High-Resolution Hazard Assessment for Tropical Cyclone-Induced Wind and Precipitation: An Analytical Framework and Application

Jiting Tang 1,2,3, Fuyu Hu 4, Yimeng Liu 1,2,3, Weiping Wang 5 and Saini Yang 1,2,3,5,*

1 Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education, Beijing Normal University, Beijing 100875, China
2 State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China
3 Academy of Disaster Reduction and Emergency Management, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
4 School of International Affairs and Public Administration, Ocean University of China, Qingdao 266100, China
5 School of National Safety and Emergency Management, Beijing Normal University, Beijing 100875, China
* Correspondence: yangsaini@bnu.edu.cn

Abstract: Intensified tropical cyclones (TCs) threaten the socioeconomic development of coastal cities. The coupling of strong wind and precipitation with the TC process usually amplifies the destructive effects of storms. Currently, an integrated analytical framework for TC hazard assessment at the city level that combines the joint statistical characteristics of multiple TC-induced hazards and local environmental features does not exist. In this study, we developed a novel hazard assessment framework with a high spatiotemporal resolution that includes a fine-tuned K-means algorithm for clustering TC tracks and a Copula model to depict the wind–precipitation joint probability distribution of different TC categories. High-resolution wind and precipitation data were used to conduct an empirical study in Shenzhen, a coastal megacity in Guangdong Province, China. The results show that the probabilities of TC-induced wind speed and precipitation exhibit significant spatial heterogeneity in Shenzhen, which can be explained by the characteristics of TC tracks and terrain environment factors. In general, the hazard intensity of TCs landing from the west side is higher than that from the east side, and the greatest TC intensity appears on the southeast coast of Shenzhen, implying that more disaster prevention efforts are needed. The proposed TC hazard assessment method provides a solid base for highly precise risk assessment at the city level.

Keywords: tropical cyclone; hazard assessment; high resolution;Copula theory; K-means clustering; Shenzhen

1. Introduction

This past decade has been marked by devastating extreme events, including Hurricane Harvey in 2017, Typhoon Lekima in 2019, and Typhoon Rai in 2021. Climate change has intensified tropical cyclones (TCs), posing a greater threat to life and property along coastal areas [1,2]. With the influence of the anthropogenic rise in greenhouse gases, cyclone track density, power dissipation, and cyclone genesis have shown robust increasing trends over the North Pacific [3,4]. In addition, the development of infrastructure in coastal areas, especially in developing countries, has accelerated. As a result, risks have increased with the urban infrastructures from more severe TCs [5]. TC hazard assessment at a high spatial resolution is an inevitable and fundamental step for risk assessment.

The demand from stakeholders for information on natural hazards has been highlighted in several reports, such as Global Assessment Reports on Disaster Risk Reduction and IPCC SREX. Comprehensive hazard assessment is essential to achieving the Millennium Development Goals, Sustainable Development Goals, and the Sendai framework.
Knowledge of natural hazards is key to developing preventative policies and taking risk-reduction measures. Such information is valuable to stakeholders from the (re)insurance, governance, and spatial planning sectors.

Many empirical models consider that TC hazards vary among different areas, provinces, cities, and even greater regions. However, the spatial or temporal resolution of previous studies has generally been rough, making it difficult to illustrate the hazards of a TC process at the city level. How do TC hazards vary within a city? What factors can explain the variation? In contrast to the extensive studies of inventory exposure to TCs with a high spatiotemporal resolution, these questions have not received adequate attention. High-resolution TC hazard assessment is urgently needed for refined risk analysis and emergency management in coastal cities. For a more accurate impact analysis of TCs, hazard assessments with a high temporal resolution are needed to depict the changes during a TC process. Additionally, hazard assessments with a high spatial resolution are also essential to match the heterogeneously distributed urban infrastructures in cities. Meanwhile, a single factor of TC, such as wind speed, cannot fully reflect the compound impacts of TCs, which usually include strong winds, rainfall, and flooding [6]. Therefore, the combined possibilities of the intensity and frequency of multiple hazards caused by TCs should be considered simultaneously. In addition, the location and direction of TCs affect the scope and distribution pattern of wind and rain in coastal cities [7,8], such that the hazards of different TC categories should be assessed separately.

To address the above issues, we developed a hazard assessment framework with a high spatiotemporal resolution to analyze regional TC hazards. In the framework, TC tracks are clustered by a fine-tuned K-means algorithm, and the wind–precipitation joint probability distribution of different TC categories is depicted by a Copula model. The proposed framework focuses on the short-term wind–rain joint occurrence probability during TC processes and enables subsequent fine-grained risk analysis. To the best of our knowledge, this is the first TC hazard assessment study with such a high temporal and spatial resolution. The outcomes can also be used to plan and schedule disaster preparedness and response operations in urban areas.

The paper is organized as follows. Section 2 reviews previous studies. Section 3 describes the proposed methodological framework, and Section 4 presents a hazard assessment application and results for Shenzhen, China. Thereafter, Sections 5 and 6 present the discussion and conclusions, respectively.

2. Literature Review

This section surveys an extensive set of TC clustering techniques and TC hazard assessment methodologies.

2.1. TC Clustering Techniques

In recent years, many clustering algorithms have been used to categorize TCs based on their features, such as TC tracks, forms, lengths, or intensities, from TC databases, among which tracks (longitudes and latitudes) have been the most widely used. The key for a clustering algorithm is clustering the data into several groups based on some similarity measures such that the total variance among the groups is minimized. The K-means algorithm is a straightforward and widely used partitioning method that seeks to assign each track to one of K groups [9–11]. However, difficulties arise because of the differentiated TC evolution processes, which result in very different point shapes and lengths. The typical K-means algorithm is not suitable for TC tracks of different lengths, and we show that this is a serious shortcoming for TCs. Hu et al. interpolated different tracks to obtain the same length of vectors [12], which added errors. Tian et al. detected interdecadal changes in the genesis of TCs in the western North Pacific based on pHash and K-means cluster analysis [13]. Chand et al. [14] and Samuel et al. [15] used the curve-clustering approach to classify TC tracks. Camargo et al. developed a special probabilistic clustering algorithm that relies on a regression mixture model to cluster TC tracks [16,17]. These techniques
provide a mixture of polynomial regression functions (curves) to fit the geographical shapes of trajectories. The Hausdorff distance is a simple dissimilarity measure and is widely used for comparing point sets [18] and image identifying [19]. The combination of the K-means algorithm and the Hausdorff distance can be a new way to cluster the irregular point sets.

2.2. TC Hazard Assessment

There are two main categories of TC hazard assessment methodologies. The first category consists of comprehensive index constructions, including the weighting, normalization, and mechanism-based methods. The weighting method gives weights to different hazard factors according to their importance, occurrence probability, or hazard intensity [20]. The normalization method normalizes a variety of indices into a single index, which is convenient for comprehensive probability analysis [21,22]. However, in these methods, the relationship among factors is characterized as linear, while it is nonlinear for actual TCs. Some researchers have constructed comprehensive indices based on the physical mechanisms or economic characteristics of TCs, such as Accumulated Cyclone Energy [23], Integrated Kinetic Energy [24], and Carvill Hurricane Index [25]. These comprehensive indices mainly focus on winds and fail to consider other TC-induced hazards (e.g., precipitation).

The other category consists of joint probability models, which construct a joint probability distribution function based on multi-hazard correlations. The Copula theory uses marginal distributions to form a joint distribution. It can solve nonlinear, nonnormal, non-symmetric, and long-tailed problems. In recent years, this theory has been widely used in multivariate analysis in the fields of extreme value theory [26], financial risk [27], wireless communication [28], drought [29,30], and flood disasters [31,32]. Some researchers have previously constructed the TC-induced wind–rain joint probability distribution function based on Copula theory. However, the data are mainly the maximum or accumulated value of annual statistics [33] or TC events [6,34] from a single meteorological station. Recent studies have used meteorological data from multiple stations to depict the regional difference of hazard intensity [35,36]. These studies have indicated that Copula theory can be an effective tool for TC hazard assessment.

3. Methodological Framework

In this paper, a novel TC hazard assessment framework is proposed and divided into three steps: determination of hazard datasets, analysis of hazard probability, and assessment of regional hazards. The analytical framework of this work is shown in Figure 1.
3.1. Hazard Datasets

3.1.1. Disaster Thresholds

The proposed framework assesses wind and rain, the two representative TC-induced hazards. Wind and rain have a significant influence on society, as they commonly lead to storm surges, floods, and debris flows. Our goal is to assess the comprehensive hazard intensity of short-term wind or rain exceeding certain thresholds during TC processes. Therefore, we set the distance threshold from the location to the TC center to determine if the wind and rain are induced by the TC and the intensity thresholds to determine if a hazard of a certain intensity will cause damage.

The thresholds used in this study include the TC impact distance threshold, wind-induced disaster threshold, and rain-induced disaster threshold. (1) TCs in the western Pacific Ocean are nearly circular circulation systems, with a typical radius of approximately 200–500 km [37]. Previous studies have shown that rainfall within 5° in longitude-latitude (approximately 500 km) from the TC center can be considered TC-induced rainfall [38], which is also consistent with on-site evidence. Therefore, we used a circle with a radius of 500 km from the TC center as the TC impact distance threshold. (2) According to the national standard of Grade of tropical cyclones https://www.codeofchina.com/standard/GBT19201-2006.html (accessed on 23 October 2022), a tropical depression is defined when the maximum 2 min mean wind speed near the TC center exceeds 10.8 m/s. Therefore, we set 10.8 m/s as the wind-induced disaster threshold. (3) According to the national standard of Grade of precipitation https://www.codeofchina.com/standard/GBT28592-2012.html (accessed on 23 October 2022), the precipitation was divided into seven levels. In this study, we chose the precipitation reaching the storm level (50 mm/24 h) as the rain-induced disaster threshold.

3.1.2. The Fine-Tuned K-Means Algorithm

In a general K-means algorithm, the data are mapped into points in a Euclidean space and clustered into several classes. These classes can be natural groups of variables, data points, or objects similar to one another in terms of some similarity measures. The clustering performance is evaluated by using the sum of squares for the error from each data point to the clustering center, as shown in Equation (1).

$$S_T = \sum_{i=1}^{K} \sum_{e_j \in P_i} ||e_j - c_i||^2$$  \hspace{1cm} (1)

Here, $K$ is the number of classes; $c_i$ is the clustering center of Class $i$; $P_i$ is the dataset for Class $i$; $e_j$ is the location attribute of TC track $j$ (the latitudes and longitudes of TC centers); and $||e_j - c_i||$ is the Euclidean distance from $e_j$ to $c_i$.

The iterative calculation process of K-means is as follows:

1. Select $K$ TC tracks as the initial clustering center.
2. Calculate the Euclidean distances from all the tracks to each clustering center and assign them to the nearest class.
3. Calculate the mean value of all data in each class as the new clustering center.
4. End if $S_T$ converges or the number of iterations reaches the preset maximum; otherwise, return to Step (2).

K-means cannot accommodate tracks of different lengths [37]. To address this problem, we fine-tuned the distance calculation in the algorithm by changing the Euclidean distance of two points to the Hausdorff distance of two point sets [38]. The Hausdorff distance is defined as follows:

$$H(A, B) = \max\{h(A, B), h(B, A)\}$$  \hspace{1cm} (2)

Here,

$$h(A, B) = \max_{a \in A} \min_{b \in B} ||a - b||$$  \hspace{1cm} (3)
\[ h(B, A) = \max_{b \in B} \min_{a \in A} \|b - a\| \quad (4) \]

\(H(A, B)\) is the bidirectional Hausdorff distance between point sets \(A\) and \(B\), \(h(A, B)\) is the unidirectional Hausdorff distance from \(A\) to \(B\), and correspondingly, \(h(B, A)\) is the Hausdorff distance from \(B\) to \(A\). \(H(A, B)\) is larger than \(h(A, B)\) and \(h(B, A)\), which measures the maximum mismatch between two point sets. In this study, the K-means clustering algorithm was reconstructed using the Hausdorff distance.

The fine-tuned K-means algorithm used in this study to fit the geographical shape of the trajectories allows the clustering to be posed and accommodates TC tracks of different lengths. The filtered TC tracks might be composed of the longitudes and latitudes of TC centers with different numbers, that is, have different lengths and shapes. The proposed method is more objective and simpler than the mixture K-means method used in previous studies.

We determined the hazard datasets in three steps. First, we filtered out the historical TC tracks by using the TC impact distance threshold. Second, we split the TC tracks according to the fine-tuned K-means clustering algorithm. Third, we filtered the meteorological data by using the wind-induced disaster threshold and rain-induced disaster threshold. Thus, the subsequent hazard analysis was carried out based on the subsets of wind speed and precipitation during the impact of different TC classes.

### 3.2. Analysis of Hazard Probability

The hazards for different TC classes were analyzed during two stages: single hazard of wind or rain and coupled hazards of wind–rain. Single hazard analysis constructs the probability density function (PDF), which indicates the probability of a hazard with different intensities. The probability distribution of a single hazard is the basis of multi-hazard joint probability analysis, commonly known as the marginal distribution. Then, the cumulative distribution function (CDF) of coupled hazards can be further generated based on the two-dimensional (2D) Copula functions.

Copulas are a family of functions that construct the joint distribution of two or more random variables with an unidentified dependence among the variables \([39,40]\). The most widely used Copula functions include two categories: Elliptic Copulas and Archimedean Copulas. Elliptic Copulas mainly include Gaussian Copula and t Copula, which are suitable when the marginal distribution obeys the standard normal distribution or Student’s T distribution. Archimedean Copulas are obtained by constructing different generator functions, including Clayton Copula, Frank Copula, Gumbel Copula, and Joe Copula (Table 1).

**Table 1.** Commonly used 2D Archimedean Copulas.

<table>
<thead>
<tr>
<th>Copulas</th>
<th>(C(u; v; \theta))</th>
<th>Limiting Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton Copula</td>
<td>[\max{u^{-\theta} + v^{-\theta} - 1; 0}]^{-1/\theta}]</td>
<td>(\theta \in [-1, \infty])</td>
</tr>
<tr>
<td>Frank Copula</td>
<td>[-\frac{1}{\theta} \log \left[ 1 + \frac{\exp(-\theta u) - 1}{\exp(-\theta v) - 1}\right]]</td>
<td>(\theta \neq 0)</td>
</tr>
<tr>
<td>Gumbel Copula</td>
<td>[-(-\log(u))^{\theta} + (-\log(v))^{\theta}]^{1/\theta}]</td>
<td>(\theta \in [1, \infty])</td>
</tr>
<tr>
<td>Joe Copula</td>
<td>[1 - \left((1 - u)^{\theta} + (1 - v)^{\theta} - (1 - u)^{\theta} (1 - v)^{\theta}\right)^{1/\theta}]</td>
<td>(\theta \in [1, \infty])</td>
</tr>
</tbody>
</table>

Here, \(u\) and \(v\) are the marginal distributions of two variables, \(\theta\) is the Copula function parameter, and \(C(u; v; \theta)\) is the 2D Archimedean Copula.

We constructed a multi-hazard joint probability model based on 2D Copula theory, which was divided into the following four steps:

1. Determine the marginal distribution of the wind speed and precipitation;
2. Measure the correlation among the hazards;
(3) Estimate the parameters of Copula functions that may be applicable;
(4) Select the appropriate Copula function to fit the joint distribution. Then, the probability of multiple hazards with different intensities can be calculated.

To choose the marginal distribution and joint distribution in Steps (1) and (4), we first determined if the variable was subject to the fitted distribution type based on the Kolmogorov-Smirnov (K-S) test. Then, we selected the best fitting function based on the Akaike information criterion (AIC) minimum, Bayesian information criteria (BIC) minimum, and log-likelihood (LogLik) maximum principle. The calculation formulas are shown in Equations (5)–(8).

\[
AIC = 2k - 2\ln(L(\theta)) \\
BIC = k\ln(n) - 2\ln(L(\theta)) \\
L(\theta) = \prod_{i=1}^{m} p(y_i; \theta) \\
\text{LogLik}(\theta) = \ln L(\theta)
\]

where \( y_i \) is the data sample, and \( p(y_i; \theta) \) is the PDF of the fitting function.

According to the single-hazard \( X \) occurrence probability \( P(X \geq x) = 1 - F(x) \), we defined two kinds of double-hazard joint occurrence probability. \( P_{X_1, X_2} \) indicates the probability of at least one hazard reaching a specific intensity, while \( P_{\cap X_1, X_2} \) indicates the probability of two hazards reaching a specific intensity simultaneously. The calculation formulas are shown in Equations (9) and (10).

\[
P_{\cup X_1, X_2} = P(X_1 \geq x_1 \text{ or } X_2 \geq x_2) = 1 - F(x_1, x_2) \\
P_{\cap X_1, X_2} = P(X_1 \geq x_1 \text{ and } X_2 \geq x_2) = 1 - F_{X_1}(x_1) - F_{X_2}(x_2) + F(x_1, x_2)
\]

where \( X_1, X_2 \) are random variables, their respective CDFs are \( F_{X_1}(x_1) \) and \( F_{X_2}(x_2) \), and the joint CDF of the Copula function is \( F(x_1, x_2) \).

In this study, we fitted the marginal distributions of the hourly wind and precipitation data separately and then fitted the joint distribution of the coupled wind and precipitation by Copula theory. Based on the single or coupled hazard fitting results of different subareas, we quantitatively depicted the characteristics of regional hazards.

4. Application and Results

4.1. Study Area and Data Source

Shenzhen, which is an economically developed and densely populated coastal megalopolis in China, has an area of 1997.47 km², a GDP of 2.69 trillion yuan and a permanent population of 13.44 million as of 2019 [1]. Shenzhen faces the South China Sea and the Pearl River. It has a subtropical oceanic climate. During summer, Shenzhen is vulnerable to frequent TCs from the western Pacific Ocean. A major reason for the high TC risk in Shenzhen is that the city is located in the TC landfall area of the western Pacific Ocean. The city is vulnerable to sudden and heavy rainfall caused by TCs, which can lead to serious communication system, and buildings has increased.

Four categories of data were used in this study: (1) the meteorological data of Shenzhen provided by the National Climate Center (NCC), including 10 m wind speed and precipitation data between 1 January 2008 and 31 December 2018 in 39 grid points with resolutions of 0.0625° × 0.0625° and 1 h, respectively [2]; (2) the TC track dataset provided by the NCC, which includes the hourly historical TCs that landed in China from 2008 to 2018. Each record includes the TC ID, year, month, day, time, longitude, and latitude of the TC centers [3]; (3) the 90 m digital elevation model (DEM) of Shenzhen from the NASA SRTM3 dataset [4]; and (4) the
data of the main rivers and water bodies in Shenzhen from the Resource and Environment Science and Data Center https://www.resdc.cn/DOI/DOI.aspx?DOIID=44 (accessed on 23 October 2022). The meteorological data and TC tracks were used to determine the hazard datasets, and the distribution of the DEM and river/water bodies served as a reference for subsequent causative analysis. Figure 2 shows the geographical location of Shenzhen and the distribution of multisource meteorological environment data.

Figure 2. Case study area: Shenzhen. (a) The geographical location; (b) The distribution of multisource environmental data.

Compared with the annual or TC events in previous studies, the temporal resolution of the hourly data used in this study is considerably enhanced. In meteorological research, one hour is typically acknowledged as a high-precision time unit. When compared to historical weather station observational data utilized for regional hazard studies, the spatial resolution of \(0.0625^\circ\) is significantly more accurate.

4.2. Preprocessing of Hazard Datasets

We filtered out the datasets according to the proposed K-means-based TC track clustering algorithm and three thresholds referred to in Step 1 in Figure 1.

First, for each grid point in Shenzhen, we calculated the distance \(d\) from the TC center in all the track records (Figure 2a) to the grid. The grid point was classified as being impacted by TCs at that moment if \(d \leq 500\) km. From 2008–2018, 74 TCs affected Shenzhen, with an average duration of 47.89 h. Then, we clustered the TC tracks affecting Shenzhen based on the fine-tuned K-means algorithm. We clustered the TC tracks into two classes, and the maximum number of iterations was 100. Figure 3 shows that the two TC classes can be roughly interpreted as landing from the west side of Shenzhen (Class 1) or the east side (Class 2). Furthermore, we retrieved the hourly wind speed and precipitation data for a grid point if the 24 h cumulative precipitation was greater than 50 mm or the wind speed was greater than 10.8 m/s. Thus, we created TC-induced wind speed and precipitation datasets for 39 grid points and 2 TC classes. Approximately 80–200 samples were recorded per grid point for each TC class. Next, we analyzed the TC hazards from the single variable (wind or precipitation) and coupled variables (joint wind–precipitation).

4.3. Hazard Analysis

In this section, we conducted the analysis of hazard probability referred to Step 2 in Figure 1. First, we found a fit for the CDF of observed wind and precipitation data. Then, we specified the Copula method to fit the joint PDFs and CDFs. Later, we appraised the performance of the hazard assessment.
wind speed was greater than 10.8 m/s. Thus, we created TC-induced wind speed and precipitation datasets for 39 grid points and 2 TC classes. Approximately 80–200 samples were recorded per grid point for each TC class. Next, we analyzed the TC hazards from the single variable for precipitation datasets. We randomly picked one grid point (denoted by \( \mathcal{A} \)) as an example to analyze a single hazard. There were 105 records of grid point \( \mathcal{A} \) for TC Class 1 and 143 records for TC Class 2. Referring to existing research results \([33–35]\), we selected the Gumbel, Weibull, Gamma, and lognormal distributions as the theoretical distribution functions to fit the hourly wind and precipitation data separately. We utilized the maximum likelihood estimate (MLE) approach to estimate the parameters. According to the AIC, BIC, and LogLik values, we found that for two TC classes, Gamma distribution was the best marginal distribution to fit the wind speed data (Figure 4a,b), whereas the lognormal distribution was suitable for fitting the precipitation data (Figure 4c,d).

For TC Class 1, the fitting Gamma CDF \( F_1(x_1) \) of the wind speed data in \( \mathcal{A} \) is

\[
F_1(x_1) = \int_0^{x_1} \frac{0.50^{2.16} x_1^{1.16} \exp(-0.50x_1)}{\Gamma(2.16)} dx_1
\tag{11}
\]

Here, \( \Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \) is the Gamma function with recursion, i.e., \( \Gamma(x+1) = x\Gamma(x) \).

The fitting lognormal CDF \( F_2(x_2) \) of the precipitation in \( \mathcal{A} \) is

\[
F_2(x_2) = \frac{1}{2} + \frac{1}{2} \text{erf} \left[ \frac{\ln(x_2) - 1.13}{1.35 \sqrt{2}} \right]
\tag{12}
\]

Here, \( \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \) is the Gaussian error function.

For TC Class 2, the fitting Gamma CDF \( F_1(x'_1) \) of the wind speed data in \( \mathcal{A} \) is

\[
F_1(x'_1) = \int_0^{x'_1} \frac{0.60^{3.02} x'_1^{2.02} \exp(-0.60x'_1)}{\Gamma(3.02)} dx'_1
\tag{13}
\]

The fitting lognormal CDF \( F_2(x'_2) \) of the precipitation in \( \mathcal{A} \) is

\[
F_2(x'_2) = \frac{1}{2} + \frac{1}{2} \text{erf} \left[ \frac{\ln(x'_2) - 1.07}{1.40 \sqrt{2}} \right]
\tag{14}
\]
Based on Equations (11)–(14), we calculated the corresponding wind speed and precipitation intensities with different probabilities (Table 2) and the probabilities of different wind speeds (Table 3) and precipitations (Table 4) for the two TC classes.

Table 2. The wind speeds and precipitations with different probabilities in \( \mathcal{A} \).

<table>
<thead>
<tr>
<th>TC Classes</th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>Wind Speed (m/s)</td>
<td>Precipitation (mm/h)</td>
</tr>
<tr>
<td>50%</td>
<td>3.67</td>
<td>3.11</td>
</tr>
<tr>
<td>20%</td>
<td>6.41</td>
<td>9.72</td>
</tr>
<tr>
<td>10%</td>
<td>8.25</td>
<td>17.64</td>
</tr>
<tr>
<td>5%</td>
<td>10.00</td>
<td>28.86</td>
</tr>
<tr>
<td>1%</td>
<td>13.87</td>
<td>72.64</td>
</tr>
<tr>
<td>0.5%</td>
<td>15.48</td>
<td>101.84</td>
</tr>
</tbody>
</table>

In addition, we found that the gamma distribution and lognormal distribution are also good fits for the TC-induced wind speed and precipitation in other grid points in Shenzhen.

4.3.2. Coupled Hazard Analysis

Using grid point \( \mathcal{A} \) as an example, the bivariate distribution of hourly wind speed and precipitation is given in Figure 5. The blue line in Figure 5 is the linear regression fit, and the shading along the lines is the confidence interval (95%). There are some points with wind speeds less than 10 m/s and 24 h cumulative precipitation greater than 50 mm. This is consistent with the facts—a typhoon may bring rainfall for several hours, but it is not always windy. At a high temporal resolution, such as one hour, the TC-induced wind and precipitation are not synchronous. The scattered distribution of the wind speed and

Figure 4. The fitting Cumulative Distribution Function of the hourly wind speed (a,b) and precipitation (c,d) in grid point \( \mathcal{A} \) for TC Class 1 (a,c) and Class 2 (b,d).

In addition, we found that the gamma distribution and lognormal distribution are also good fits for the TC-induced wind speed and precipitation in other grid points in Shenzhen.
precipitation in Figure 5 caught our attention. For TC Class 1, the wind–rain correlation was $P_{\text{west}} = 0.13$, passing the significance test of 0.05. The distributions of the hourly wind and rain data were asymmetric and nonnormal, as indicated by the kurtosis values of the wind speed and precipitation, which were 4.65 and 2.17, and their skewness values, which were 2.07 and 1.71, respectively. For TC Class 2, the wind–rain correlation was $P_{\text{east}} = 0.17$. The kurtosis values of the wind speed and precipitation were 6.99 and 0.84, and their skewness values were 1.99 and 1.29, respectively. Therefore, we selected Archimedean Copulas to fit the joint probability distribution and estimated the parameters by MLE. We found that the Joe Copula outperformed the other three Copulas in fitting the joint probability of the TC-Class-1-induced wind speed and precipitation in grid point $A$, but the Clayton Copula performed better for fitting the TC-Class-2-induced wind and precipitation (Table 5). Figure 6 shows the Copula fitting results in grid point $A$.

Table 3. The probabilities of different wind speeds in $A$.

<table>
<thead>
<tr>
<th>Grade of TCs</th>
<th>Wind Speed (m/s)</th>
<th>Wind Scale</th>
<th>Probability (Class 1)</th>
<th>Probability (Class 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tropical depression</td>
<td>10.8–17.1</td>
<td>6–7</td>
<td>0.036</td>
<td>0.045</td>
</tr>
<tr>
<td>Tropical storm</td>
<td>17.2–24.4</td>
<td>8–9</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>Severe tropical storm</td>
<td>24.5–32.6</td>
<td>10–11</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Typhoon</td>
<td>32.7–41.4</td>
<td>12–13</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Strong typhoon</td>
<td>41.5–50.9</td>
<td>14–15</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Super typhoon</td>
<td>$\geq$51.0</td>
<td>$\geq$16</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

/ indicates that the calculated value is too small, the same below.

Table 4. The probabilities of different precipitations in $A$.

<table>
<thead>
<tr>
<th>Grade of Precipitation</th>
<th>Precipitation (mm/h)</th>
<th>Probability (Class 1)</th>
<th>Probability (Class 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light rain</td>
<td>$\leq$2.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Moderate rain</td>
<td>2.6–8.0</td>
<td>0.564</td>
<td>0.544</td>
</tr>
<tr>
<td>Heavy rain</td>
<td>8.1–15.9</td>
<td>0.243</td>
<td>0.236</td>
</tr>
<tr>
<td>Torrential rain</td>
<td>16.0–49.9</td>
<td>0.113</td>
<td>0.112</td>
</tr>
<tr>
<td>Heavy downpour</td>
<td>50.0–99.9</td>
<td>0.020</td>
<td>0.021</td>
</tr>
<tr>
<td>Rainstorm</td>
<td>$\geq$100.0</td>
<td>0.005</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Table 5. Comparison of AIC, BIC, LogLik results for Copulas in $A$.

<table>
<thead>
<tr>
<th>TC Class</th>
<th>Copulas</th>
<th>Fitting $\theta$</th>
<th>AIC</th>
<th>BIC</th>
<th>LogLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Clayton</td>
<td>0.32</td>
<td>−6.96</td>
<td>−4.00</td>
<td>4.48</td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>0.88</td>
<td>−0.66</td>
<td>2.31</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>Gumbel</td>
<td>1.15</td>
<td>−5.45</td>
<td>−2.48</td>
<td>3.72</td>
</tr>
<tr>
<td></td>
<td>Joe</td>
<td>1.26</td>
<td>−7.85</td>
<td>−4.89</td>
<td>4.92</td>
</tr>
<tr>
<td>Class 2</td>
<td>Clayton</td>
<td>0.23</td>
<td>−1.02</td>
<td>−1.63</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>0.95</td>
<td>−0.64</td>
<td>2.02</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td>Gumbel</td>
<td>1.12</td>
<td>−1.00</td>
<td>−1.65</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td>Joe</td>
<td>1.18</td>
<td>−0.67</td>
<td>−1.98</td>
<td>1.34</td>
</tr>
</tbody>
</table>
the probability of occurrence of a rainstorm or tropical depression in one hour is \( P(0.05, 0.05) = 0.136 \).

Based on the fitting Copula and marginal distribution of two single hazards, we obtained the joint occurrence probability of double hazards in \( A \). When the TC lands in western Shenzhen (Class 1), the probability of the simultaneous occurrence of a rainstorm and a tropical depression in one hour during the TC process is \( P(0.05, 0.05) = 0.021 \), passing the significance test of 0.05. When the TC lands on the east side (Class 2), the probability of the simultaneous occurrence of a rainstorm and a tropical depression in one hour during the TC process is \( P(0.05, 0.05) = 0.017 \).

### Marginal Distribution

<table>
<thead>
<tr>
<th>Copula</th>
<th>AIC</th>
<th>BIC</th>
<th>LogLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joe</td>
<td>7.85</td>
<td>1.71</td>
<td>6.99</td>
</tr>
<tr>
<td>Frank</td>
<td>2.48</td>
<td>3.72</td>
<td>4.00</td>
</tr>
<tr>
<td>Clayton</td>
<td>2.07</td>
<td>1.71</td>
<td>2.17</td>
</tr>
<tr>
<td>Archimed</td>
<td>2.02</td>
<td>1.65</td>
<td>1.98</td>
</tr>
</tbody>
</table>

Based on the fitting results above, we calculated descriptive statistics, including the mean value, standard deviation, and skewness values. For TC Class 1, the mean of the wind speed and precipitation were 4.65 and 2.17, respectively, and their skewness values were 1.99 and 1.29, respectively. Therefore, we selected Archimed Copula to fit the joint probability distribution and estimated the parameters by MLE. We found that the Joe Copula outperformed the other three Copulas in fitting the data, with the best performance in fitting the hourly mean value, standard deviation, and skewness values.

### Figure 5

**Figure 5.** Bivariate distribution of wind speed and precipitation in grid point \( A \) for TC Class 1 (a) and Class 2 (b). The blue line is the linear regression fit, and the shading along the lines is the confidence interval (95%).

### Figure 6

**Figure 6.** The fitting results of the coupled hazards in \( A \). (a) Joe Copula PDF for TC Class 1; (b) Joe Copula CDF for TC Class 1; (c) Clayton Copula PDF for TC Class 2; (d) Clayton Copula CDF for TC Class 2.
storm and a tropical depression in one hour during the TC process is \( P_{\cap}(10.8,16.0) = 0.021 \), and the probability of occurrence of a rainstorm or tropical depression in one hour is \( P_{\cup}(10.8,16.0) = 0.136 \). When the TC lands on the east side (Class 2), \( P_{\cap}(10.8,16.0) = 0.017 \), \( P_{\cup}(10.8,16.0) = 0.125 \). That is, for grid point \( A \), the TCs landing on the west side of Shenzhen have a higher intensity than those landing on the east side. The results may help prepare for events that have not yet been observed in the past but nonetheless can be expected in the future.

### 4.3.3. The Performance of Hazard Assessment

Based on the fitting results above, we calculated descriptive statistics, including the mean value, standard deviation (\( s_d \)), coefficient of skewness (\( c_s \)), and their relative error (RE, in \% \([41] \), to demonstrate the performance of the hazard assessment (Table 6). Here, the descriptive statistics were calculated by Monte Carlo simulation, which consisted of three steps: (1) generating 5000 random numbers obeying the fitted Copula functions, (2) performing a computation of descriptive statistics using the random numbers, and (3) averaging the results across 10 repetitions to eliminate randomness.

<table>
<thead>
<tr>
<th>TC Classes</th>
<th>Statistics</th>
<th>Wind Speed</th>
<th>Precipitation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Single Factor Fitting</td>
<td>Coupled Factor Fitting</td>
</tr>
<tr>
<td>Class 1</td>
<td>Mean</td>
<td>4.32</td>
<td>4.31</td>
</tr>
<tr>
<td></td>
<td>( s_d )</td>
<td>2.94</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>( c_s )</td>
<td>1.36</td>
<td>1.49</td>
</tr>
<tr>
<td>Class 2</td>
<td>Mean</td>
<td>5.04</td>
<td>4.25</td>
</tr>
<tr>
<td></td>
<td>( s_d )</td>
<td>2.90</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>( c_s )</td>
<td>2.15</td>
<td>2.93</td>
</tr>
</tbody>
</table>

Table 6 shows the fitting results of the single and coupled hazards. The descriptive statistics of the single fitting factor were calculated from Equations (11)–(14). The descriptive statistics of the fitting coupled factors were calculated from \( C(u,v;\theta) \) of Joe Copula and Clayton Copula in Table 1; here, \( \theta \) is the bold fitting \( \theta \) in Table 5, and \( u, v \) are taken from Equations (11)–(14). The descriptive statistics of the empirical values were calculated from preprocessed hazard datasets in Section 4.2. Then, we compared the RE of the descriptive statistics between the empirical values and fitting values. Table 6 demonstrates that the hazard assessment performed well because the REs for more than half of the statistics were less than 30%. Comparing the wind speed and precipitation, we found that the fitting REs of the precipitation are larger than those of the wind speed because the wind speed values are mostly concentrated in a smaller range (approximately 0–14 m/s). While the precipitation span is larger (approximately 0–40 mm/h), the extreme precipitation values amplify the bias. Comparing the three statistics, we found that the REs of the mean are small, while the REs of \( s_d \) and \( c_s \) are large. Because the statistics of the fitting values were generated based on simulation data with randomness, the generated extreme values have a smaller impact on the mean value but a greater impact on \( s_d \) and \( c_s \).

It should be noted that the uncertainty of the meteorological gridded data, which are interpolated based on the observational data from the weather stations, affects the goodness of the fitting results.

### 4.4. The Spatial Heterogeneity of TC Hazards

Based on the analysis results of single and coupled hazard probabilities, we further assessed the regional hazards for specific infrastructure risks referred to in Step 3 in Figure 1. Using road traffic as an example, we analyzed the spatial distribution of TC hazards and identified significant heterogeneity. We set the TC disaster threshold for road traffic based
on the Technical Specification for Highway Travel Information Service as 10 mm/h for the precipitation and 8 m/s for the wind speed.

For the 39 grid points in Shenzhen, we used Gamma distribution to fit the wind speed and a lognormal distribution to fit the precipitation for the two TC classes. Then, we calculated the probability of wind speeds over 8 m/s as P(west_wind) and P(east_wind) and the probability of precipitation over 10 mm/h as P(west_rain) and P(east_rain). Figure 7 displays the spatial distribution.

Figure 7a,c demonstrate that the spatial distributions of P(west_wind) and P(east_wind) have a similar pattern regardless of the TC landing location. The east is larger than the west, and the south is larger than the north because the water vapor and surface environment are distinct between the sea and the land, and TCs cause much stronger winds in coastal areas compared to inland areas. In Shenzhen, the TC-induced winds in the west have been significantly weakened due to the blockage of the Pearl River estuary and Hong Kong. The eastern region, particularly Dapeng District, lacks natural barriers and is frequently exposed to severe winds.

Most of western Shenzhen and some of eastern Shenzhen have a higher probability of experiencing rainfall over 10 mm/h when a TC arrives from the west (Figure 7b), whereas small areas of eastern Shenzhen experience rainfall exceeding 10 mm/h when a TC arrives from the east (Figure 7d). This higher probability is due to the terrain and TC tracks. Shenzhen is generally in low hilly areas with mild terraces in between, with a high elevation in the southeast and a low elevation in the northwest. When addressing TCs that are landing from the west, Shenzhen’s western plains and rivers promote significant rainfall, which gradually decreases until it hits gentle hills. The southeast mountains would obstruct the flow while facing TCs arriving from the east. As a result, the elevation causes the wind to be weaker and the rainfall to be heavier on the windward sides. Dense air flow lines at the top and sides of the mountains cause the wind to blow faster. As a result, in the
area east of Shenzhen, the likelihood of strong winds would decrease, while the likelihood of heavy rain would change from south to north.

In conclusion, there is a significant spatial difference in the probability of a single hazard affecting road traffic. The most widespread effects of TC-induced rainfall occurred when it arrived from the west. The spatial heterogeneity of TC hazards is influenced by multiple factors of the meteorological environment. These findings validate the necessity and illustrate the importance of high-resolution hazard assessments, especially for accurate urban disaster management.

Repeating the calculation process, we tested the fitting performance of Archimedean Copulas for 39 grid points in Shenzhen. Joe Copula and Clayton Copula fit 27 and 12 grid points, respectively, for TCs landing in western Shenzhen (Class 1). For TCs landing in eastern Shenzhen (Class 2), Clayton Copula had a good fitting performance on the combined probability of 28 grid points, and Joe Copula fit well for 11 grid points. Most of the hourly data in our dataset are on the low side, and Clayton Copula is sensitive to changes at the lower tail of the variables. Joe Copula can describe the intense wind and rain because it is more sensitive to changes in the upper tail of the variable. This higher sensitivity indicates that in some areas, a TC-induced rainfall process may be accompanied by extremely high winds and rainfall. We estimated the two double-hazard joint occurrence probabilities for the two TC classes as P(west_and), P(west_or), P(east_and), and P(east_or) based on the best fitted Copula for each grid. The spatial distributions are given in Figure 8.

![Figure 8. Spatial distribution of the probability of coupled hazards for road traffic. (a) P(west_and); (b) P(west_or); (c) P(east_and); (d) P(east_or).]
5. Discussion

The trend in hazard assessment is toward high resolutions for refined risk analysis and emergency management. Figure 5 shows the weak correlation between the hourly wind speed and precipitation during the impact of TCs—this is surprising, but it also makes sense. A typhoon may bring rainfall for several hours, but it is not always windy. At a high temporal resolution, such as one hour, the TC-induced wind and precipitation are not synchronous. In previous studies, a higher correlation has been found by using maximum or cumulative values based on TC events or annual statistics. This finding suggests that we need to consider the time scale differences of TC hazard assessment. Here, we compared the Pearson correlation coefficients of wind and rain over various time scales. Coefficients were used instead of Copulas because Copula fitting results may vary for different time scales. The Pearson correlation coefficient represents the degree of linear correlation between winds and rains. A larger correlation indicates a larger risk of strong winds and heavy rains. Based on the filtered hourly data reaching three thresholds in Section 4.2, we calculated the maximum wind speed and cumulative precipitation for 39 grid points at four time scales, including the hour, rainfall process, TC event, and year. Here, a rainfall process is a set of continuous records with hourly precipitation data greater than 0. We then obtained their Pearson coefficients.

Figure 9 shows a higher and more dispersed correlation when using the cumulative data (rainfall process, TC event, and year scales) than the hourly data. We speculate that the reason is that the temporal accumulation amplifies the compound hazard severity and local spatial heterogeneity. If we evaluate the TC process hazards based on the overall statistics, the higher correlation between multiple hazards would lead to a higher disaster risk. This finding suggests that we need to pay attention to the time scale of TC hazard analysis. For research on the impact of TCs on urban infrastructures that change rapidly, such as urban road transportation, the evolution of a TC process matters. Hazard analysis with a high temporal resolution can help improve the accuracy of determining how TC processes affect urban infrastructures.

From the hazard assessment results based on the high-resolution TC-induced wind rain data in Shenzhen, we found that the probability of the TC-induced wind speed and precipitation displayed a significant spatial heterogeneity, which can be explained by the TC tracks and terrain environment factors. The findings show that the TC hazard assessment results in different regions are customized. However, the proposed framework can be generalized.
6. Conclusions

This study developed a high-resolution framework to assess the regional hazards of wind and rain. This is a key step toward city-scale risk management and infrastructure resilience analysis during TC processes.

The proposed framework was applied to Shenzhen, China. Based on the hazard assessment results, we identified a significant spatial heterogeneity of TC hazards for urban traffic within the city and addressed the importance of high-resolution hazard assessment in urban disaster management. The southeast coast of Shenzhen is the most likely to be affected by TCs, and disaster prevention efforts should be increased. The hazard of TCs arriving from the west is greater than that of TCs arriving from the east, mostly due to the geographical location and terrain environment.

Our framework has the following three advantages: (1) the framework integrates single hazards and coupled hazards at the city level and is combined with joint statistical characteristics of TC hazards and local environmental features; (2) the fine-tuned K-means clustering algorithm provides a simple method to cluster TC tracks of various shapes and lengths and maintains the characteristics of the original data; and (3) the Copula-based joint probability model provides considerable flexibility for additional multiple hazards and can also be expanded to the joint probability distribution of more dimensions.

There are also certain limitations to this study. When studying joint probability based on Copula theory, we need to select the marginal distribution function and joint probability distribution. The outcomes of the joint probability distribution can vary with different models and fitting methods. Discussing the uncertainty of the results in Copula-based hazard analysis is worthwhile. Several extensions to the methodology are anticipated in our future work. We will analyze historical data based on a larger time scale and more dimensions to improve the fitting performance.

Our study can better reveal the interaction between a TC system and the physical environment of Shenzhen. The expected assessment outputs can be used for informed decision making and as a reference for disaster risk reduction.

Author Contributions: Conceptualization, S.Y.; Methodology, J.T. and W.W.; Supervision, S.Y.; Validation, Y.L.; Writing—original draft, J.T.; Writing—review and editing, F.H. and S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China, grant number 2018YFC1508903; and by the Scientific Research Start-up Funding of Ocean University of China, grant number 862201013145.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from the National Climate Center.

Acknowledgments: Many thanks to reviewers for their valuable suggestions.

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

References


10. Fang, W.; Zhang, H. Zonation and Scaling of Tropical Cyclone Hazards based on Spatial Clustering for Coastal China. Nat. Hazards 2021, 109, 1271–1295. [CrossRef]


34. Um, M.; Joo, K.; Nam, W.; Heo, J. A Comparative Study to Determine the Optimal Copula Model for the Wind Speed and Precipitation of Typhoons. Int. J. Clim. 2017, 37, 2051–2062. [CrossRef]


