An Objective Track Similarity Index and Its Preliminary Application to Predicting Precipitation of Landfalling Tropical Cyclones

FUMIN REN, WENYU QIU, CHENCHEN DING, XIANLING JIANG, LIGUANG WU, YINGLONG XU, AND YIHONG DUAN

State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China
Key Laboratory of Meteorological Disaster, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, China
College of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu, China
Hainan Meteorological Observatory, Haikou, China
National Typhoon and Marine Weather Forecast Center, Beijing, China

(Manuscript received 22 January 2018, in final form 4 September 2018)

ABSTRACT

Combining dynamical model output and statistical information in historical observations is an innovative approach to predicting severe or extreme weather. In this study, in order to examine a dynamical–statistical method for precipitation forecasting of landfalling tropical cyclones (TC), an objective TC track similarity area index (TSAI) is developed. TSAI represents an area of the enclosed scope surrounded by two TC tracks and two line segments connecting the initiating and ending points of the two tracks. The smaller the TSAI value, the greater the similarity of the two TC tracks, where a value of 0 indicates that the two tracks overlap completely. The TSAI is then preliminarily applied to a precipitation forecast test of landfalling TCs over South China. Given the considerable progress made in TC track forecasting over past few decades, TC track forecast products are also used. Through this test, a track-similarity-based landfalling TC precipitation dynamical–statistical ensemble forecast (LTP_DSEF) model is established, which consists of four steps: adopting the predicted TC track, determining the TC track similarity, checking the seasonal similarity, and making an ensemble prediction. Its application to the precipitation forecasts of landfalling TCs over South China reveals that the LTP_DSEF model is superior to three numerical weather prediction models (i.e., ECMWF, GFS, and T639/China), especially for intense precipitation at large thresholds (i.e., 100 or 250 mm) in both the training (2012–14) and independent (2015–16) samples.

1. Introduction

There has been considerable progress made in the numerical forecasting of tropical cyclone (TC) tracks during the past few decades (Fraedrich et al. 2003; Langmack et al. 2012; Cangialosi and Franklin 2015). This forecast improvement can be attributed mainly to better numerical weather prediction (NWP) model performance, such as the reasonable representation of the large-scale circulation as well as ensemble forecasting (Fraedrich et al. 2003; Zhang and Krishnamurti 1997; Goerss 2000). However, our ability to predict precipitation associated with landfalling tropical cyclones (LTCs) with NWP models, which is one of the major challenges facing today’s TC scientific community (Chen et al. 2010; Woo et al. 2014), is still very limited—especially in predicting the area coverage, intensity, and finescale distribution of precipitation (Tuleya et al. 2007; Marchok et al. 2007; Huang et al. 2009; Wang et al. 2012). Therefore, improving the reliability and accuracy of LTC precipitation forecasts remains as one of our most challenging tasks.

The first solution to the above challenging task is to improve the NWP models. It is well known that there have been considerable advances in model core development, parameterization of physical processes, and data assimilation (Chen and Xue 2004; Bauer et al. 2015), which have resulted from a steady accumulation of scientific knowledge and technological advances over many years (Bauer et al. 2015). It is becoming increasingly difficult to improve the forecasting skill of NWP models (Wang and Chen 2007).
On the other hand, dynamical–statistical models, which are the combination of a dynamical model and a statistical relation, have been developed as one important step forward in weather prediction. For example, some statistical–dynamical models, for example, SHIPS (DeMaria and Kaplan 1994, 1999; DeMaria et al. 2005), Logistic Growth Equation Model (LGEM; DeMaria 2009), and extensions for rapid intensification (Kaplan and DeMaria 2003; Kaplan et al. 2010, 2015; Rozoff and Kossin 2011; Rozoff et al. 2015), have been developed for TC intensity prediction. For the prediction of LTC precipitation, the studies adopting dynamical–statistical models can be sorted into three types. First, using historical observations of LTC precipitation and applying a climate mean, LTC precipitation forecasts can be obtained, based on the NWP forecast of LTC tracks (Marks et al. 2002; Lee et al. 2006; Lonfat et al. 2007). Second, assuming constant rainfall rates during the landfalling process, the LTC precipitation prediction can be estimated by integrating the rainfall rates obtained at the initial time along the NWP forecast LTC track for the forecast period (Kidder et al. 2005; Liu 2009; Ebert et al. 2011). A third type of the LTC precipitation forecasts involves applying analog forecast techniques. Based on analog analyses of a target TC within a forecast period with historical cases mainly in the large-scale circulation fields, previous studies (e.g., Zhong et al. 2009; Li and Zhao 2009) have made LTC precipitation forecasts using historical LTC precipitation and an index-based ensemble scheme of LTC similarity but without including LTC tracks. One good example is the Climatology-Based Quantitative Rainfall tool (CLIQR; http://www.wpc.ncep.noaa.gov/tropical/rain/web/cliqr.html) that has been operationally used at the Weather Prediction Center.

Nevertheless, it still remains uncertain as to how the dynamical–statistical approach could improve LTC precipitation forecasts. Clearly, a key issue is determining how to effectively combine the advantages of dynamical models with the valuable information from historical observations. The abovementioned first two types of studies focus on the NWP forecasts of LTC tracks, which is one of the most important factors for TC precipitation, whereas the third type focuses on analog forecast techniques but without using the model-predicted LTC tracks. Since today’s model-predicted TC tracks have become more reliable (Fraedrich et al. 2003; Langmack et al. 2012; Cangialosi and Franklin 2015), it is highly desirable to incorporate the model-predicted LTC tracks into any analog forecast technique.

To achieve this, the first step is to identify the analog TC tracks from historical data for a target TC track. Then, a suitable TC track similarity index must be found. It is well known that a TC track is an end product of multiscale interactions with a variety of physical factors. Chen and Ding (1979) summarized three criteria that TC track similarities should meet: seasonal, geographical, and TC moving direction and speed similarity. Zhong (2002) and Zhong et al. (2007, 2012) defined a nonlinear TC track similarity index based on multiple similarities, including landfalling time, the initial position of a TC, TC central pressure, and environmental fields. Wang et al. (2006) proposed a spatial similarity index (SSI) based on GIS technology, which is the ratio of a polygon area constituted by two TC tracks inside a specific region to an area of interest. Xu et al. (2013) proposed a track similarity criterion by averaging the Euclidean distance at all key points, such as sharp turning points, between two TCs. Liu et al. (2006) studied the algorithm of TC track similarity deviation. Based on GIS and distance functions, Scheitlin et al. (2010, 2013) developed a polyline averaging technique, which has the potential to identify track similarity. In operational TC prediction in China, whether or not TCs pass through a fixed region is also used as a criterion for identifying TC track similarity. Among the above indices, some deal with only the geometric shape between two TC tracks, while others take into consideration not only the geometric shape between two TC tracks but also other elements such as the seasonal, TC moving direction, and speed similarities, and even environmental fields. Although the studies above have examined TC track similarity indices, few suitable indices have been developed with comprehensive consideration of the realistic processes involved in TC tracks and with direct application in an operational setting.

Thus, the major purpose of this study is to develop a new index for identifying TC track similarity, which we call the tropical cyclone track similarity area index (TSAI). TSAI refers to areas enclosed by two TC tracks with a segment connecting the two initiating points and another segment connecting the two ending points of the two TC tracks. As compared to the previous research into track similarity techniques, an advantage of the TSAI is its simple physical implication (i.e., an area between two tracks). Section 2 describes the TSAI concept and presents the algorithms in terms of geometries to calculate the areas (TSAI) covered by two TC tracks on an (x, y) plane. Section 3 shows an application to LTC precipitation forecasting. A summary and concluding remarks are given in the final section.

2. Tropical cyclone track similarity area index

a. Zonal and meridional similarity patterns

TCs move mainly northward in the western North Pacific (WNP) basin. For 2115 TCs during 1949-2012
observed by the Shanghai Typhoon Institute, 2004 TCs (94.7% of the total) travel northward, which means that the latitude of the initiating point of a TC is less than that of its ending point, or both of the two points are located at the same latitude. Only 112 TCs (5.3% of the total) move southward.

To determine the similarity patterns, we define a latitudinal extreme point, which is the northernmost point or the southernmost point within a cyclone track. There are two latitudinal extreme points for each track. There are 1092 TCs (51.6% of the total TCs) with endpoints (either the initiating point or the ending point of a TC track) coincident with latitudinal extreme points. A concept of “the segmentation ratio of the latitudinal extreme point,” which can be used to define how close to the endpoints a latitudinal extreme point is, has been proposed and is discussed in further detail in the appendix, section b(1). If the criteria are not strict, which means that we are not only considering the endpoints to be the latitudinal extreme points, there are 1667 TCs (78.8% of the total TCs) whose latitudinal extreme points are close to the endpoints of the tracks (defined as where the segmentation ratio of the latitudinal extreme point is less than 0.2). Then, these TCs can be taken as north–south-moving cases, which belong to a meridional pattern (Fig. 1, left). Meanwhile, the remaining TCs (22.2% of the total), which have no latitudinal extreme points close to the endpoints, are considered as cases of east–west movement, which belong to a zonal pattern (Fig. 1, right). Only when the general directions\(^1\) of the two tracks are the same, is it meaningful to discuss track similarity. That is, for a zonal pattern, the general directions of the two tracks should both be eastward or westward, while for a meridional pattern, the general directions of the two tracks should both be northward or southward.

Generally, we focus on the TC track similarity within a specific region or designated region. In this case, the above considerations can still apply to the segments of the TC track. For any two TC tracks, or track segments (dotted lines in Fig. 1), if the two initiating and the two ending points are connected by two lines (thick dashed lines in Fig. 1), the two TC tracks or track segments constitute a closed region of which the area \(S\) can be calculated. The smaller \(S\) is, the more similarity there is between the two TC tracks, with \(S = 0\) meaning that the two tracks or track segments perfectly match each other. Therefore, for any two TC tracks, or track segments, the area of the enclosed scope surrounded by the two TC tracks and the two line segments, which connect the two initiating and the two ending points, can effectively represent their similarity. This area, as a similarity index of TC tracks, is named TSAI.

\(\text{b. TSAI calculation method}\)

The detailed calculation method of TSAI is presented in the appendix, and the calculation of TSAI includes five steps (Fig. 2): preprocessing TC tracks, identification of a track pattern, track idealization, and calculation and determination of the similarity index. Taking Typhoons Nina (197506) and Talim (200513), here we show how to calculate the TSAI of the two typhoons. The two observed TC tracks are shown in Fig. 3a.

The first step is preprocessing TC tracks, including determining the tracks within a designated region and the simplification of complex tracks. The designated region is indicated by the blue rectangle in Fig. 3b. The two tracks within the region can be determined as a result of the first substep (Fig. 3b). Since the two track segments are simple, the second substep (“simplification of complex tracks”) makes no change to the two tracks shown in Fig. 3b. In the second step (“identification of track pattern”), segmentation ratios \(r\) of the latitudinal extreme points of the two tracks are first calculated. For the two tracks, \(r = 0\), which means that neither of the two tracks have any latitudinal extreme points close to the endpoints. Then, the conditions for zonal pattern similarity are not met while those for meridional pattern similarity are met. Therefore, in the third step (“track idealization”), the meridional pattern track idealization is carried out. Due to the characteristics of the two tracks in the designated region, the idealized tracks are the same as those in Fig. 3b without any modification. In the fourth step (“calculation of the similarity index”), following the process of this step, the meridional pattern TSAI \(S_{\text{lon}}\) is calculated and \(S_{\text{lon}} = 40825.9\text{ km}^2\). Then, according to the flowchart, the fifth step (“determination of TSAI”) is taken directly. As \(n = 0\), which means that

\(^1\) See “General direction” in the appendix, section b(1).
the two tracks do not have any latitudinal extreme points that are close to the endpoints, TSAI = $S_{\text{lon}}$.

c. Sensitivity of TSAI to external parameters

The calculation of TSAI involves both the simplifying tracks and the selections of the three external parameters: the designated region, the threshold $r_0$ of the segmentation ratio of a latitudinal extreme point, and the threshold $c_0$ of the overlap percentage of two TC tracks in the general direction (north–south or east–west). Considering the simplifying tracks can be carried out automatically when necessary and in most cases it does not need to be done, in this section, only the influence of the three parameters on the similarity results is discussed.

We investigate the sensitivity of these parameters in TSAI sample calculations using the Shanghai Typhoon Institute TC best track dataset.

1) SIMILARITY REGION

The method can be applied to both full track similarity and designated region similarity. Full track similarity means that the similarity region can be large enough for the two TC tracks of concern, while designated region similarity means the similarity region is focused on a designated area.
Typhoon Nina (1975) caused serious damage in Zhumadian in Henan Province after making two landfalls across Taiwan Island and along the coast of Fujian Province before moving inland. Figure 4 shows Typhoon Nina and the five most similar TCs according to TSAI within the different similarity regions. Figure 4a presents the results of full track similarity. All five analog TCs move in a similar direction (southeast–northwest) to that of Typhoon Nina, although for full track similarity one of the five TCs makes no landfall on Taiwan Island and another makes no landfall at Fujian.

Figure 4b shows the results of similarity before landfall, with the similarity being assessed by focusing on the blue box. All five analog TCs make a first landfall in Taiwan and move across north-central Taiwan and then make a second landfall along the central to southern Fujian Province, before the tracks diverge.

Figure 4c displays the results of similarity after landfall. All five TCs move across south-central Fujian Province and enter Jiangxi Province, while the tracks diverge before the TCs make landfall at Fujian.

Therefore, according to the different similarity regions, the TSAI can provide different but reasonable TCs.

2) **THRESHOLD \( r_0 \) OF THE SEGMENTATION RATIO OF A LATITUDINAL EXTREME POINT**

The threshold of the segmentation ratio of a latitudinal extreme point can affect whether or not a latitudinal extreme point of a track is close to the endpoints. Therefore, this will affect the determination of TSAI.

Figure 5 shows the strong Tropical Storm Bilis (2006) and the five most similar TCs according to TSAI, with different thresholds \( r_0 \) of the segmentation ratio of a latitudinal extreme point under full track similarity. Figures 5a and 5b show the results with \( r_0 = 0.2 \) and 0.25, respectively, and the two results are clearly different. A segmentation ratio of 0.25 leads to the two southwestward-turning tracks after landfall (Fig. 5b), whereas a segmentation ratio of 0.2 does not (Fig. 5a). Further analysis shows that the segmentation ratio of the latitudinal extreme point of Bilis is 0.23, and the segmentation ratios of the latitude extreme points of the two TCs just mentioned before are both less than 0.20. The value of \( r_0 \) can therefore affect the similarity results to a certain extent. The reason is that, in step 4 (determination of TSAI) for the TSAI between Bilis and any of the two TCs, when \( r_0 = 0.25 \), TSAI = \( S_{lon} \), and when \( r_0 = 0.20 \), TSAI = \( \max(S_{lat}, S_{lon}) \), and this influences the results directly.

3) **THRESHOLD \( c_0 \) OF THE OVERLAP PERCENTAGE OF THE TWO TC TRACKS**

The overlap percentage of two TC tracks \( c \) has the range \([0, 1.0]\). Based on experience, the suggested threshold \( c_0 \) is between 0.4 and 0.8.

Figure 6 presents the tracks of Supertyphoon Haitang (2005) and the five most similar TCs based on TSAI with different thresholds \( c_0 \) for TC tracks over the blue designated region for landfall similarity. When \( c_0 = 0.4 \) (Fig. 6a), the lengths of the five TC tracks are different in the designated region, where the initiating point of a TC track is within the region. When \( c_0 = 0.8 \) (Fig. 6b), all five TC tracks move across the designated region and show a higher degree of similarity.

3. Preliminary application in LTC precipitation forecasting

In this section we describe an application of TSAI for the forecasting of accumulated precipitation associated with LTCs. This development has taken place over one year of research, resulting in a technique called the track-similarity-based landfall tropical cyclone...
precipitation dynamical–statistical ensemble forecast model (LTP_DSEF). Here, we present the main idea of the LTP_DSEF technique and the key results of experimental prediction.

In the test, we focus on South China (Fig. 7) and TCs that produced more than 100 mm of daily precipitation at least one station in the region. For verification of the precipitation forecast for the LTP_DSEF model, we adopted a regional NWP model approach, as these models have relatively high resolution and perform better when predicting precipitation. However, as no suitable regional NWP model’s results are available, three global NWP models are adopted for this work. The period 2012–16, during which data are available for three dynamical models: the European Centre for Medium-Range Weather Forecasts (ECMWF) model, the Global Forecast System (GFS) run by the National Weather Service (NWS), and the global spectral model in National Meteorological Center/China Meteorological Administration (T639), with horizontal resolutions of $0.25^\circ \times 0.25^\circ$, $1^\circ \times 1^\circ$, and $1.125^\circ \times 1.125^\circ$, respectively, is selected as the analysis period, with 2012–14 designated for the training sample and 2015–16 for the independent sample. There are a total of 27 TCs: 17 in 2012–14 and 10 in 2015–16. The TC dataset, which is for identifying analog TCs, is the best track data from the CMA Tropical Cyclone Database (http://tcdata.typhoon.org.cn/en/zjlssj_sm.html) over the period 1958–2016. The training sample differs from the independent sample in that a TC of the former can have analog TCs that occur after the TC under investigation while a TC of the latter cannot. To perform the verification, the precipitation forecasts created by the three NWP models are interpolated directly with the algorithm of inverse–distance-weighted interpolation onto the 191 stations across South China (Fig. 7).

To identify TC precipitation, the objective synoptic analysis technique (OSAT; Ren et al. 2001, 2007, 2011),
which uses the distance from the TC center and the closeness and continuity between neighboring raining stations to identify TC-influenced rain belts that may extend from 500 to 1100 km away from a TC center, and the daily precipitation data during 1958–2016 of the 191 stations are applied. To assess the forecast effect, it is important to choose a suitable skill measure. Considering that the threat score (TS) is the primary operational-forecast verification metric in China, TS is selected as the skill measure in this study. TS values range between 0 and 1, with 1 indicating a perfect score, and its formula is as follows: 

$$TS = \frac{\text{hits}}{\text{hits} + \text{misses} + \text{false alarms}}$$

(https://www.cawcr.gov.au/projects/verification/). TS is used for different thresholds (such as 0.1, 10, 25, 50, 100, and 250 mm) of LTC accumulated precipitation over the station network in Fig. 7.

The LTP_DSEF technique includes four steps:

1) Predict the TC track. It is suggested that the TC track prediction of an NWP model be absorbed directly. Based on data availability, the TC track prediction used here is the forecast of the National Meteorological Center/China Meteorological Administration, which in turn is based on the TC track prediction of the NWP models. For a predicted track at an initial time, the lead time of the track prediction ranges from 12 to 120 h, depending on the initial time and the status of the TC.

2) Identify TC track similarity. There are two substeps. The first is to construct a complete TC track, which consists of the observed and the predicted tracks. The second is to apply TSAI to identify TC track similarity within a similarity region, which is a rectangular area covering the domain where the predicted track

![Fig. 5. Strong Tropical Storm Bilis (2006) and the five most similar TCs according to TSAI with different thresholds \( r_0 \) of the segmentation ratio of a latitudinal extreme point under full track similarity for (a) \( r_0 = 0.2 \) and (b) \( r_0 = 0.25 \).](image1)

![Fig. 6. The tracks of Supertyphoon Haitang (2005) and the five most similar TCs according to TSAI with different thresholds \( c_0 \) of overlap percentages of the two TC tracks under landfall similarity for (a) \( c_0 = 0.4 \) and (b) \( c_0 = 0.8 \).](image2)
extends. Then, according to TSAI, a number \( N \) of the most similar TC tracks can be identified.

3) Consider seasonal similarity. Three time periods are selected (July–September, May–November, and the whole year) for seasonal similarity. This means that, using the results derived by identifying the TC track similarity, the first \( n (n < N) \) most similar TCs can be determined after considering seasonal similarity.

4) Develop ensemble predictions of the accumulated precipitation using ensembles of TCs with the most similar tracks. Two methodologies are used in the development: the ensemble mean rainfall or selection of the maximum rainfall value for each station.

In this test, there are seven parameters in the LTP_DSEF model (Table 1): initial time \( (P_1) \), similarity region \( (P_2) \), threshold of the segmentation ratio of a latitudinal extreme point \( (P_3) \), threshold of the overlap percentage of the two TC tracks \( (P_4) \), seasonal similarity \( (P_5) \), number of the most similar TCs \( (P_6) \), and an ensemble scheme \( (P_7) \). Considering these numerous parameters with different values or settings, which are listed in Table 1, ideally there are a total of 103,680 \( (= 4 \times 15 \times 3 \times 6 \times 3 \times 16 \times 2) \) different schemes. However, because some of the 27 TCs produce rainfall soon after genesis and those TCs can influence the number of values for parameters \( P_1 \) and \( P_2 \), this will result in a decrease in the total number of schemes for the test. To avoid a small total number of schemes occurring, the 27 TCs have been divided into short-track TCs (6) and long-track TCs (21), as in Table 2, where a short-track TC is defined as a TC with the number of values for parameter \( P_1 \) being 1, and only the 21 long-track TCs (15 for the training sample and 6 for the independent sample) are used for the test. Therefore, the final total number of schemes is 15,552.

To obtain suitable schemes for both the training and the independent samples, the TS for the precipitation above the different thresholds \( (r \geq R_0) \), where \( R_0 \) has six values of 0.1, 10, 25, 50, 100, and 250 mm) is calculated according to the LTC accumulated precipitation over the station network in Fig. 7. Figure 8 shows the training sample–independent sample TC \( (r \geq 100 \text{ mm}) \) cross-section distribution for the 15,552 different schemes in the TC precipitation prediction of the LTP_DSEF model. All the schemes can be sorted into two groups: one is the ensemble mean (gray points), while the other is the maximum value ensemble scheme (orange points), with the average of the orange points \( (0.170, 0.192) \) being much better than that of the gray points \( (0.097, 0.081) \). This means that the maximum value ensemble is much better than the ensemble mean for intense precipitation forecasts. From an ensemble scheme perspective, the maximum value ensemble shares some similarities with the probability matching procedure outlined by Ebert (2001), which was shown to improve precipitation forecast skill in his study. In Fig. 8, the three open circles represent the three dynamical models (ECMWF, GFS, and T639), and the two dotted lines indicate the highest values of the training sample TSt\(_{100} \) (0.168) and the independent sample TSi\(_{100} \) (0.238), both of which are from GFS. A total of 259 schemes are better than the dynamical models and are located in the first quadrant of the two dotted lines.

To identify the schemes within the total 259 that have better-than-NWP-model performance in the accumulated precipitation \( \geq 250 \text{ mm} \), we prepared Fig. 9. The two dotted lines represent the highest values of the training sample TSt\(_{250} \) (0.043) and the independent sample TSi\(_{250} \) (0.0), for the three dynamical models (ECMWF, GFS, and T639), which are also both from GFS. There are a total of 202 schemes that are better than the dynamical models and are located in the first quadrant of the two dotted lines (Fig. 9). This means that the 202 schemes show better-than-NWP-model performance in the accumulated precipitation of \( \geq 100 \) and \( \geq 250 \text{ mm} \) in the TC precipitation prediction of the LTP_DSEF model.

To identify the best choice among the 202 schemes and considering precipitation of \( \geq 100 \text{ mm} \) while already including that of \( \geq 250 \text{ mm} \), we find that TSt\(_{100} \) and TSi\(_{100} \) are suitable for doing this. Figure 10 presents the training sample–independent sample TS (represented by TSt\(_{100} \) and TSi\(_{100} \), respectively) cross-section distribution for the 202 schemes. The scheme marked with the open circle, which has the largest value of TSt\(_{100} \) + TSi\(_{100} \), is selected as the best scheme. In this scheme, the seasonal similarity is the whole year, the number of the most similar TCs is nine, the ensemble prediction scheme is the maximum, the initial time is the latest one, and the other three parameters are those of TSAI (Table 3).
**Figure 11** presents a comparison of the training and independent sample TSs for precipitation above different thresholds for the best scheme of LTP_DSEF and the three dynamical models. Although the LTP_DSEF model does not show any advantages over the three dynamical models at small precipitation thresholds (0.1–25 mm), it shows better prediction ability than the three dynamical models at large precipitation thresholds (100–250 mm) in the training and the independent samples. For example, for $100 \text{ mm}$ precipitation, the TS values for the three dynamical models range between 0.115 and 0.168 (between 0.171 and 0.238) in the training (independent) sample, while that of the best scheme of the LTP_DSEF model is 0.198 (0.266). As for the reasons for the above results, in addition to the potential ability of the LTP_DSEF technique for predicting heavy precipitation, part of the reason may be that the dynamical models limit their abilities to predict heavy precipitation because of low resolution.

To inspect the stability of the above results, a new set of tests, which randomly choose a set of the TCs for training and the remainder for testing, have been performed. Based on the 21 TCs and selecting four consecutive TCs for training, there are five new tests, with TEST1 using the last four TCs, TEST2 using the 14th–17th TCs, TEST3 using the 10th–13th TCs, TEST4 using the 6th–9th TCs, and TEST5 using the 2nd–5th TCs for training. The best schemes of the above five tests, noting that is unnecessary that they be the same, can be easily identified. Results of the best schemes of the five new tests have been compared with these of the three dynamical models (ECMWF, GFS, and T639) (figure omitted). Generally, the LTP_DSEF model shows better performance than do the dynamical models, with the results of the training sample being more stable than those of the independent sample, which may probably be attributed mainly to the large difference between the sample size (17) in the training sample and in the independent sample (4).

### 4. Summary and discussion

Based on the above analysis and discussion, we offer the following conclusions.

1) A new tropical cyclone track similarity index (i.e., the TSAI) has been developed. For any two TC tracks, TSAI has a simple meaning, which is the area of the enclosed scope surrounded by the two TC tracks and the two line segments connecting the first and the last two points of the two TC tracks. The smaller the value of TSAI, the greater the similarity for the two tracks. The detailed calculation process of TSAI has also been presented, including five steps: preprocessing TC tracks, identification of the track pattern, track idealization, calculation of the similarity index, and determination of TSAI. In addition, there are three adjustable parameters: the similarity region, the threshold of the segmentation ratio of a latitudinal extreme point, and the threshold of the overlap percentage of two TC tracks.
Using examples of sensitivity tests for the external parameters, we demonstrate that TSAI has good capability in characterizing TC track similarity.

2) During the preliminary application of TSAI to develop dynamical–statistical analog methods for predicting LTC precipitation, a new technique—the LTP_DSEF model—was developed. The LTP_DSEF model includes four steps: predict the TC track, identify TC track similarity, consider the seasonal similarity, and develop an ensemble prediction of the LTC accumulated precipitation. There are seven parameters in the LTP_DSEF model.

3) The forecasting test for the LTP_DSEF model for South China during 2012–16, with 2012–14 used as the training sample and 2015–16 as the independent sample, reveals that, for the LTC accumulated precipitation forecast, the LTP_DSEF model is superior to three NWP models (ECMWF, GFS, and T639), especially at large precipitation thresholds (100 or 250 mm) in the training and independent samples. For example, for 100-mm precipitation, the TS value for the best scheme of the LTP_DSEF model is 0.198 (0.266), which is better than those of the three dynamical models, at 0.115–0.168 (0.171–0.238) in the training (independent) sample.

Some issues concerning the calculation of TSAI must be mentioned. First, a similarity region needs to be specified. Generally, there are three types of the similarity regions: full track similarity, similarity before landfall, and similarity after landfall. Second, application of the TSAI method does require simplification of the TC tracks. In practice, and in most cases, simplification is not necessary, especially for application within a designated region, such as the landfall areas in the examples in Figs. 4 and 6. However, occasionally, for a singular TC track with twisting or even spiral parts, TSAIs between this track and other TC tracks may be limited due to track simplification, such that manual verification is required. Third, in rare cases, TSAI cannot be calculated when the general directions of the two tracks are opposite, which means that there is no similarity between the two TC tracks. For example, for the zonal similarity pattern, there exists no similarity between a TC with a general westward direction and another TC whose general direction of motion is from west to east. Fourth, TSAI only represents track pattern similarity. For application in analog rainfall prediction, additional factors will need to be considered, including factors of the TC itself, such as landfall seasonal similarity, TC intensity, TC forward speed, and TC structure, as well as environmental factors such as monsoon similarity, subtropical high similarity, low-level jet similarity, and vertical wind shear similarity. Further analysis is needed to incorporate these effects (Velden and Leslie 1991).

Compared with existing track similarity indices, such as Wang et al. (2006), Xu et al. (2013), Liu et al. (2006), Zhong (2002), and Zhong et al. (2007), TSAI has three distinctive features: (i) it has a clear meaning in that it represents the area of the enclosed scope surrounded by the two TC tracks, (ii) it is objective with a more detailed introduction, and (iii) it is easily applied in studies, as well.
as in weather and climate operations. It should be noted that the concept of the area between two TC tracks, which is the core meaning of TSAI, is actually included in the SSI index (Wang et al. 2006). However, since the concept is not independently proposed and the calculating technology is GIS, there are limits to the application of SSI. For the application of TSAI, the first thing to be thought of is possibly the track forecast. However, as considerable progress has been made by numerical models in track forecasts (Leslie and Fraedrich 1990; Zhang and Krishnamurti 1997; Goerss 2000; Fraedrich et al. 2003; Langmack et al. 2012; Cangialosi and Franklin 2015), it is unnecessary to apply TSAI when forecasting TC tracks. Considering the successful preliminary application of TSAI, as reported upon in section 3, TC precipitation forecasting is an important field where TSAI can be used. However, more complete validation of the performance of the LTP_DSEF model is needed, and more factors that influence TC precipitation, such as TC characteristics (e.g., TC intensity and translation speed) and environmental characteristics (e.g., intensity of the summer monsoon, status of the WNP subtropical high), should be taken into consideration in the LTP_DSEF model in the future.

We believe that further studies on the TSAI and its application in dynamical–statistical analog prediction for LTC precipitation will deliver upon its great potential in both research and operational applications.

Acknowledgments. The authors would like to express their sincere thanks to the two anonymous reviewers; the editor, Prof. Dalin Zhang; and Dr. John L. McBride for their helpful suggestions and comments to improve the manuscript, as well as the Shanghai Typhoon Institute and National Meteorological Information Center/China Meteorological Administration for sharing the tropical cyclone best-track dataset and daily precipitation data, respectively. This work was supported by the National Natural Science Foundation of China (Grants 41675042...
APPENDIX

TSAI Calculation Method

The detailed calculation processes for these steps are outlined in this appendix.

a. Preprocessing TC tracks (Fig. 2, top box)

In general, we are interested in the TC track similarity within a designated region, and occasionally a TC track may be so singular that it has twisting or even spiral parts. To facilitate the calculation of TSAI, it is necessary to preprocess the original TC tracks. The preprocessing of TC tracks includes two steps: (i) determining the tracks within a designated region and (ii) simplification of the complex tracks (Fig. 2).

1) Determining the tracks within a designated region\textsuperscript{A1}

As shown in Fig. A1a, when a designated region (ABCD) is defined, the track within the region is determined through the following steps.

First, identify all the TC points within the designated region (open circles in Fig. A1a). Second, calculate all the points of intersection (filled circles in Fig. A1a) of the track and the border in the designated region. Third, exclude all the points (open circles with crosses in Fig. A1a) that are outside the designated region. Finally, connect the open and filled circles in the original time order to produce a new track within the designated area (Fig. A1b).

2) Simplification of complex tracks

Figure A2 is a schematic diagram of the simplification of a complex track. For a given TC track that is within the designated area and composed of many observation locations (points), the simplification of the track can be completed with the following four steps. First, determine the bizarre points. Suppose the track has \( m \) points in total. For point \( j (j = 1, \ldots, m) \) of the track, calculate the distances between point \( j \) and its two adjacent points \( (j - 1 \) and \( j + 1) \).\textsuperscript{A2} and select the longer one, \( d_j \). If there is another point \( (k) \) that makes the distance between itself and point \( j \) less than or equal to \( d_j \), point \( j \) is called a bizarre point (filled points in Fig. A2). Then, exclude all the bizarre points (or outliers). Third, to reduce the bending degree of the track, exclude the short line segment that is divided by the outliers, has less than three points, and does not contain the first or the ending point of the original track (such as CD). Finally, connect the remaining line segments again in the original time order to produce a simplified track. For example, after the simplification, points A and B become two adjacent points in time.

b. Identification of the track pattern

The TC track pattern includes individual TC track shapes and the relationship between the two TC track shapes.

1) Concepts

To identify the track pattern, three concepts, or parameters, are defined.

General direction: If the difference between the latitude (longitude) of the ending point and that of the initiating point is greater than or equal to zero, the general direction of the track is defined as northward (eastward) in the north–south (east–west) direction. When the difference is less than zero, the general direction of the track is defined as southward (westward) in the north–south (east–west) direction.

\textsuperscript{A1} When working with the TC track similarity for the whole lifetime of the TC, this step can be omitted.

\textsuperscript{A2} For the first and ending points of a TC track, the number of adjacent points is one. The number of adjacent points is two for the other points.
Segmentation ratio of a latitudinal extreme point:
For a TC track AB with A and B being the two endpoints, assume that a latitudinal extreme point \( A^3 \) (C) divides the track into two parts: AC and BC. Assume \( L_{AB} \) is the length of track AB, and \( L_C \) is the length of the shorter of the segments AC and BC. Then, the segmentation ratio of a latitudinal extreme point is defined as

\[
r = \frac{L_C}{L_{AB}}.
\] (A1)

By this definition, the value range of \( r \) is \([0, 0.5]\), where 0 means that the latitudinal extreme point is an endpoint, and 0.5 means that the latitudinal extreme point is located in the middle of the track.

For a certain threshold \( r_0 \), if a TC track does not have any latitudinal extreme points with \( r \geq r_0 \), it means that all the latitudinal extreme points of the track are close to endpoints. Generally, \( r_0 \) does not have a typical value, and a suitable value of \( r_0 \) can be gained through a sensitivity or operational test.

Overlap percentage: For any two TC tracks, assume \( L_0 \) is the length of the longer track (the track with open circles in Fig. A3). Based on the general direction of this track (in the north–south or east–west directions), assume that the track length of the segment AB (Fig. A3) that this track overlaps with is \( L_{\text{overlap}} \). Then, the overlap percentage of the two TC tracks in this direction is defined as

\[
c = \frac{L_{\text{overlap}}}{L_0}.
\] (A2)

The value range of \( c \) is \([0, 1.0]\), where 0 means that the two tracks do not overlap, and 1 indicates that the two TC tracks overlap completely.

2) ZONAL PATTERN SIMILARITY CRITERION
In the identification of track pattern box in Fig. 2, the first question addressed is that of zonal pattern similarity.

When the two tracks satisfy the following three conditions, we move to the middle-left box for zonal pattern track idealization. Otherwise, we move to the middle-right box (in Fig. 2) to consider whether there is meridional pattern similarity.

The three zonal pattern similarity conditions are

(i) at least one TC track has a latitudinal extreme point that is not close to the endpoints,
(ii) the general directions of the two tracks are the same in the east–west direction, and
(iii) for a given threshold \( c_0 \), the overlap percentage \( c \geq c_0 \).

Generally, \( c_0 \) does not have a typical value, and a suitable value can be gained through a sensitivity or operational test.

---

A3 "Latitude extreme point" refers to the maximum or minimum of a TC track latitude.
3) MERIDIONAL PATTERN SIMILARITY CRITERION

When the two tracks satisfy both of the following two conditions, they are considered to have meridional pattern similarity:

(i) the general directions of the two tracks are the same in the north–south direction and
(ii) for a given threshold $c_0$ ($c_0$ is generally around 0.5), the overlap percentage $c \geq c_0$.

When these conditions are met, we move to the middle-left box in Fig. 2 for meridional pattern track idealization. If they are not met, then the tracks have neither zonal nor meridional similarity, and we move to the bottom-left corner in Fig. 2: these two tracks are “not similar.”

c. Track idealization

To facilitate the calculation of TSAI, approximation or simplification of the TC tracks (i.e., track idealization) is needed. Track idealization can be divided into two types: meridional and zonal pattern.

1) MERIDIONAL PATTERN TRACK IDEALIZATION

Figure A4 shows a schematic diagram of meridional pattern similarity idealization. There are two steps:

(i) Unification of track direction: Adjust all the points of the track according to latitude in ascending (northward movement) or descending (southward movement) order, so that the moving direction between any two points is consistent with the general direction of the track (Fig. A4b).

(ii) The second simplification of complex tracks: After the unification of the track direction, the track (Fig. A4b) may be severely distorted. Therefore, to facilitate the calculation of TSAI, it is necessary to repeat the step in section a(2) of this appendix, which is called the second simplification of complex tracks. The new track after a second simplification is presented in Fig. A4c.

value of $c_0$ can be gained through a sensitivity or operational test.

FIG. A4. Schematic diagram of the meridional pattern track idealization. (a) The track after preprocessing of the TC track. (b) The track after the unification of the track direction. (c) The new track after the second simplification of the track, with the solid thin gray line AB replacing the original dashed one as part of the track.

FIG. A5. Schematic diagram of the zonal pattern track idealization. (a) Cutting lines along the longitude at the latitude extreme points and the endpoints. (b) Several triangles ($\Delta$) and quadrangles ($\square$) together approximate the enclosed scope for the calculation of TSAI. (c) A quadrangle is divided into two triangles, and (d) a triangle can be taken as a scope surrounded by two idealized tracks.
2) ZONAL PATTERN TRACK IDEALIZATION

Figure A5 shows the schematic diagram of zonal pattern track idealization. There are four steps in this process:

(i) Cut lines along the longitude at the latitudinal extreme points and the endpoints. At each latitudinal extreme point and endpoint, draw a cutting line along the longitude and calculate the points of intersection of the cutting line and the two tracks. At most, there are eight cutting lines, which are shown in Fig. A5a.

(ii) Several triangles (∆) and quadrangles (□) together approximate the enclosed scope for the calculation of TSAI. In Fig. A5b, connect the adjacent points of intersection of the same track with line segments. In rare cases, some intersections (represented by the small black boxes) may exist between two line segments. Then, several triangles and quadrangles, which are surrounded by the cutting lines and the line segments, form and they can together approximate the enclosed scope for the calculation of TSAI.

(iii) A quadrangle can be divided into two triangles. For a quadrangle (Fig. A5c), its diagonal (fine dotted line segment in Fig. A5c) can divide it into two triangles.

(iv) A triangle can be taken as a scope surrounded by two idealized tracks. For any triangle (Fig. A5d, left), when the latitude maximum and minimum points of the triangle are determined, the triangle can be taken as a scope surrounded by two idealized tracks of meridional pattern similarity (Fig. A5d, right).

After the above four steps, the zonal pattern TSAI is transformed into a sum of areas of all the triangles, and any triangle can be taken as an area surrounded by two idealized tracks of meridional pattern similarity. Therefore, the zonal pattern track idealization is finally converted to meridional pattern track idealization.

d. Calculation of the similarity index (Fig. 2, bottom-right box)

Based on the above analysis, the TSAI of the two TC tracks may include a meridional pattern TSAI and a zonal pattern TSAI, and the calculation of the zonal pattern TSAI depends on the algorithm of the meridional pattern TSAI. Therefore, once the algorithm of the meridional pattern TSAI is established, the zonal pattern TSAI is also solved.

The algorithm of the meridional pattern TSAI includes three steps: (i) slicing the scope, (ii) calculation of the area of a single slice, and (iii) addition of all the slice areas.

1) SLICING THE SCOPE

For the two idealized tracks, Fig. A6 shows a schematic diagram of slicing the scope surrounded by the two tracks and the two line segments connecting the first two and the last two points of the two TC tracks.

First, at each point of the two tracks, slice the scope with a cutting line and calculate the points of intersection of the cutting line and the two tracks. Then, the scope can be divided into a number of slices (Fig. A6), which can be sorted into three types of geometric graphs: triangle, trapezoid, and double triangle (Fig. A7).

2) CALCULATION OF THE AREA OF A SINGLE SLICE

For slice $i$, its area can be calculated as

$$S_i = \begin{cases} AC \times BD/2, & \text{triangle} \\ (AP + BQ) \times BD/2, & \text{trapezoid} \\ (AP \times MD + BQ \times ME)/2, & \text{double triangle} \end{cases}$$

(A3)
Equation (A3) shows the three formulas for calculating the areas of a triangle (Fig. A7a), trapezoid (Fig. A7b), and double triangle (Fig. A7c). Only the calculation of the area of a double triangle is complex. First, the location of point M must be calculated. Figure A8 shows a schematic diagram of the intersection of the two line segments AB and PQ, which are assumed to be straight to more readily facilitate TSAI computation. Assume that \( x \) and \( y \) represent the longitude and latitude, and that the positions of A, B, P, and Q are \((x_a, y_a)\), \((x_b, y_b)\), \((x_p, y_p)\), and \((x_q, y_q)\), respectively. Then, the location of M \((x_m, y_m)\) can be determined by two steps.

First, calculate the equations of the straight lines AB and PQ according to the two point formula:

\[
y = y_a + k_{ab}(x - x_a) \quad \text{and} \quad y = y_p + k_{pq}(x - x_p),
\]

where \( k_{ab} = (y_b - y_a)/(x_b - x_a) \) and \( k_{pq} = (y_q - y_p)/(x_q - x_p) \).

Second, calculate the location of the intersection M \((x_m, y_m)\) of the two straight lines [(A4) and (A5)]:

\[
x_m = \frac{y_p - y_a + x_a k_{ab} - x_p k_{pq}}{k_{ab} - k_{pq}} \quad \text{and} \quad (A6)
\]

\[
y_m = y_a + k_{ab} \frac{y_p - y_a + x_a k_{ab} - x_p k_{pq}}{k_{ab} - k_{pq}} \quad \text{and} \quad (A7)
\]

Then, the area of the double triangle can be easily determined.

3) ADDING ALL THE SLICE AREAS

When the areas of all the slices have been determined, add all the areas together. Suppose there are \( L \) slices in total, the area of the scope surrounded by the two tracks can be expressed as

\[
S = \sum_{i=1}^{L} S_i. \quad (A8)
\]

e. Determining TSAI

Based on the algorithm of the meridional pattern TSAI, the meridional pattern TSAI \( S_{\text{lon}} \), and the zonal pattern TSAI \( S_{\text{lat}} \), the units of which are kilometers squared, are determined.

To determine the value of TSAI, we define the parameter \( n \), which is the size of the subset of the two tracks whose latitude extreme points are not close to the endpoints. Then, the value of TSAI is determined as

(i) \( n = 2 \), TSAI = \( S_{\text{lat}} \);

(ii) \( n = 1 \), TSAI = \( \max(S_{\text{lat}}, S_{\text{lon}}) \); that is, select the larger one; and

(iii) \( n = 0 \), TSAI = \( S_{\text{lon}} \).

REFERENCES


