A quantitative method for the similarity assessment of typhoon tracks

Yangchen Di¹ · Mingyue Lu² · Min Chen³,⁶,⁷ · Zhangjian Chen⁴ · Zaiyang Ma³,⁶,⁷ · Manzhu Yu⁵

Received: 21 August 2021 / Accepted: 28 December 2021 / Published online: 5 January 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
Typhoons are one of the most dangerous types of natural hazards; they are always developed in the western and southwestern Pacific Ocean and pose economic and human security threats to the Pacific Rim annually. Therefore, many scholars in related fields devote themselves to finding an effective way to analyze and forecast typhoon tracks to prevent disasters. Similarity analysis of typhoon tracks can provide great help for typhoon prediction. In this paper, a model for typhoon similarity analysis is proposed to effectively measure and quantify the similarity between two historical typhoon tracks based on the dynamic time warping algorithm, in which five typhoon elements—namely, longitude, latitude, central pressure, expanded Beaufort scale, and movement speed—are integrated to derive a final similarity percentage indicating the similarity level. At the end of this paper, case studies concerning historical typhoons and the ongoing Typhoon 202,106 In-Fa are also conducted to verify the validity and effectiveness of the proposed model. The results show that the proposed model can effectively provide a quantitative similarity of two typhoon tracks when functioning well on ongoing typhoons with a cutoff rule and supplying promising support for typhoon prediction simultaneously.

Keywords Typhoon tracks · Similarity assessment · Time series · Dynamic time warping (DTW)

✉ Mingyue Lu
lumingyue@nuist.edu.cn

¹ Changwang School of Honors, Nanjing University of Information Science & Technology, Nanjing 210044, China
² School of Geography Science, Nanjing University of Information Science & Technology, Nanjing 210044, China
³ Key Laboratory of Virtual Geographic Environment (Ministry of Education of PRC), Nanjing Normal University, Nanjing 210023, China
⁴ Zhejiang Academy of Surveying and Mapping, Hangzhou 311100, China
⁵ Department of Geography, The Pennsylvania State University, University Park, PA 16802, USA
⁶ State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China
⁷ Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
1 Introduction

A typhoon is a kind of tropical cyclone whose central wind speed can easily exceed 17.2 m/s. Typhoons develop on the surface of the western and southwestern Pacific Ocean and bring ample water vapor to the Pacific Rim, where they usually cause severe convective weather, such as heavy rains and storm surges. Thus, typhoons are an accepted devastating natural phenomenon that imperil human safety and economic development in coastal areas. For instance, in 2019, super Typhoon 201,909 Lekima caused 48 deaths or disappearances in China, with an economic loss of 40.71 billion yuan (Zhou 2021). Therefore, typhoons have become a hot topic because of their destructiveness, and many previous studies in this field have focused on typhoon track analysis for the sake of issuing early warnings of developing typhoons and reducing injuries and social loss.

Some studies on typhoon tracks pay close attention to track prediction, whose main methods originate from numerical prediction. Numerical prediction is also known as forecasting technology based on typhoon dynamics, and all analyses are centered on dynamic equations involving meteorological factors and physical processes. Hong et al. (2007) serialized the 500 hPa geopotential height field of the T106 numerical forecast product and applied the empirical orthogonal function to these data to optimize the parameters of the T106 model in parallel and globally. Zhang et al. (2017) attempted to incorporate large-scale climate change models, such as the Pacific Meridional Model, the Atlantic Meridional Model, and the North Atlantic Sea Surface Anomaly Temperature, as forecast factors in the statistical dynamics approach and improved the effects to a certain degree. Other research attempted to think outside the meteorological methods that basically generalize the generation and development principles of typhoons by various kinds of meteorological models. Such studies have also become an important part of typhoon analysis. In the 2010s, artificial intelligence (AI) achieved great success in typhoon forecasting. Hong et al. (2017) applied convolutional long short-term memory to a new typhoon path prediction model, which is helpful in processing spatial data over time. Gao et al. (2018) trained a long short-term memory (LSTM) neural network to establish a forecast model based on machine learning for the prediction of typhoon tracks. Jiang et al. (2018) proposed two algorithms based on machine learning neural networks that can be used with atmosphere-only typhoon forecast models to provide flow-dependent typhoon-induced sea surface temperature cooling for improving typhoon predictions. Alemany (2019) rasterized the region of the Atlantic Ocean by a resolution of 1° × 1°, coded these grids with fixed numbers into which points of typhoon paths were categorized, and then predicted the grid number by an LSTM model. With the rapid development of geographic information systems (GISs), typhoon analysis studies are inspired with new methods in GIS (Moayedi et al. 2019; Alizadeh et al. 2020), for geographic methods also have great potential in terms of enhancing the disaster management (Li et al. 2021a, b) and reproduction (Li et al. 2021a, b). Geographers have come up with the idea that the similarity between two typhoons is essential to typhoon research because the similarity analysis of historical typhoons is a great support and supplement to typhoon track prediction. According to the First Law of Geography—everything is related to everything else, but near things are more related than distant things—proposed by Tobler (1970), similar typhoons are highly likely to travel along the similar trajectory with similar natural properties. Therefore, similar tracks can exert a referential effect to predict tracks of ongoing typhoons. Zou et al. (2008) built a system based on the key point similarity method and buffer spatial analysis for fast forecasting of typhoon track trends within 24–48 h. This study determined the track similarity via the quadratic ratio of
the distance between corresponding key points and the radius of the circular buffer, and it was considered pioneering work in GIS-aided typhoon prediction. Hsu et al. (2014) developed a GIS-based decision support system, which integrated the real-time rainfall monitor and forecast information, for enhancing the emergency operations during typhoon attacks. Wu et al. (2014) first put forward Information-Matter-Element Model and established a GIS system for rainfall-based landslide hazard prediction. Bui et al. (2019) proposed and verified a new soft computing approach based on multivariate adaptive regression splines and particle swarm optimization for spatial prediction of flash flood susceptible areas at high-frequency tropical typhoon area. Wang et al. (2021) semi-quantified the risk assessment of storm surge under different typhoon intensities and performed the assessment in ArcGIS by the coupled model ADCIRC–SWAN and the Jelesnianski method.

However, methods utilizing GIS technology result in the following problems: On one hand, more stress is placed on geodesic elements such as longitudes and latitudes in the GIS field, which does not give enough consideration to other factors of typhoons; on the other hand, the processing of geographic data consumes more time and memory resources than normal data without geoinformation. For this reason, some scholars try to introduce concise mathematical models that lower the calculation power demand and quicken operation speed. She et al. (2014) abstracted the typhoon track as a plane curve and evaluated similarity in terms of the shapes of typhoon tracks and numerical indices of track points. Huang et al. (2019) vectorized typhoon data and applied principal component analysis and dynamic time warping (DTW) to portray similarity. However, these studies mostly lack explanations for the final results, and some of them are inclined to check the integrity of typhoon data and interpolate the missing data instead of performing similarity analysis. Thus, this paper tries to fill the gap of interpreting similarity intelligibly.

Looking at typhoons that have developed in the past decade, this paper serializes the original typhoon data, calculates the transitional similarity distance between two typhoons based on the DTW algorithm and, for the first time, proposes a new metric that quantifies the similarity degree by normalizing the corresponding similarity distance with the modified hyperbolic tangent function.

The other sections of this paper are organized as follows. Section 2 introduces the typhoon data used throughout and establishes the model for subsequent analysis. Section 3 presents the detailed procedure of the algorithm this paper utilizes based on DTW. Section 4 conducts experiments designed and discusses the results and parameters, with trying to filter similar typhoons of the currently ongoing Typhoon 202,106 In-Fa. Finally, Sect. 5 concludes the current research and proposes future research.

2 Typhoon data

2.1 Data introduction

The typhoon data from 2010 to 2020 utilized in this paper were obtained from the Real Time Release System of Typhoon Track (http://typhoon.zjwater.gov.cn/default.aspx) provided by the Ministry of Water Resources of Zhejiang Province, China. The data are transformed into a JSON-formatted file after data gathering and pretreatment for further analysis, which consists of 272 key-value pairs, as shown in Fig. 1.

Each key-value pair is composed of one key and one value and coded as “"key": value,” where the key is the index of the typhoon, and the value is a list of lists ordered by time,
each of which contains 6 elements denoted as “$\langle T_i, \text{Lon}_i, \text{Lat}_i, P_i, L_i, V_i \rangle$” representing the time elapsed after generation in hours (h), the longitude and the latitude of the central cyclone (unitless), the central pressure in hectopascals (hPa), the grade of wind speed in the expanded version of Beaufort scale (unitless), and the move speed of the typhoon in kilometers per hour (Km/h), respectively.

Considering that the original data were not organized with an equal time interval, which technically means that the interval of early observation records was 6h, while the interval of more current records, in recent five years, has been shortened to 3h usually or 1h when a typhoon is going to make landfall or intensify its wind speed, attached to each datum item as a time indicator ($T_i$) is, please notice that it is not involved in subsequent calculation, and this will be explained later when introducing the method.

### 2.2 Modeling and data preprocessing

To describe typhoons comprehensively, a more concise definition is given as

$$T = \{p_1, p_2, p_3, \ldots, p_n\}$$

(1)
where $T$ is the typhoon time series; $n$ is the length of the typhoon time series; and each $p_i$ ($i = 1, 2, 3, \ldots, n$) is a multidimensional vector consisting of the same number of observation values of a certain typhoon at different times, which makes $T$ a multidimensional time series.

Owing to the lack of observational measurements, some components in a $p_i$ marked as $-1$ are indicated as missing, and this $p_i$ is called a missing vector, while a vector whose components are all nonnegative is viewed as a complete vector. To lower the error between a series with missing vectors and one without, two criteria adopted for preprocessing are followed:

(a) If and only if the number of missing vectors in a certain typhoon time series is greater than or equal to 4, it is considered nugatory in future research and is to be discarded.
(b) Otherwise, a simple moving average is applied to the typhoon time series to make all vectors complete. This application can be formulated as

$$p_k = \frac{1}{n} \sum_{1 \leq |i| \leq 2} p_i$$

where $p_k$ is the missing vector; each $p_i$ is a complete vector that is adjacent to $p_k$ within 2 indices; and $n$ is the number of these complete vectors.

### 3 Methodology

In this section, a method of estimating and quantifying the similarity between two typhoon time series based on the dynamic time warping (DTW) algorithm is expounded in detail. DTW was first proposed in the 1960s (Bellman and Kalaba 1959) and developed rapidly in the field of speech recognition in the 1970s (Sakoe and Chiba 1978; Myers et al. 1980). This algorithm is a high-performance algorithm with noteworthy robustness that adopts the idea of dynamic programming to accumulate the shortest warping distances of two time series whose lengths are unequal (Müller 2007). Due to the inequality of existence time of different typhoons, the lengths of the corresponding time series are not equal. However, conventional time series processing algorithms require the same length of time series, which brings about the potential fitting and resampling of the time series when forcing the length of the two typhoon time series to be the same and thus results in the decline of the accuracy. Therefore, the application of the DTW algorithm can better avoid the problems caused by conventional algorithms.

Data that the DTW algorithm generally handles are sampled with equal time intervals, but as mentioned in Sect. 2.1, the observation intervals of typhoons are usually unequal. Giorgino (2009) pointed out that DTW is “stretch-insensitive,” for it is insensitive to prolonged duration of tones when applied to speech recognition. Consequently, inconsistent time intervals of typhoons can be viewed as different degrees of stretching to make the timeline equally spaced. And for this reason, time indicators are trivial factors in the DTW algorithm.

#### 3.1 Purpose and symbol definition

Considering two given historical typhoons, this section proposes a method to analyze the corresponding typhoon time series, which the analysis indicates the difference between.
To simplify the description of the method, some definitions of symbols are given as follows.

(a) The two historical typhoon time series are denoted as

\[ B = \{ b_1, b_2, b_3, \ldots, b_n \} \]

(3)

\[ C = \{ c_1, c_2, c_3, \ldots, c_m \} \]

(4)

where \( B \) is named the base time series whose length is \( n \) and \( C \) is named the compared time series whose length is \( m \), and most of the time this nomenclature is circumstantial. Each component of \( B \) and \( C \), namely \( b_i \) and \( c_i \), is a multidimensional vector mentioned in Sect. 2.

(b) The distance function between two multidimensional vectors \( p_1 \) and \( p_2 \) is defined as

\[ \text{dis}(p_1, p_2) = \sum_{i=1}^{n} f_i (p_{1i} - p_{2i})^2 \]

(5)

where \( p_{1i} \) is the \( i \)-th component of the vector \( p_1 \); \( p_{2i} \) is the \( i \)-th component of the vector \( p_2 \); \( f_i \) is the positive coefficient of the weight of the \( i \)-th part and \( \sum_{i=1}^{n} f_i = 1 \); and \( n \) is the number of dimensions of the vector.

### 3.2 Similarity analysis

(a) Construct a distance function matrix \( \text{Dis} \) with a size of \( n \times m \) as:

\[
\text{Dis}_{n \times m} = \begin{bmatrix}
\text{dis}(b_1, c_1) & \text{dis}(b_1, c_2) & \cdots & \text{dis}(b_1, c_m) \\
\text{dis}(b_2, c_1) & \text{dis}(b_2, c_2) & \cdots & \text{dis}(b_2, c_m) \\
\vdots & \vdots & \ddots & \vdots \\
\text{dis}(b_n, c_1) & \text{dis}(b_n, c_2) & \cdots & \text{dis}(b_n, c_m)
\end{bmatrix}
\]

(6)

(b) Construct a cost function matrix \( \text{Acc} \) with a size of \( n \times m \) by filling elements from top left to bottom right according to the rule (Senin 2008):

\[
\text{Acc}_{i,j} = \begin{cases} 
\text{Dis}_{1,1} & i = j \\
\text{Acc}_{1,j-1} + \text{Dis}_{1,j} & i = 1, j > 1 \\
\text{Acc}_{i-1,1} + \text{Dis}_{i,1} & i > 1, j = 1 \\
\min\{\text{Acc}_{i-1,j-1}, \text{Acc}_{i-1,j}, \text{Acc}_{ij-1}\} + \text{Dis}_{ij} & i > 1, j > 1
\end{cases}
\]

(7)

where the subscripts \( i, j \) indicate the element located at the \( i \)-th row and the \( j \)-th column of the matrix.

(c) The elements in the typhoon time series are multidimensional vectors with linear units. Considering that all elements in the cost function matrix, \( \text{Acc} \), are quadratic in dimension, the square root of the last one of \( \text{Acc} \) is taken as the similarity distance to make the unit of the result linear, namely,

\[
\sqrt{\text{Acc}_{n,m}}
\]

(8)
3.3 Normalization of similarity

Soboleva and Beskorovainyi. (2008) introduced a modified hyperbolic tangent function originally for multiobjective optimization in decision-making. This function is defined as:

$$m \tanh (x; a, b, c, d) = \frac{e^{ax} - e^{bx}}{e^{cx} + e^{dx}}$$  \hspace{1cm} (9)

which can be viewed as a universal form of the hyperbolic tangent function as it turns into the normal \( \tanh \) function when \( a = b = c = d = 1 \). This function is widely used in AI fields as a kind of activation functions to increase the nonlinearity of neural networks (Golev et al. 2017).

Given that the \( m \tanh \) function can inherit the features of the \( \tanh \) function better when 4 parameters \( a, b, c, \) and \( d \) are all equal to one positive real number, which is marked as \( q \), a simplified function for normalizing the similarity distance to a number ranging in \((0, 1]\) is defined as:

$$t(x; p) = \frac{p^x - p^{-x}}{p^x + p^{-x}}$$  \hspace{1cm} (10)

where \( p = e^q > 1 \) and \( p \) is named the simplification parameter.

To derive a clear and comprehensible index for evaluating the similarity between the two typhoon time series, a warped function named the similarity percentage is defined as:

$$f(x; p) = \left[ 1 - t(x; p) \right] \times 100\%$$  \hspace{1cm} (11)

where \( x \) is the similarity distance educed by Formula (8), and \( p \) is the simplification parameter mentioned in Formula (10).

4 Experiments

4.1 Experiment design

The typhoon time series that this section utilizes are extracted from Sect. 2.1. Typhoon 202,018 Molave is chosen as the base time series while typhoons that developed in 2019 or 2020 make up a compared time series set with a size of 52 in which each typhoon is chosen to be a compared time series in sequence. Specifically, 29 typhoons developed in 2019 are indexed from 201,901 to 201,929, and 23 typhoons developed in 2020 are indexed from 202,001 to 202,023.

In Sect. 3.1, Formula (5) introduced \( n \) coefficients \( f_i \) (\( i = 1, 2, 3, \ldots, n \)), where \( n = 5 \) since vectors involved in the distance function are all five dimensional, as mentioned in Sect. 2.1 but without time indicators as explained in the beginning of Sect. 3, which can be considered the significance of different elements of typhoons. Among the five elements, the longitude and the latitude are required, and the other three elements are optional, which is not fixed and can be adjusted flexibly in practice use. According to empirical recognition of typhoons and experts’ advice, within this experiment, the importance of geodesic elements (longitude and latitude) is emphasized, while the importance of dynamical elements (movement speed) is abated instead. In consideration of typhoon presence as a synoptic
system, meteorological elements (central pressures and Beaufort scales) are also weighted more than dynamical elements but less than geodesic elements. In summary, these coefficients are assigned values, as shown in Table 1.

Therefore, the expansion form of the distance function is

$$\text{dis}(p_1, p_2) = 0.26(Lon_1 - Lon_2)^2 + 0.26(Lat_1 - Lat_2)^2$$

$$+ 0.22(P_1 - P_2)^2 + 0.14(L_1 - L_2)^2 + 0.12(V_1 - V_2)^2$$

(12)

Meanwhile, the simplification parameter $p$ introduced in Formula (10) is assigned with 1.005, which will be discussed later.

### 4.2 Result presentation and analysis

After calculating the similarity distance and the similarity percentage between the base time series and each typhoon in the compared time series set, 6 typhoons listed in Table 2 are selected as representative typhoons and plotted in Fig. 2. Please notice the base position of Typhoon 202,018’s time series, hence, the zero similarity distance and the 100-percent similarity percentage for a time series are absolutely the same as itself and nothing vary.

The fundamental principle of similarity normalization is that the shorter the similarity distance between the compared time series and the base one is, the higher the percentage this pair takes, since the similarity distance is a positively correlated metric measuring the contrast of two time series, and the normalization function is strictly monotonically decreasing, which is corroborated in Fig. 2.

The paths of these 6 typhoons are mapped in Fig. 3 using the Mercator projection. From the map, it can be intuitively perceived that Typhoon B, which has an up-to-80 similarity

| Symbol | Meaning | Category       | Value |
|--------|---------|----------------|-------|
| $f_1$  | Longitude         | Geodesic     | 0.26  |
| $f_2$  | Latitude          | Geodesic     | 0.26  |
| $f_3$  | Central pressure  | Meteorological | 0.22  |
| $f_4$  | Expanded Beaufort Scale | Meteorological | 0.14  |
| $f_5$  | Move speed        | Dynamical    | 0.12  |

### Table 2 The indices and names of the selected typhoons are listed in this table

| Typhoon code | Typhoon index | Typhoon name |
|--------------|---------------|--------------|
| A            | 202,018       | Molave       |
| B            | 202,022       | Vamco        |
| C            | 202,007       | Higos        |
| D            | 202,004       | Hagupit      |
| E            | 202,008       | Bavi         |
| F            | 201,909       | Lekima       |

Codes are also attached to these typhoons for simpler formulation.
Fig. 2 The similarity distances between the base typhoon and each of six selected typhoons

Fig. 3 The paths of selected typhoons annotated with their codes and similarity percentages are displayed on the map. The abbreviation “SP” means “Similarity Percentage”
percentage, is the closest to Base Typhoon A, while northern-trajectory Typhoons D, E, and F with longer similarity distances are huddled together and perpendicular to the base typhoon overall, with Typhoon C following a similar but outlying path from the base one.

In this experiment, a laptop equipped with an Intel® Core™ i5-8265U CPU @ 1.60 GHz, an 8 GB RAM @ 2400 MHz, and a 256 GB SSD is installed with Windows 10 Home 64 Bits, Python 3.8.0, and gcc (MinGW.org GCC-8.2.0–3) 8.2.0; codes are programmed in Python-C hybrid style in which Python is used to load and pretreat data, and C-Language is intended to accelerate the calculation of similarity distances and percentages. The Python package `time` is used to determine the runtime of each programming language. Conclusively, with the typhoon time series of Typhoon A selected as the base time series and all other 52 time series in 2019 or 2020 forming the compared time series set, during 50 repeated experiments, the Python side took 6.829 ms and the C-Language side took 60.730 ms on average, which means that results are calculated in an instant.

### 4.3 Parameter discussion

In Fig. 3, under the circumstance where the geographic similarity is emphasized, as mentioned in Sect. 4.1, it is slightly contradictory that Typhoon E has a higher similarity percentage than Typhoon F, while E is farther away from A (the base typhoon) than F. The reason is that the similarity can also be affected by elements apart from geodesic ones. Section 4.1 introduces the significance rank of coefficients of different elements and assigns different values. The idea of this process is that the more significant some element is, the higher the expectation of exerting more prominent discrimination on the corresponding factor. A re-experiment is conducted to correct the visual contradiction without the influence of meteorological and dynamical elements. The comparison is shown in Table 3, which illustrates that the extent of elements’ influences can be customized by modifying their coefficients.

In addition, the five coefficients mentioned in Table 1 can be assigned different values according to different uses. From the perspective of meteorologists, the similarity in meteorology is more significant than in geographic morphology, and the assignment is shown in Table 4.

Typhoon 201,909 Lekima is selected as the base time series, and the compared time series set is expanded to cover all typhoons ranging from 2018 to 2020 except discarded typhoons. With the simplification parameters $p$ remaining unchanged, the top 4 typhoons, including the base typhoon, with the shortest similarity distances, are chosen as representative typhoons, and their similarity information is plotted in Fig. 4. These typhoons

| Typhoon code | Coefficient assignment | Similarity metrics |
|--------------|------------------------|--------------------|
|              | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | Distance | Percentage (%) |
| Before       | E     | 0.26  | 0.26  | 0.22  | 0.14  | 0.12     | 125.22   | 44.6          |
|              | F     |       |       |       |       |          | 165.40   | 32.2          |
| After        | E     | 0.5   | 0.5   | 0     | 0     | 0        | 151.46   | 36.2          |
|              | F     |       |       |       |       |          | 149.70   | 36.7          |
are Typhoon 201,909 Lekima coded as “G,” Typhoon 202,010 Haishen coded as “H,” Typhoon 201,808 Maria coded as “I,” and Typhoon 202,009 Maysak coded as “J.”

Similar to the origin experiment, the similarity percentage of the base time series remains 100% because of the zero similarity distance. Paths of these typhoons are mapped in Fig. 5.

Base Typhoon W developed at 17:00 (UTC+8, and the same below) on August 4, 2019, on the ocean surface east of the Philippines and then traveled northwestward, rapidly intensifying from a strong tropical storm to a super typhoon within 24 h. At 1:45 on August 10, it made landfall off the coast of Wenling, Zhejiang Province, China, with a maximum wind speed of 52 m/s rated Grade 16 and a minimum central pressure of 930 hPa, which gradually weakened after landfall. Typhoon Z was developed at 17:00 on August 28, 2020, on the ocean surface east of the Philippines and moved southwest. At 5:00 on September 1, it intensified into the first super typhoon of 2020 and then shifted to the east-northwest and began to weaken, making landfall off the coast of Busan, Gyeongsangnam-do, South Korea at approximately 1:30 on the 3rd with a wind speed of 42 m/s rated Grade 14 and central pressure of 950 hPa. Typhoon Y was developed at 20:00 on September 1, 2020, in the northwest Pacific Ocean, intensified to a super typhoon at 5:00 on the 4th and landed on the southern coast of

| Symbol | Meaning                   | Significance | Value |
|--------|---------------------------|--------------|-------|
| $f_1$  | Longitude                 | 3            | 0.16  |
| $f_2$  | Latitude                  | 3            | 0.16  |
| $f_3$  | Central pressure          | 2            | 0.26  |
| $f_4$  | Expanded Beaufort scale   | 1            | 0.28  |
| $f_5$  | Move speed                | 5            | 0.14  |

Fig. 4 The similarity distances between the base typhoon and each of four selected typhoons
South Korea at approximately 7:30 on the 7th with wind speeds of 40 m/s rated Grade 13 and central pressure of 955 hPa. Typhoon X developed at 20:00 on July 4, 2018, moved northwest, continued intensifying, and eventually intensified to a super typhoon at 5:00 on the 6th, making landfall on the coast of Huangqi Peninsula, Lianjiang, Fujian Province, China, at 9:10 on the 11th with wind speed of 42 m/s rated Grade 14 and central pressure of 960 hPa.

It is obvious that the top 3 most similar typhoons to super Typhoon 201,909 (W) are all categorized as super typhoons, which demonstrates the universality of the new coefficient setting. However, a deficiency is reflected in Fig. 5: The similarity percentages are not clearly discriminative. Since the similarity percentage is calculated by Formula (11), the simplification parameter $p$ is the decisive parameter varying the result. Situations of Formula (11) with different simplification parameters are plotted in Fig. 6.

It is easy to prove that

$$\forall x > 0, f(x; p_1) < f(x; p_2)$$

where $p_1 > p_2 > 1$. This proof can be interpreted that the plot of $f(x; p)$ will nest toward the origin point $(0, 0\%)$ with the increase of $p$. In addition, the function of $p$ is to stretch the plot in different ranges, as plot $f(x; 1.4)$ can be applied to handle similarity distances ranging in $(0, 10]$. Since the purpose of switching the simplification parameter $p$ is to distribute similarity distances to the range of $(0, 1]$, combined with data ranging in $(0, 500]$ in the original experiment, it can achieve a remarkable effect when $p < 1.4$. 

Fig. 5 The paths of the most similar typhoons, annotated with their codes and similarity percentages, and the base typhoon are displayed on the map. The abbreviation “SP” means “Similarity Percentage”
4.4 Similarity analysis on ongoing typhoon

Typhoon 202,106 In-Fa generated at 2:00 on July 18, 2021, in the northwest Pacific Ocean. It traveled anticlockwise and then moved northwest when it was approximately at the point of 125.6°E, 23.5°N at 1:00 on the 23rd. Given that Typhoon 202,106 is an ongoing typhoon till now (19:00, July 24, 2021), when it is chosen as the base typhoon, a cutoff rule is proposed as follows to extract valid parts of historical typhoons: The valid typhoon extraction begins with the point of generation and ends with the point that is nearest to the latest point of the ongoing typhoon under the WGS84 (World Geodetic System 1984), covering all points within this segment.

Since an ongoing typhoon is the first half of the entire typhoon, while data of the entire typhoon are incomplete before its extinction, the track is the main factor that needs to be taken into consideration and coefficients of longitudes and latitudes are assigned as 0.5 and other elements are ignored within this experiment. With the simplification parameters $p$ remaining 1.005, Typhoon 202,106 is selected as the base typhoon, and the top 5 similar typhoons are selected as representative typhoons, and their similarity distances and percentages are listed in Table 5.

Tracks of these 6 typhoons are also mapped in Fig. 7. The cutoff points are located where solid lines turn into dashed ones, which means that solid segments are valid typhoon extraction, and dashed segments are rest parts of corresponding historical typhoons.

From Fig. 7, all five similar typhoons take similarity percentages over 90% and can be considered more potential in helping predict the ongoing Typhoon K. While typhoon
L, O, and P made their landfall off the coast of Zhejiang Province, China, Typhoon K is also expected to land at the same area. Furthermore, with the development of Typhoon K, the algorithm will adjust the similarity calculation, and by then, Typhoon M and N would fade out of the list of similar typhoons.

Therefore, this experiment also corroborates that the method with the cutoff rule functions well in handling ongoing typhoons and finding similar typhoons that have certain referential value.
5 Conclusion

The definition of typhoon similarity has not been clearly and accurately determined by academic circles. As a result, the method proposed in this paper contributes to first quantifying and normalizing the similarity between two historical typhoons based on the dynamic time warping algorithm. Experiments illustrate that the method can make a distinction between typhoons and filter the relatively more similar typhoons under certain parameter specifications, such as Typhoon B with an 82.6% similarity and Typhoon C with a 68.3% similarity, and also demonstrate the feasibility that this method is able to process various types of historical and ongoing typhoon time series data while maintaining appreciable objectivity and accuracy. Moreover, the method shows great potential in seeking similar typhoons with ongoing typhoons and supporting typhoon prediction. The parameters involved with this algorithm are highly customizable in light of users’ needs.

However, this research still lacks a rational explanation for the similarity distance in physics and a systematic criterion to assign parameters reasonably and purposefully. In the future, a series of parameter recommendations need to be formulated for usages in different circumstances and fields. With the increase in typhoon data collected and the support of machine learning technology, a method for forecasting the path of a developing typhoon based on this research also requires investigation for practical application.

Acknowledgements Many thanks to anonymous reviewers for their valuable comments. This paper was supported by the NSFC Project (41871285).

Funding The NSFC Project (41871285).

Data availability The data used in this paper are available at: https://myetyet.coding.net/public/typhoon-similarity/typhoon-similarity/git/files/master/zjwater.

Code availability The code used in this paper is available at: https://myetyet.coding.net/public/typhoon-similarity/typhoon-similarity/git/files/master/exp.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Alemany S, Beltran J, Perez A, Ganzfried S (2019) Predicting hurricane trajectories using a recurrent neural network. Proc AAAI Conf Artif Intel 33(01):468–475
Alizadeh D, Alesheikh AA, Sharif M (2020) Prediction of vessels locations and maritime traffic using similarity measurement of trajectory. Ann GIS. https://doi.org/10.1080/19475683.2020.1840434
Bellman R, Kalaba R (1959) On adaptive control processes. IRE Trans Autom Control 4(2):1–9
Bui DT, Hoang ND, Pham TD, Ngo PTT, Hoa PV, Minh NQ, Tran XT, Samui P (2019) A new intelligence approach based on GIS-based multivariate adaptive regression splines and metaheuristic optimization for predicting flash flood susceptible areas at high-frequency tropical typhoon area. J Hydrol 575:314–326
Gao S, Zhao P, Pan B, Li Y, Zhou M, Xu J, Zhong S, Shi Z (2018) A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network. Acta Oceanol Sin 37(5):8–12
Giorgino T (2009) Computing and visualizing dynamic time warping alignments in R: the dtw package. J Stat Softw 31:1–24
Golev A, Iliev A, Kyurkchiev N (2017) A note on the Soboleva’ modified hyperbolic tangent activation function. Int J Innov Sci Eng Technol 4(6):2348–7968
Hong M, Zhang R, Liu K, Yu D (2007) A medium/long-range forecast of Pacific subtropical high based on dynamic statistic model reconstruction. Acta Meteorol Sin 21(3):376–384
Hong S, Kim S, Joh M, Song SK (2017) Globenet: convolutional neural networks for typhoon eye tracking from remote sensing imagery. http://arxiv.org/abs/1708.03417
Hsu M, Chen A, Chen L, Lee C, Lin F, Huang C (2014) A GIS-based decision support system for typhoon emergency response in Taiwan. Geotech Geol Eng 29:7–12
Huang D, Zheng X, Zhao D, Wang L (2019) Similarity measure method of multiple typhoon time series. Appl Res Comput 36(03):844–850
Jiang G, Xu J, Wei J (2018) A deep learning algorithm of neural network for the parameterization of typhoon-ocean feedback in typhoon forecast models. Geophys Res Lett 45(8):3706–3716
Li W, Zhu J, Fu L, Zhu Q, Xie Y, Hu Y (2021a) An augmented representation method of debris flow scenes to improve public perception. Int J Geogr Inf Sci 35(8):1521–1544
Li W, Zhu J, Fu L, Zhu Q, Guo Y, Gong Y (2021b) A rapid 3D reproduction system of dam-break floods constrained by post-disaster information. Environ Model Softw 139:104994
Moayedi A, Abbaspour RA, Chehreghan A (2019) An evaluation of the efficiency of similarity functions in density-based clustering of spatial trajectories. Ann GIS 25(4):313–327
Müller M (2007) Dynamic time warping, information retrieval for music and motion. Springer Berlin Heidelberg, pp 69–84
Myers C, Rabiner L, Rosenberg A (1980) Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. IEEE Trans Acoust Speech Signal Process 28(6):623–635
Sakoe H, Chiba S (1978) Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Process 26(1):43–49
Senin P (2008) Dynamic time warping algorithm review, Ph.D. thesis, Information and Computer Science Department, University of Hawaii at Manoa Honolulu, USA
She L, He L, Miao N (2014) Similar paths analysis of historical typhoons in Ningbo Based on WebGIS. Zhejiang Hydrotech 42(01):24–26
Soboleva EV, Beskorovainyi VV (2008) The utility function in problems of structural optimization of distributed objects. Kharkiv University of Aair Force 121
Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46(1):234–240
Wang S, Mu L, Yao Z, Gao J, Zhao E, Wang L (2021) Assessing and zoning of typhoon storm surge risk with a geographic information system (GIS) technique: a case study of the coastal area of Huizhou. Nat Hazard Earth Sys 21(1):439–462
Wu Y, Chen L, Cheng C, Yin K, Töröké Á (2014) GIS-based landslide hazard predicting system and its real-time test during a typhoon, Zhejiang Province, Southeast China. Eng Geol 175:9–21
Zhang W, Vecchi GA, Villarini G, Murakami H, Gudgel R, Yang X (2017) Statistical-dynamical seasonal forecast of western North Pacific and East Asia landfalling tropical cyclones using the GFDL FLOR coupled climate model. J Clim 30(6):2209–2232.
Zhou C, Chen P, Yang S, Zheng F, Yu H, Tang J, Lu Y, Chen G, Lu X, Zhang X, Sun J (2021) The impact of Typhoon Lekima (2019) on East China: a postevent survey in Wenzhou City and Taizhou City. Frontiers of Earth Science, 1–12
Zou L, Ren A, Xu F, Zhang X (2008) Typhoon track forecasting based on GIS spatial analyses. J Tsinghua Univ (science and Technology) 48(12):2036–2040

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.