0.125. For lifetime, the trend in Fig. 1 is similar to that in Fig. 2; however, in Fig. 2, the correlation coefficients with the number of deaths (0.083) and with the number of collapsed buildings (0.061) are low. These findings may reflect the result of human defense against TCs; early detection of TCs increases the amount of time available for response and implementation of protection measures. For the number of collapsed buildings and the number of deaths (Fig. 2c–d), the average values of the correlation coefficients are low (0.220 and 0.116, respectively). These results may indicate that coastal cities have fully taken into account TC damage risk in architectural design and construction between 2015 and 2019. Reinforcement of buildings has been accompanied by the implementation of effective governmental measures to minimize damage risk and reduce casualties.

In Fig. 2, affected population has the highest average correlation coefficient (0.606) of the five parameters; this is 0.080 higher than the average correlation coefficients of the other four parameters in Fig. 2. The average correlation coefficient of affected population in Fig. 2 is also 0.196 higher than that in Fig. 1. For affected population, the average correlation coefficient is 0.483, which is higher than that in Fig. 1 (0.437). For direct economic loss, the average correlation coefficient is 0.413, which is lower than that in Fig. 1 (0.454). The PPDC model can adequately characterize affected population, agricultural disaster area and direct economic loss caused by TCs. Because coastal cities have fully considered the impact of TCs on architectural design and construction, the correlation coefficients for number of collapsed buildings and number of deaths are relatively low. In Fig. 2 and for the number of collapsed buildings, the average value of correlation coefficients is 0.234, and the average correlation coefficients of the other four parameters is 0.220, which is lower than the corresponding value in Fig. 1. In Fig. 2 and for the number of deaths, the average value of correlation coefficients is 0.128, and the average correlation coefficients of the other four parameters is 0.116, which is lower than the corresponding value in Fig. 1. The government has taken an active role in TC defense; casualties have been reduced as a result of these defense measures. The PPDC model can provide stable indicators that can be useful for the assessment of TC damage risk. Model output can also reflect changes that arise from mitigation efforts, and can be useful to support decisions for disaster prevention and mitigation.

3.2 Examination of different indices through case studies

We compared PPDC output with the intensity of the damage caused by three TCs between 2016 and 2018. For each of the six indices, we calculated the mean for each TC, and the mean for all three TCs (hereafter referred to as sample mean). Figures 3a and 4 show Typhoon Nida (2016), which started at 12.2° N, 127.6° E at 12:00 UTC on July 29, moved northwesterly, swept over the northern Philippines, and landed in Guangdong Province in China as a typhoon. The maximum wind speed of Nida was 30.9 m/s, which is 1.7 m/s lower than the sample mean; the minimum central pressure was 975 hPa, which is equivalent to the sample mean; the maximum rainstorm ratio was 0.1536, which is 76.04% that of the sample mean; the surface rainfall was 2.9×10^5 mm, which is 75.73% that of the sample mean; track length was 1,844 km, which is equivalent to the sample mean; lifetime was 78 hours, which is 23 hours lower than the sample mean (Fig. 5). Nida affected 470,000 people, damaged 40,000 hectares of agricultural land, and resulted in the collapse of 200 buildings, direct economic loss of 624.8 million yuan, and no casualties. The maximum wind speed, maximum rainstorm ratio, cumulative rainfall, and TC lifetime closely reflect the intensity of the disaster, while minimum central pressure and track length overestimate disaster intensity. Output from the PPDC model is 0.226, which is 61.41% that of the sample mean (Fig. 6); PPDC model output classifies Nida as a general TC, which is consistent with the damage caused by Nida.

Figures 3b and 7 show Typhoon Khanun (2017), which started at 16.6° N, 131.4° E at 12:00 UTC on October 11, crossed the northern part of the Philippines into the South China Sea in a westward direction, and landed as a severe tropical storm. The maximum wind speed of Khanun was 38.6 m/s, which is 6 m/s larger than the sample mean; the minimum central pressure was 955 hPa, which is 13 hPa lower than the sample mean; the maximum rainstorm ratio was 0.2643, which is 130.84% that of the sample mean; the surface rainfall was 3.6×10^5 mm, which is equivalent to the sample mean; track length was 1,688 km, which is 89.44% that of the sample mean; lifetime was 84 hours, which is 17 hours below than the sample mean. Khanun caused direct economic losses of about 1 billion yuan, affected approximately 570,000 people, damaged approximately 100,000 hectares of agricultural land, and resulted in the collapse of a small number of buildings. The maximum wind speed and central air pressure closely reflect the intensity of the disaster; lifetime and track length underestimate disaster intensity; the maximum rainstorm ratio overestimates disaster intensity. Output from the PPDC model is 0.3125 (Fig. 6), which indicates a strong TC and

is consistent with the actual situation. All indicators and PPDC model output closely reflect the intensity of the damage caused by Khanun.

Figures 3c and 8 show Typhoon Mangkhut (2018), which started at 12.9° N, 165.3° E at 12:00 UTC on September 7, moved westerly, landed in the northern Philippines, continued to move northwesterly, and finally landed in Guangdong Province in China as a strong typhoon. The maximum wind speed of Mangkhut was 56.6 m/s, which is 24.0 m/s higher than the sample mean; minimum central pressure was 905 hPa, which is lower than the sample mean by 58.0 hPa; maximum rainstorm ratio was 0.2645, which is higher than the sample mean; cumulative surface rainfall was 1.4×10^6 mm, which is 281% that of the sample mean; track length was 6,549 km, which is 247% that of the sample mean; lifetime was 234 hours, which is 124% that of the sample mean (Fig. 5). Mangkhut caused severe damage, affected 2.95 million people, and damaged approximately 260,000 hectares of agricultural land; it caused casualties and the collapse of approximately 1,400 buildings. Direct economic loss was approximately 14.4 billion yuan, which was the fourth largest economic loss of Guangdong since 2000. The maximum rainstorm ratio underestimates the intensity of the disaster; the maximum wind speed, central minimum pressure, cumulative rainfall, maximum rainstorm ratio, and track length overestimate disaster intensity. Output from the PPDC model is 0.8251 (Fig. 6), which is 124% that of the sample mean, which indicates a very strong TC and closely reflects the intensity of the damage caused by Mangkhut.

4. Discussion and conclusions

Using IMERG precipitation estimates from the TRMM satellite, CMA TC best track data, and data from Digital Typhoon, we examined 62 TCs that landed in Guangdong Province in China between 2000 and 2019. We assessed the TC damage using the indices of minimum central pressure, maximum wind speed, maximum rainstorm ratio, cumulative rainfall, TC track length and lifetime and a PPDC model. Our conclusions are as follows:

(1) Although a single index may provide correct information on the intensity of certain types of TC damage, a comprehensive damage risk assessment cannot be obtained from individual indices alone. For minimum central pressure and maximum wind speed, correlation coefficients with various types of TC damage are high. The maximum rainstorm ratio is only highly correlated with

342 the number of deaths. Tropical cyclone lifetime is only correlated with affected population.
343 Cumulative rainfall is only correlated with affected population, agricultural area and direct economy.
344 Direct economic loss is correlated with most indicators.

345 (2) The projection direction vector of the PPDC model was $\vec{a}^{*}$ = (0.1667, 0.0933, 0.1667,
346 0.1733, 0.2000, 0.2417). Output from the PPDC model can stably and adequately reflect the
347 intensity of various types of TC damage, and especially that of direct economic loss, agricultural
348 disaster area and affected population.

349 (3) Model validation improved the correlation of each of the six indices with the five types of
350 damage. For minimum central pressure, maximum wind speed, cumulative surface rainfall, and
351 track length, correlations with affected population, agricultural disaster area and direct economic
352 loss increased after model validation. The maximum rainstorm rate is only correlated with the
353 number of collapsed buildings. Output of the PPDC model for affected population and agricultural
354 disaster area also improved after model validation. The PPDC model is a stable tool for the
355 assessment of TC damage risk.

356 (4) We used the data from three TCs—Nida, Khanun, and Mangkhut—to examine the
357 limitations of single indices. Output from the PPDC model can adequately reflect the intensity of
358 the damage caused by the TCs. The PPDC can provide the scientific basis to support disaster
359 prevention and mitigation.

360 Projection pursuit dynamic cluster is a new and objective method for typhoon damage risk
361 assessment. However, the model in this study needs to be further improved to reflect the complexity
362 of TC damage risk. Future improvements include the use of indices that are related to physical
363 quantities and increasing the sample size in model calculations. Output from our PPDC model
364 includes high and low eigenvalues, which are not categorized according to intensity or grade. Future
365 studies can develop models that can evaluate and classify landfalling TCs according to the damage
366 caused by the TCs.

Declarations

Funding

This study was supported by the Fundamental Research Funds of the Special Program for Key Research and Development of Guangdong Province [Grant No. 2019B111101002], Guangzhou Science and Technology Planning Project [Grant No. 201903010036], China Postdoctoral Science Foundation [Grant No. 2020M683021], National Natural Science Foundation of China [Grant Nos. 42075004, 41875021, and 41830533].

Conflicts of Interest

The authors declare no conflicts of interest.

Availability of data and material

The TC best-track dataset related to this article can be found at [http://tcdata.typhoon.org.cn], hosted by the Shanghai Typhoon Institute of the China Meteorological Administration (Ying et al. 2014).

Code availability

Not applicable.

Authors' contributions

Conceptualization: Shumin Chen, Weibiao Li;
Methodology: Changjian Ni;
Formal analysis and investigation: Zhongkuo Zhao;
Writing - original draft preparation: Chaoyong Tu;
Writing - review and editing: Chaoyong Tu, Shumin Chen;
Funding acquisition: Shumin Chen, Weibiao Li;
Supervision: Shumin Chen.

References

Chen Y, Huang X (2005) An optimization method based on chaotic immune evolutionary algorithm. In: Wang L, Chen K, Ong YS (Eds) Advances in Natural Computation. ICNC 2005. Lecture Notes in Computer Science, 3611, Springer Berlin Heidelberg, pp 890–894. https://doi.org/10.1007/11539117_125

Chen Y, Zhang A, Zhang Y, Cui C, Wan R, Wang B, Fu Y (2020) A heavy precipitation event in the Yangtze River basin led by an eastward moving Tibetan Plateau cloud system in the summer of 2016. J Geophys Res Atmos 125(15). https://doi.org/10.1029/2020JD032429

Choun Y, Kim M (2019) Logic tree approach for probabilistic typhoon wind hazard assessment. Nucl Eng Technol 51(2):607–617. https://doi.org/10.1016/j.net.2018.11.006

Defu L, Hao J, Fengqing W, Yongfa W, Lingling Y (2009) Prediction of typhoon triggered sea hazards in China. OCEANS 2009-EUROPE 1–7. https://doi.org/10.1109/OCEANSE.2009.5278158

Defu L, Guilin L, Hongda S, Fengqing W, Ziyu C (2014) Extreme sea hazards statistics and engineering applications. OCEANS 2014-TAIPEI 1–8. https://doi.org/10.1109/OCEANS-TAIPEI.2014.6964285

Fang G, Pang W, Zhao L, Rawal P, Cao S, Ge Y (2021) Toward a refined estimation of typhoon wind hazards: Parametric modeling and upstream terrain effects. J Wind Eng Industrial Aerodynamics 209. https://doi.org/10.1016/j.jweia.2020.104460

Friedman JH, Tukey JW (1974) A projection pursuit algorithm for exploratory data analysis. IEEE Transactions on Computers C-23(9):881–890. https://doi.org/10.1109/T-C.1974.224051

Glover DM, Hopke PK (1994) Exploration of multivariate atmospheric particulate compositional data by projection pursuit. Atmos Environ 28(8):1411–1424. https://doi.org/10.1016/1352-2310(94)90204-6

Guan S, Li S, Hou Y, Hu P, Liu Z, Feng J (2018) Increasing threat of landfalling typhoons in the

western North Pacific between 1974 and 2013. Int J App Earth Observation and Geoinformation68:279–286. https://doi.org/10.1016/j.jag.2017.12.017

Guo Y, Hou Y, Liu Z, Du M (2020) Risk prediction of coastal hazards induced by typhoon: a case Study in the coastal region of Shenzhen, China. Remote Sens 12(11):1731. https://doi.org/10.3390/rs12111731

Guo ZL, Wang S, Zhuang J (2006) A novel immune evolutionary algorithm incorporating chaos optimization. Pattern Recognit Lett 27(1):2–8. https://doi.org/10.1016/j.patrec.2005.06.014

Higa M, Tanahara S, Adachi Y, Ishiki N, Nakama S, Yamada H, Ito K, Kitamoto A, Miyata R (2021) Domain knowledge integration into deep learning for typhoon intensity classification. Scientific Rep 11(1). https://doi.org/10.1038/s41598-021-92286-w

Hoarau K, Lander M, De Guzman R, Guard C, Barba R (2017) Did Typhoon Haiyan have a new record-minimum pressure? Weather 72(10):291–295. https://doi.org/10.1002/wea.3045

Hong X, Kareem A, Li J (2020) Validation of the fast intensity model for typhoon and its application to the estimation of typhoon wind hazard for the southeast coast of China. J. Wind Eng. Industrial Aerodynamics 206. https://doi.org/10.1016/j.jweia.2020.104379

Jiang H, Halverson JB, Zipser EJ (2008) Influence of environmental moisture on TRMM-derived tropical cyclone precipitation over land and ocean. Geophys Res Lett 35(17). https://doi.org/10.1029/2008GL034658

Knaff JA, Zehr RM (2007) Reexamination of tropical cyclone wind–pressure relationships. Weather and Forecasting 22(1):71–88. https://doi.org/10.1175/WAF965.1

Li SH, Hong HP (2016) Typhoon wind hazard estimation for China using an empirical track model. Nat Hazards 82(2):1009–1029. https://doi.org/10.1007/s11069-016-2231-2

Liou Y, Liu J, Liu CP, Liu C (2018) Season-dependent distributions and profiles of seven super-typhoons (2014) in the northwestern Pacific Ocean from satellite cloud images. IEEE Transactions on Geoscience and Remote Sens 56(5):2949–2957. https://doi.org/10.1109/TGRS.2017.2787606

443 Liu G, Li X, Wang J, Kou Y, Wang X (2020) Research on the statistical characteristics of typhoon
444 frequency. Ocean Eng 209, 107489. https://doi.org/10.1016/j.oceaneng.2020.107489

445 Mei W, Xie S (2016) Intensification of landfalling typhoons over the northwest Pacific since the late
446 1970s. Nat Geoscience 9(10):753–757. https://doi.org/10.1038/ngeo2792

447 Meng ZY, Chen LS, Xu XD (2002) Recent progress on tropical cyclone research in China. Adv Atm
448 Sci 19(1):103–110. https://doi.org/10.1080/15481603.2021.1908675

449 Meng ZY, Zhang F, Luo D, Tan Z, Fang J, Sun J, Shen X, Zhang Y, Wang S, Han W, Zhao K, Zhu L,
450 Hu Y, Xue H, Ma Y, Zhang L, Nie J, Zhou R, Li S, Liu H, Zhu Y (2019) Review of Chinese
451 atmospheric science research over the past 70 years: Synoptic meteorology. Sci China Earth
452 Sci 62(12):1946–1991. https://doi.org/10.1007/s11430-019-9534-6

453 Nakamura J, Lall U, Kushnir Y, Rajagopalan B (2015) HITS: Hurricane Intensity and Track
454 Simulator with North Atlantic Ocean applications for risk assessment. J Appl Meteorol
455 Climatol 54(7):1620–1636. https://doi.org/10.1175/JAMC-D-14-0141.1

456 Song J, Duan Y, Klotzbach PJ (2020) Revisiting the relationship between tropical cyclone size and
457 intensity over the western North Pacific. Geophys Res Lett 47(13).
458 https://doi.org/10.1029/2020GL088217

459 Su S, Kuo H, Hsu L, Yang Y (2012) Temporal and spatial characteristics of typhoon extreme
460 rainfall in Taiwan. J Meteorol Soc Japan Ser II 90(5):721–736.
461 https://doi.org/10.2151/jmsj.2012-510

462 Sun W, Sun Y, Zhang Y, Qiu Q, Wang T, Wang Y (2018) Ground validation of GPM IMERG
463 rainfall products over the capital circle in northeast China on rainstorm monitoring. Remote
464 Sens Agric Ecosyst Hydrol XX 10783. https://doi.org/10.1117/12.2500087

465 Wang S, Ni C (2008) Application of projection pursuit dynamic cluster model in regional partition
466 of water resources in China. Water Resour Manag 22(10):1421–1429.
467 https://doi.org/10.1007/s11269-007-9234-4

468 Xing L, Hu D, Tang LT (2011) Development of typhoon disaster risk evaluation and early warning
469 system integrating real-time rainfall data from the satellite. 2011 19th International

Conference on Geoinformatics 1–5. https://doi.org/10.1109/GeoInformatics.2011.5981067

[dataset] Ying M, Zhang W, Yu H, Lu X, Feng J, Fan Y, Zhu Y, Chen D (2014) An overview of the China Meteorological Administration tropical cyclone database. J Atmos Ocean Technol 31(2):287–301. https://doi.org/10.1175/jtech-d-12-00119.1.

Yuan S, Wang C, Mu B, Zhou F, Duan W (2021) Typhoon intensity forecasting based on LSTM Using the rolling forecast method. Algorithms 14(3):83. https://doi.org/10.3390/a14030083

Zhang D, Zhang Y, Hu T, Xie B, Xu J (2014) A Comparison of HY-2 and QuikSCAT Vector Wind Products for Tropical Cyclone Track and Intensity Development Monitoring. IEEE Geoscience and Remote Sens Lett 11(8):1365–1369. https://doi.org/10.1109/LGRS.2013.2293496

Zhou Y, Matyas CJ (2021) Regionalization of precipitation associated with tropical cyclones using spatial metrics and satellite precipitation. GIScience & Remote Sens 58(4):542–561. https://doi.org/10.1080/15481603.2021.1908675

Zong Y, Chen X (1999) Typhoon hazards in the Shanghai area. Disasters 23(1):66–80. https://doi.org/10.1111/1467-7717.00105

Figure captions

Fig.1: Correlation coefficients between landfalling Tropical Cyclone (TC) classification indices and (a) affected population, (b) agricultural disaster area, (c) number of collapsed buildings, (d) number of deaths, and (e) direct economic loss. Correlation coefficients were calculated from the data of 45 TCs from 2000–2014.

Fig.2: Correlation coefficients between landfalling TC classification indices and (a) affected population, (b) agricultural disaster area, (c) number of collapsed buildings, (d) number of deaths, and (e) direct economic loss. Correlation coefficients were calculated from the data of 17 TCs from 2015–2019.

Fig.3: TC track length, maximum wind speed, and lifetime for (a) Nida (2016), (b) Khanun (2017), and (c) Mangkhut (2018).

Fig.4: The daily precipitation changes of TC Nida in 2016 for (a) July 30, (b) July 31, (c) August 1, and (d) August 2. The red dot indicates cyclone location at 06:00 UTC of each day. The black line shows cyclone track. Colors indicate cumulative daily precipitation.

Fig.5: Average landfalling TC classification indices for three TCs.

Fig.6: Average output from the projection pursuit dynamic cluster (PPDC) model for three TCs.

Fig.7: The daily precipitation changes of TC Khanun in 2017 for (a) October 12, (b) October 13, (c) October 14, and (d) October 15. The red dot indicates cyclone location at 06:00 UTC of each day. The black line shows cyclone track. Colors indicate cumulative daily precipitation.

Fig.8: The daily precipitation changes of TC Mangkhut in 2018 for (a) September 7, (b) September 9, (c) September 11, (d) September 13, (e) September 15, and (f) September 16. The red dot indicates cyclone location at 06:00 UTC of each day. The black line shows cyclone track. Colors indicate cumulative daily precipitation.

Figures

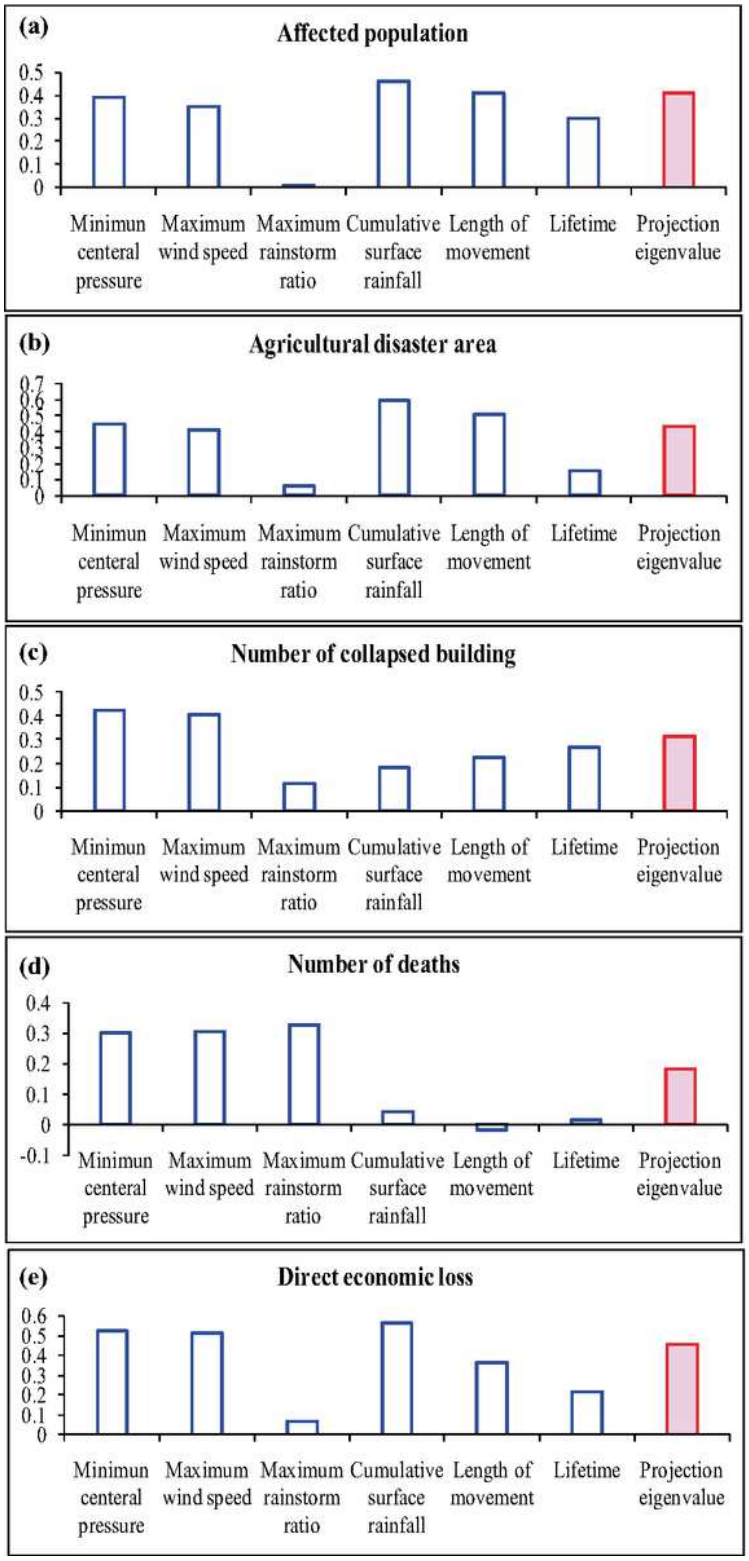

Figure 1

Correlation coefficients between landfalling Tropical Cyclone (TC) classification indices and (a) affected population, (b) agricultural disaster area, (c) number of collapsed buildings, (d) number of deaths, and (e) direct economic loss. Correlation coefficients were calculated from the data of 45 TCs from 2000–2014.

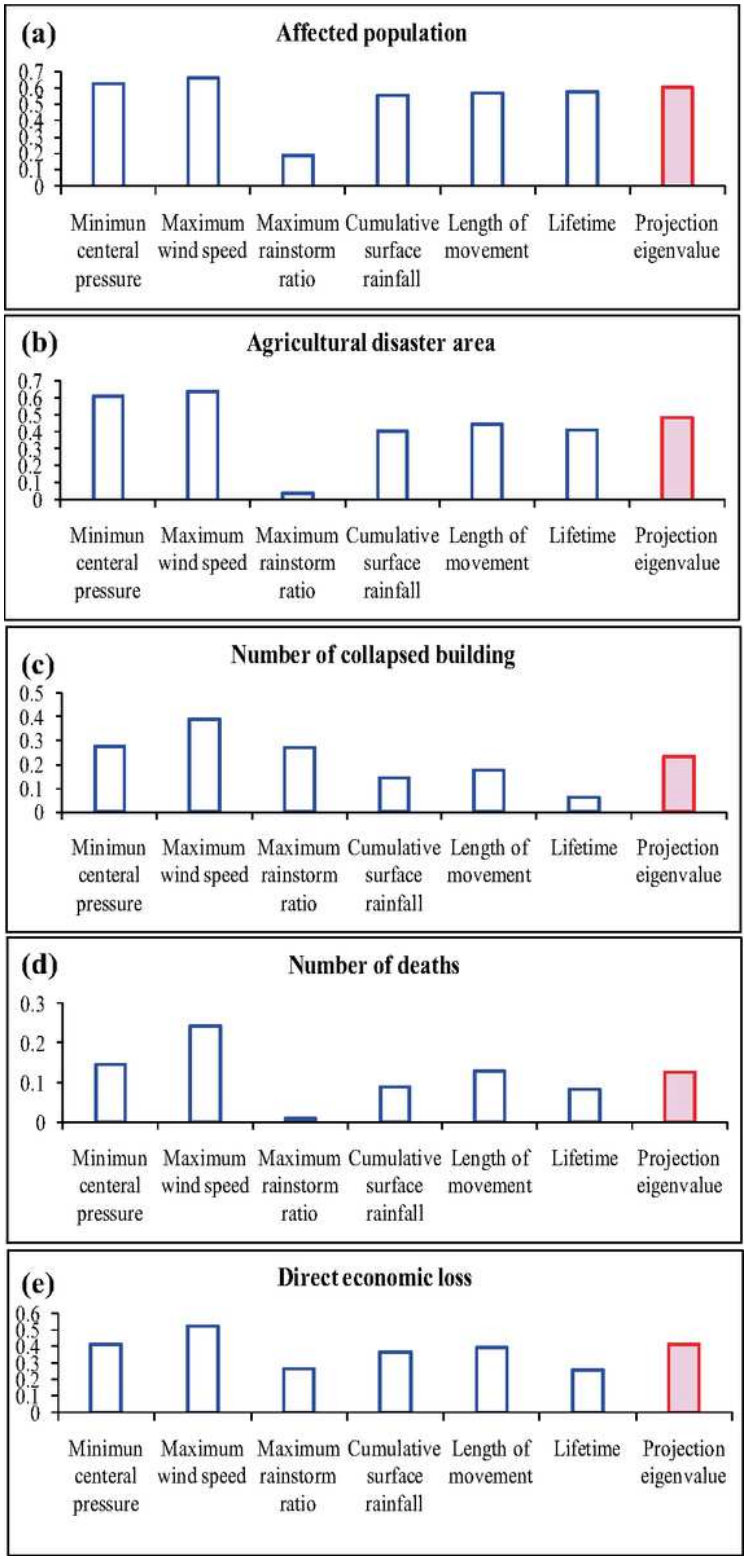

Figure 2

Correlation coefficients between landfalling TC classification indices and (a) affected population, (b) agricultural disaster area, (c) number of collapsed buildings, (d) number of deaths, and (e) direct economic loss. Correlation coefficients were calculated from the data of 17 TCs from 2015–2019.

Figure 3

TC track length, maximum wind speed, and lifetime for (a) Nida (2016), (b) Khanun (2017), and (c) Mangkhut (2018).

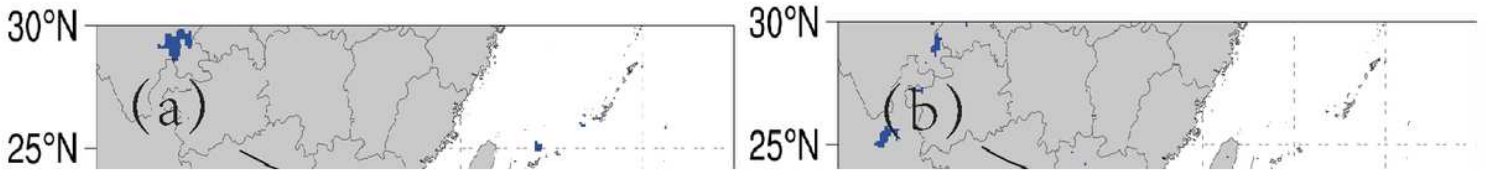

Figure 4

The daily precipitation changes of TC Nida in 2016 for (a) July 30, (b) July 31, (c) August 1, and (d) August 2. The red dot indicates cyclone location at 06:00 UTC of each day. The black line shows cyclone track. Colors indicate cumulative daily precipitation.

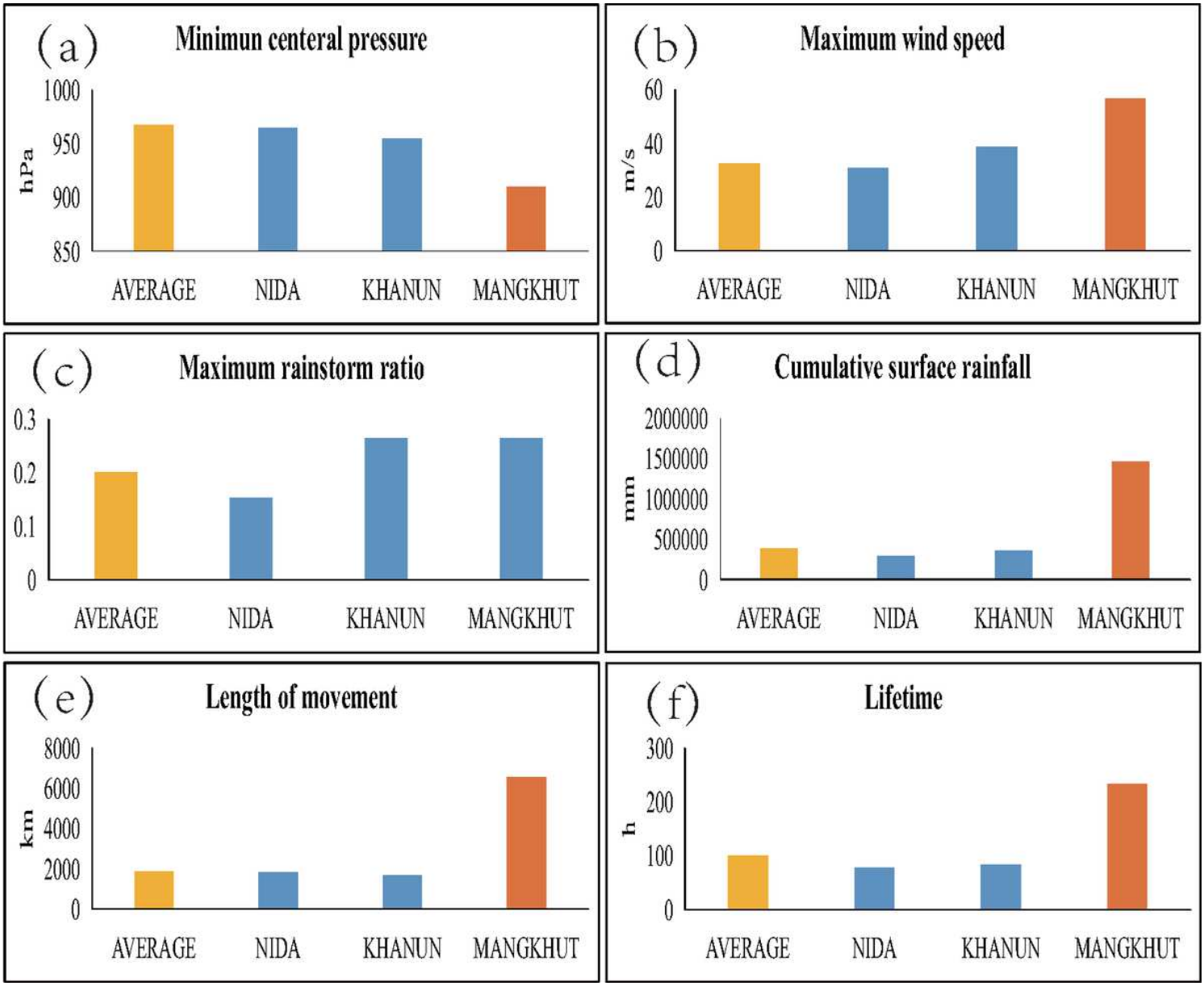

Figure 5

Average landfalling TC classification indices for three TCs.

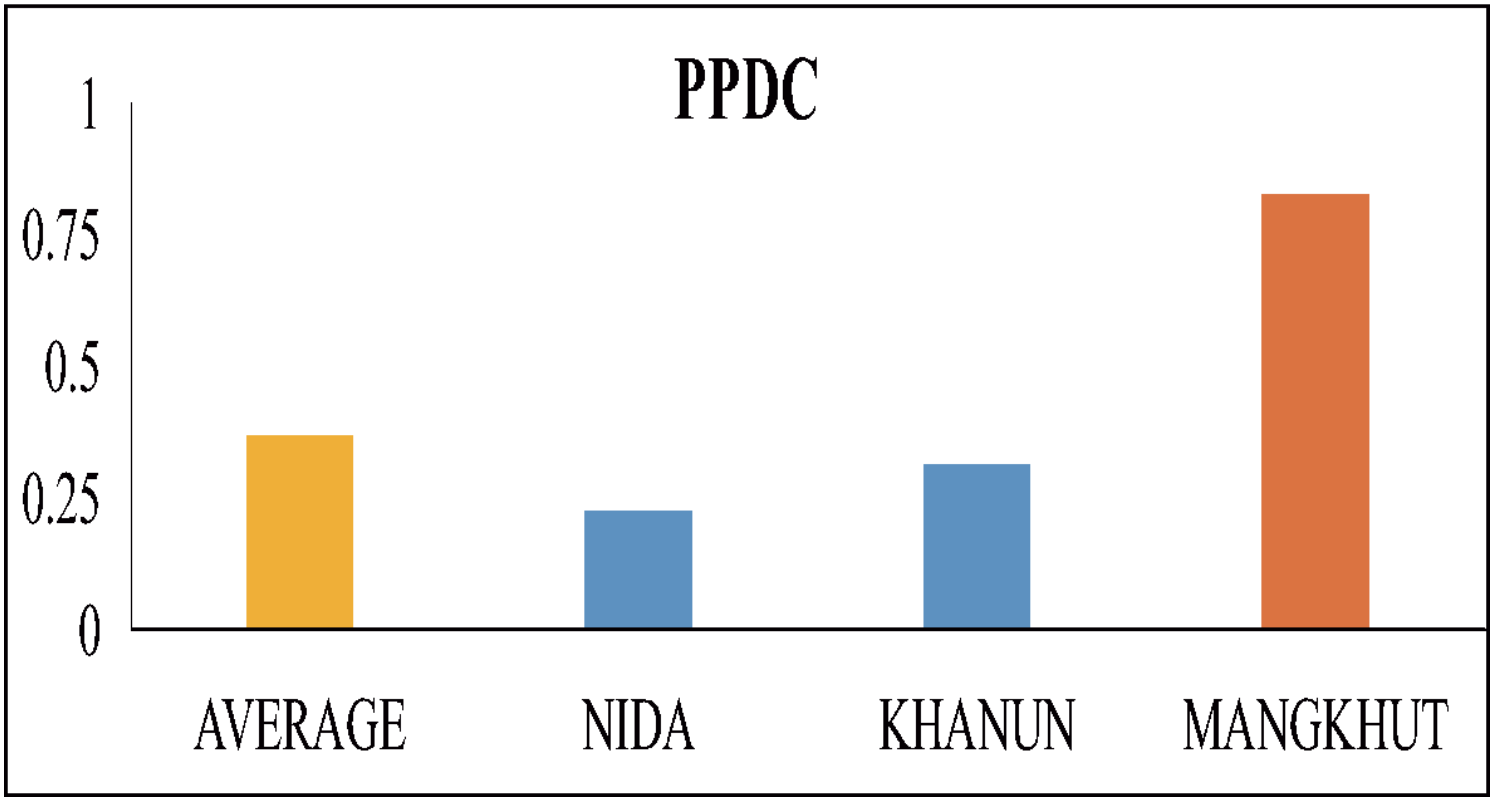

Figure 6

Average output from the projection pursuit dynamic cluster (PPDC) model for three TCs.

Figure 7

The daily precipitation changes of TC Khanun in 2017 for (a) October 12, (b) October 13, (c) October 14, and (d) October 15. The red dot indicates cyclone location at 06:00 UTC of each day. The black line shows cyclone track. Colors indicate cumulative daily precipitation.

Figure 8

The daily precipitation changes of TC Mangkhut in 2018 for (a) September 7, (b) September 9, (c) September 11, (d) September 13, (e) September 15, and (f) September 16. The red dot indicates cyclone location at 06:00 UTC of each day. The black line shows cyclone track. Colors indicate cumulative daily precipitation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx