Article ID: 1006-8775(2012) 03-0322-10

A SIMILARITY SCHEME FOR QUANTITATIVE FORECAST OF PRECIPITATION OF TYPHOONS

ZHONG Yuan (钟　元)[1], PAN Jin-song (潘劲松)[2], ZHU Hong (朱　红)[3], CHEN Wei-feng (陈卫锋)[3], CHEN Shi-chun (陈世春)[3], LIANG Ming-zhu (梁明珠)[3]

(1. Zhejiang Meteorological Research Institute, Hangzhou 310017, China; 2. Zhejiang Meteorological Observatory, Hangzhou 310017, China; 3. Huzhou Meteorological bureau, Huzhou Zhejiang 313000, china)

Abstract: A quantitative scheme is put forward in our work of forecasting the storm rainfall of typhoons for specific sites. Using the initial parameters, weather situations and physical quantities as well as numerical weather prediction products, the scheme constructs multivariate, objective and similarity criteria for environmental factors for the time between the current and forthcoming moment within the domain of forecast. Through defining a non-linear similarity index, this work presents a comprehensive assessment of the similarity between historical samples of typhoons and those being forecast in terms of continuous dynamic states under the multivariate criteria in order to identify similar samples. The historical rainfall records of the similar samples are used to run weighted summarization of the similarity index to determine site-specific and quantitative forecasts of future typhoon rainfall. Samples resembling the typhoon being forecast are selected by defining a non-linear similarity index composed of multiple criteria. Trial tests have demonstrated that this scheme has positive prediction skill.

Key words: weather forecast; forecasting methods; typhoon; storm precipitation; site-specific and quantitative forecast; similarity

CLC number: P444 **Document code:** A

1 INTRODUCTION

It remains one of the difficult issues to give site-specific and quantitative forecasts of the storm rainfall of the tropical cyclone (TC, including the tropical storm, severe tropical storm, typhoon and super typhoon, the same below), due to, on the one hand, the interactions between large-scale synoptic systems and meso- and fine-scale systems, between low-latitude systems and mid- and high-latitude systems, and between the environmental field, TC circulation and local conditions, as well as complicated TC precipitation itself[1-9], and on the other, the fact that a storm rain could last from 24 hours to 5–7 days, even 10 days, non-stop. As both of its temporal and spatial scale is so complicated, it is difficult to have accurate determination of either the mechanisms responsible for the precipitation triggered by the TC or the key factors and processes, making it hard to upgrade forecast accuracy.

Remarkable progress has been made in numerical weather prediction (NWP) in that large-scale background has been forecast with increasing accuracy, durations of forecast validity have been increased constantly, and there have been more and more numerical models dealing with the TC rainfall forecast. However, most of these models are useful in addressing the nowcast and short-term (3-6 h) forecast of TC rainfall, leaving little progress in site-specific and quantitative forecast of the storm rainfall[10-11]. The existing NWP only gives direct forecast of rainfall for gridpoints while the majority of observation sites are not located at the gridpoints, making interpolation the only way to forecast the rainfall for specific observation sites. As a matter of fact, the interpolation is an effective means of forecast for continuous predictants (such as temperature) but not necessarily so for those with large discreteness (such as rainfall). As the spatial distribution of rainfall does not always conform to existing methods of interpolation, which are designed artificially, forecasting with secondary approximation obviously reduces the accuracy.

Due to the present difficulty in direct forecasting of the storm rainfall of TC, indirect forecasting is an

Received 2009-12-25; **Revised** 2012-04-27; **Accepted** 2012-07-15
Foundation item: Specialized Research Project for Social Welfare from Ministry of Science and Technology of China (2005DIB3J104); Science and Technology Planning Project for Zhejiang Province (2007C23065)
Biography: ZHONG Yuan, research fellow, primarily undertaking research on typhoons.
Corresponding author: ZHONG Yuan, e-mail: yzhong686@hotmail.com

alternative to solve the problem and similarity forecast is quite effective in this aspect. As NWP is relatively accurate in short-term prediction of large-scale situations, it is able to provide information useful for forecasting future environmental fields. The key now is how to construct a reasonable forecasting scheme and apply the information to perform site-specific and quantitative forecast of the storm rainfall. An similarity scheme was once put forward by Zhong and Jin[12] to deal with the wind and rain brought about by the TC offshore. Unfortunately, the similar conditions it identified were incomplete and the forecasts it made were also not as satisfactory as they should due to constraints at that time—the only NWP product available was the 500 hPa geopotential height. The current work is just another attempt to solve this problem.

2 SAMPLE DATA AND DOMAIN OF SIMILARITY

Precipitation from the TC results from a given pattern of synoptic surroundings. When similar processes are identified in which interactions take place between the TC and the environmental field, it is possible to locate similar processes of rain. Of course, being similar is not only for the time prior to the point of forecast but also for that of the future evolution, a feature that is more essential for any prediction. Having similar preceding situations does not necessarily mean that there will be similar developments in the future as well; the future is exactly what forecasting should be based on. In view of it, the NWP—representative of dynamic processes—is the foundation of the methodology this work has employed.

Previously, a similar state is just for a particular point of time, a point of stationary similarity that is unable to reflect the evolution of events. To determine the similarity in evolution, i.e., dynamic similarity, the variations of discriminating factors for TC parameters at the time of forecast (t = 0), 6 h prior to it (t = –6), and 12 h prior to it (t = –12) are examined. For the dynamic similarity of the environmental field of these factors, samples at the time of forecast, 24 h, 48 h, and 72 h after it are taken for both the forecasting and historical periods.

The forecasting sample of the environmental field, $X(t)_{0\ k,\ l}$ is obtained from the NWP value of the synoptic situation field, $\hat{X}(t)_{k,l}$,

$$X(t)_{0,k,l} = \hat{X}(t)_{k,l}\,,$$

$$t=0,\ 24,\ 48,\ 72, \quad K, \quad l=1,\ 2,\ \cdots\cdots,\ L \qquad (1)$$

The historical sample of the environmental field, $X(t)_{j,\ k,\ l}$, is obtained from the historical record, $Xr(t)_{k,l}$, and $X(t)_{j,\ k,\ l}=Xr(t)_{k,\ l}$. t=0, 24, 48, 72, j=1, 2, ……, J, k=1, 2,……, K, l=1, 2, ……, L (2)

where t is the time, subscript 0 the forecast sample, subscript j the order of the historical sample, J the total of the historical sample, k the order of longitudinal gridpoints, K the total of longitudinal gridpoints, l the order of latitudinal gridpoints, and L the total of latitudinal gridpoints.

Two similar domains of the environmental field are taken based on the characteristics of the factors. A fixed domain, from 10–50°N and 90–150°E with the resolution at 2.5×2.5, is provided for large-scale weather situations, such as mid-level geopotential heights and upper-level stream flow. To reflect the TC motion, a movable domain is set for such environmental fields as low-level circulation, mid-level steering field, vertical motion and moisture transport. Centered on the TC center, the domain is a quasi-square that is 10 degrees of latitude from the center to either its south or north side and 10 degrees of longitude from the center to either its east or west side. It moves along with the TC (Figure 1).

The small circular spot in the figure is the location of the TC center available every 6 hours while the large one indicates where the TC center is at the time of forecast. The solid-line frame is the domain of movement similarity for TC Matsa (0509), while the dashed-line frame stands for that of TC Dujuan (0313), at the time of forecast. The bold, solid line is the 24-h watch line against the TC that has the potential to make landfall in China.

A total of 559 samples of TCs are selected that crossed the 24-h watch line and made landfall in China from 1949 to 2005. In each of the samples, the TC parameters (time, location, center pressure and maximum wind speed near the eye) are derived from the *Typhoons Yearbook* and *Tropical Cyclones Yearbook*. For the environmental field, the data are taken from 6-hourly daily reanalysis from the National Centers for Environmental Protection/National Center for Atmospheric Research (NCEP/NCAR, USA). For the forecast field, the data are taken from the European Center for Medium-Range Weather Forecast (ECMWF) and Beijing Meteorological Center (the NWP product T_{213}).

3 IDENTIFICATION BY SIMILARITY

As the rain brought about by a landfalling TC is the result of interactions between its own internal force, the environmental field and land surface, it is difficult to decide whether similarity can be achieved with only a few simple criteria. Instead, multiple criteria—composed of the TC parameters themselves as well as multi-time, multi-level environmental elements—should be used to conduct comprehensive assessment to identify reasonable similarity. For this purpose, the NWP results for the environmental field are used to introduce as criteria to the similarity forecast of TC rainfall so as to determine the dynamic similarity over the time prior to and after the point of forecast. The key of similarity forecast is to discriminate similar TC samples from numerous historical candidates. The methods introduced as follows are used in this work.

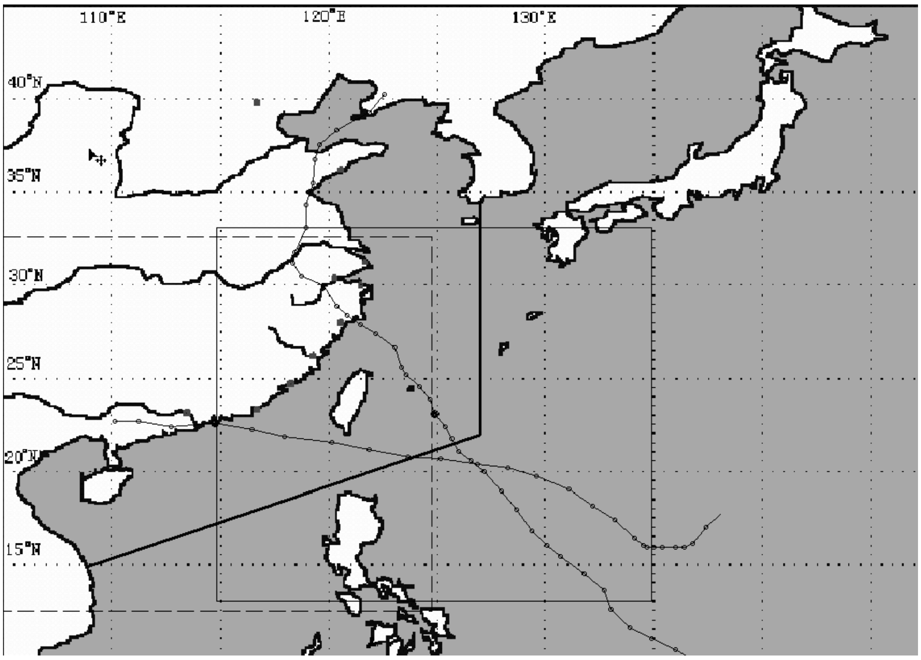

Figure 1. The domains of movement similarity in the forecast scheme.

3.1 *Similarity in the distribution of spatial plane field*

When the historical sample is $X_{j,k,l}$ and forecast sample is $X_{0,k,l}$ in a spatial plane field in which the total of zonal circles is K and the total of meridional circles is L for the gridpoints of a similar domain, the degree of similarity between them is determined through $M(M=K\times L)$ variables of the two samples and expressed by the cosine between two vectors in the M-dimensional space. Or, it has a similar coefficient of

$$Sx_j=\cos\theta_j=\frac{\sum_{l=1}^{L}\sum_{k=1}^{K}X_{j,k,l}\cdot X_{0,k,l}}{\sqrt{\sum_{l=1}^{L}\sum_{k=1}^{K}X_{j,k,l}^2\sum_{l=1}^{L}\sum_{k=1}^{K}X_{0,k,j}^2}}$$

$$j=1,2,\ldots\ldots,J,\ k=1,2,\ldots\ldots,K,\ l=1,2,\ldots\ldots,L \tag{3}$$

The larger the similarity coefficient, the more similar the two samples are. It seems a little coarse to determine the degree of similarity using Eq. (3) as computation for similarity through the whole field would result in excessive smoothing and make it difficult to shed light on particular details. To avoid the shortfalls, two similarity coefficients, $Sz_{j,k}$ and $Sm_{j,l}$, one normalized on the zonal direction and the other on the meridional direction, are first sought as in

$$Sz_{j,k}=\frac{\sum_{l=1}^{L}(X_{j,k,l}-\overline{X}_{j,k})(X_{0,k,l}-\overline{X}_{0,k})}{\sqrt{\sum_{l=1}^{L}(X_{j,k,l}-\overline{X}_{j,k})^2\sum_{l=1}^{L}(X_{0,k,l}-\overline{X}_{0,k})^2}}$$

$$j=1,2,\ldots\ldots,J,\quad k=1,2,\ldots\ldots,K,\quad l=1,\ 2,\ldots\ldots,\ L \tag{4}$$

where $\overline{X}_{j,k}=\frac{1}{L}\sum_{l=1}^{L}X_{j,k,l}$, $\overline{X}_{0,k}=\frac{1}{L}\sum_{l=1}^{L}X_{0,k,l}$

$$Sm_{j,l}=\frac{\sum_{k=1}^{K}(X_{j,k,l}-\overline{X}_{j,l})(X_{0,k,l}-\overline{X}_{0,l})}{\sqrt{\sum_{k=1}^{K}(X_{j,k,l}-\overline{X}_{j,l})^2\sum_{k=1}^{K}(X_{0,k,l}-\overline{X}_{0,l})^2}} \tag{5}$$

where $\overline{X}_{j,l}=\frac{1}{K}\sum_{k=1}^{K}X_{j,k,l}$, $\overline{X}_{0,l}=\frac{1}{K}\sum_{k=1}^{K}X_{0,k,l}$.

Then, $Sz_{j,k}$ for each of the zonal circles is averaged by the number of zonal circles to obtain the zonal similarity coefficient of the field $\overline{Sz}_j$,

$$\overline{Sz}_j=\frac{1}{K}\sum_{k=1}^{K}Sz_{j,k} \tag{6}$$

and $Sz_{j,k}$ for each of the meridional circles is averaged by the number of meridional circles to obtain the zonal similarity coefficient of the field $\overline{Sm}_j$,

$$\overline{Sm}_j=\frac{1}{L}\sum_{l=1}^{L}Sm_{j,l} \tag{7}$$

For the spatial plane field of factor X, its similarity coefficient Sx_j is

$$Sx_j=\overline{Sz}_j+\overline{Sm}_j. \tag{8}$$

3.2 *Similarity in the distance of spatial plane field*

The distance between the spatial plane field of the historical sample and that of the forecast sample is actually the difference in nature; the smaller (larger) the difference, the more (less) similar they are. In this work, the Euclidean distance is used to assess the similarity in distance between the spatial plane fields.

$$Dx_j = \sqrt{\sum_{k=1}^{K}\sum_{l=1}^{L}(X_{j,k,l} - X_{0,k,l})^2}$$

$$j=1, 2, \ldots\ldots, J, \quad k=1, 2, \ldots\ldots, K, \quad l=1, 2, \ldots\ldots, L \qquad (9)$$

3.3 *Similarity in the elements of spatial points*

For an element at a given fixed point in space, the distance of X_j between the historical sample x and forecast sample X_0 depicts their difference in nature. The smaller the difference, the more similar they are, and vice versa. Here in this work, the absolute distance, ΔX_j, is used to judge how similar the elements are in spatial points.

$$\Delta X_j = X_j - X_0 \text{ or } \Delta X_j = | X_j - X_0 |, \quad j=1, 2, \cdots\cdots, J. \qquad (10)$$

3.4 *Discrimination of similarity*

From a number of selected factors that are related to the rainfall of landfalling TCs, criteria for similarity are determined based on their characteristics using Eqs. (3)–(10) and their similarity indexes are assessed. Upon comprehensively assessing all of the similarity criteria, historical TC samples that bear the most similarity are located with the help of the forecast model.

4 INDEXES OF SIMILARITY

When similarity criteria are constructed using multiple factors, it is difficult to perform comprehensive assessment due to inconsistent dimension and range of value. Therefore, a similarity index is defined that sets a unified standard by which the extent of similarity can be known for criteria that differ by a wide range. Besides, numerous historical samples are distinguished from each other based on how similar they are to isolate the samples with the most similarity.

For the series of the ith factor, when the similarity function $f_{i,j}(x, x_0)$ for a given similarity relation is obtained from the historical sample $x_{i,j}$ and the forecast sample $x_{i,0}$, it could be either the function relation of the similarity coefficient Sx_j or the Euclidean distance Dx_j, or the absolute distance Δx. The two subscripts *i and j* are the order of the criterion and the sample order of the criterion factor, respectively.

With the maximum $\max(f)_i$ and minimum $\min(f)_i$ determined for the series of the ith factor, the extreme difference is divided into ten equal parts as in

$$\Delta D_i = [\max(f)_i - \min(f)_i]/10,$$

C_i, the threshold value—which can be either a maximum or minimum value depending on the similarity function $f_{i,j}(x, x_0)$—is determined for the series of the criterion factor. With ΔD_i set as the unit of distance and when $Df_{i,j}$ is the absolute distance between point (i, j) in the similarity space, whose similarity function is $f_{i,j}(x, x_0)$, and the threshold value C_i, i.e.

$$Df_{i,j} = | f_{i,j}(x, x_0) - C_i |, \qquad (11)$$

A similarity index can then be defined as in

$$SI_{i,j} = \begin{cases} 10, & Df_{i,j} = 0, \\ 10-k, & (k-1)\dfrac{\Delta D_i}{4} < Df_{i,j} \le k\dfrac{\Delta D_i}{4}, \quad k=1,2,3,4, \\ 6-k, & (k+1)\dfrac{\Delta D_i}{2} < Df_{i,j} \le (k+2)\dfrac{\Delta D_i}{2}, \quad k=1,2, \\ 4-k, & (3k+1)\dfrac{\Delta D_i}{2} < Df_{i,j} \le (3k+4)\dfrac{\Delta D_i}{2}, \quad k=1,2, \\ 1, & 5\Delta D_i < Df_{i,j} \le 7\Delta D_i, \\ 0, & 7\Delta D_i < Df_{i,j} \le 10\Delta D_i, \end{cases} \qquad (12)$$

As shown in the definition, the similarity index $SI_{i,j}$ is not proportional to the equal-length section of ΔD_i. Among its results, only a few $Df_{i,j}$ are small, i.e. just a few historical samples are close to the forecast samples and thus acquire indexes of high similarity, while a majority of the historical samples are given indexes of low similarity. As indicated in the statistics, factors of criteria are quasi-positively distributed by the frequency of equal-length sections. However, when non-linear transform is conducted according to a distribution function that has unequal-length sections, historical samples with high indexes appear at rates 20% lower than the original series that are distributed by equal-length sections, those with medium-range indexes at rates 50% lower, while those with low indexes at rates 70% higher. It is now clear that most of the criteria factors are assigned with values of low similarity indexes when processed with non-linear transformation following equal-length sectional distribution functions. As a result, similarity weights of the similar samples are increased while those of dissimilar samples are decreased, distinguishing the historical samples by the degree of similarity.

5 CRITERIA OF SIMILARITY

As criteria of similarity are key to locating rain processes from similar TCs, their selection and construction should be such that they much reflect on the characteristics and laws of behavior. For this purpose, some TC parameters, such as the initial position, center pressure, moving speed and direction, are selected that are able to describe the intensity and internal force of the TC, the initial state for future evolution and movement, and its interactions with the surroundings[13, 14].

Sea level pressure describes the TC intensity and range of impact while temperature is a factor related to the warm-core of the TC. As temperature is essentially integral to the life cycle of the TC, rainfall from the TC could be affected.

TC rainfall is closely related to its track. As the rain area varies with the TC movement, wind fields at both upper and lower levels and mid-level steering airflow are also important environmental factors that could play a part in the TC rainfall[15, 16].

Rainfall is related to water vapor and sustained rain depends greatly on continuous supply of water vapor. As a result, mid- and lower-level fields of moisture, airflow, moisture flux and its convergence are all indispensable factors of the criteria. When lifted by airflow, water vapor creates rainfall, suggesting close relationships between environmental vertical velocities and the TC rainfall.

On the basis of the key factors presented above and currently available NWP products, 15 factors were chosen to construct the following 28 criteria of similarity. They will be studied one by one as follows (the subscript j in the equation stands for the order of the historical samples and J for the total number of the historical samples).

5.1 *Initial position of the TC*

The reason why the initial TC position is made a criterion is that it is related to the future direction of movement and most of the TC rainfall is distributed off the two sides of the track.

Criterion 1 is the spherical distance between the position of its historical sample and the current position of the forecast TC, i.e.

$$f(1)_j = R \cdot \arccos[\sin\varphi_j \sin\varphi_0 + \cos\varphi_j \cos\varphi_0 \cos(\lambda_j - \lambda_0)];$$

$$C(1)=\min[f(1)_j],$$

where R is the Earth's radius, ϕ_j, λ_j are the latitude and longitude of the historical sample, respectively, and ϕ_0, λ_0 the latitude and longitude of the forecast sample.

5.2 *Initial TC center pressure*

Center pressure is the measurement of TC intensity, whose variation shows whether the TC is intensifying or weakening, and changes in internal TC force have an impact on the environmental field. That is why it is chosen one of the criteria for similarity.

Criterion 2 takes as its parameters the center pressure p at the initial time and its variations at 6 h and 12 h, as in

$$f(2, 0)_j = P_j,\ C(2, 0) = P_0,\ f(2, t)_j = P_j(0) - P_j(t),$$

$$C(2, t)_j = P_0(0) - P_0(t),\ t=-6, -12$$

5.3 *Initial TC velocity*

The reason whey initial TC velocity is used as one of the criteria is that it exerts direct impacts on the movement of TC to influence its track and rainfall.

Criterion 3 takes as its parameter the average moving velocity VV_J over the time within 6 h from the initial time, as in

$$f(3)_j = VV_j = R\ \theta_j/6;\ C(3)=\min[f(3)_j]$$

in which

$$\theta_j = \arccos[\sin\varphi_{0,j} \sin\varphi_{-6,j} + \cos\varphi_{0,j} \cos\varphi_{-6,j} \cos(\lambda_{0,j} - \lambda_{-6,j})].$$

5.4 *Initial moving direction of TC*

As the initial moving direction could have direct impacts on the movement of TC, it is used as one of the criteria for similarity.

Criterion 4 takes as its parameter the average moving direction α_j over the time within 6 h from the initial time, as in

$$f(4)_j = \alpha_j = \mathrm{arctg}[(\phi_{0,j} - \phi_{6,j})/(\lambda_{0,j} - \lambda_{6,j})];$$

$$C(4)=\min[f(4)_j]$$

5.5 *Distribution of 500 hPa geopotential heights*

500 hPa geopotential heights describe the mid-level atmospheric circulation. Under most circumstances, the TC track is subject to the subtropical high pressure in the western North Pacific. That is why it is used as one of the criteria for similarity.

On gridpoints with a resolution of 2.5°×2.5° and within the range of 10–50°N, 90–150°E in Northern Hemisphere, zonal circles take K=17 and meridional circles take L=25, historical samples take $X(t)_{j,k,l}=H(t)_{j,k,l}$ and forecast samples take $X(t)_{0,k,l}=H(t)_{0,k,l}$. Based on Eqs. (3)–(8), the similarity coefficient $Sh(t)_j$ is derived for the 0, 24, 46 and 72 h after the time of forecast. Then, the dynamic similarity coefficient for the 500 hPa geopotential heights Sh_j is integrated from $Sh(t)_j$ over the four hours prior to and after the time, as in

$$Sh_j = \frac{1}{4}\sum_{t=0}^{72} Sh(t)_j,\ t=0, 24, 48, 72.$$

In other words, Criterion 5 is composed of Sh_j and its threshold $C(5)$, as in

$$f(5)_j = Sh_j,\ C(5) = \max[f(5)_j]$$

5.6 *Distance in the 500 hPa geopotential heights*

The geopotential heights show the nature of a given point in the geopotential height field and similar distance in the heights suggests similar thermodynamic background between two TCs. That is why it serves as one of the criteria for similarity. Taking the domains, gridpoints and samples that are the same as section 5.5 and based on Eq. (9), the Euclidean distance between the historical and forecast samples at 0, 24, 48 and 72 h are computed. Then, the dynamic distance coefficient for the 500 hPa geopotential heights Dh_j is integrated from $Dh(t)_j$ over the four hours prior to and after the time, as in

$$Dh_j = \frac{1}{4}\sum_{t=0}^{72} Dh(t)_j, \quad t=0, 24, 48, 72.$$

Criterion 6 is composed of the dynamic distance coefficient for the 500 hPa geopotential heights and its threshold, as in

$$f(6)_j = Dh_j, \ C(6)=\min[f(6)_j].$$

5.7 *Criteria of similarity for other environmental field*

With the domains, gridpoints and samples that are the same as sections 5.5 and 5.6, Criteria 7–16 are constructed for the distribution and distance for the fields of 850 hPa temperature, 925 hPa humidity, 200 hPa stream flow, 925 stream flow and 200 hPa vertical velocity, respectively.

For the gridpoints within a movable 20°×20° domain of similarity and at a resolution of 2.5°×2.5°, K = 9 number of latitude circles and L = 9 number of longitude circles are taken. With the methods the same as sections 5.5–5.6, Criteria 17–28 are constructed for the distribution and distance of sea surface pressure, mid-level steering flow, mid- and lower-level vertical velocity, low-level moisture flux and moisture flux divergence, respectively.

6 SIMILARITY FORECAST

6.1 *Comprehensive assessment of multi-variable, objective criteria of similarity*

For general assessment of the extent of similarity, indexes of similarity are first sought based on Eq. (12), and then the similarity series are integrated for all of the criteria with an index of

$$SI_j = \sum_{i=1}^{28} W_i SI(i)_j,$$

$$i=1, 2, \cdots\cdots, 28, \ j=1, 2, \cdots\cdots, J,$$

where W_i is the weight for each of the criteria $SI(i)_j$. During the initial phase of model forecast, W_i=2 is taken for six of the criteria that are closely related to the TC, namely, initial position, initial moving direction, mid-level steering flow, low-level vertical velocity, low-level moisture flux and moisture while W_i=1 is taken for the rest of the criteria. With multiple runs of forecast, a considerable amount of forecast samples and conclusions are accumulated. On the basis of that, empirical or semi-empirical expressions can be analyzed that are with relative weights of each of the criteria for similarity.

6.2 *Determination of similar samples*

Realigning all of the J historical samples according to the value of SI_j, the value of SI_m in the new series $\{N_m\}$ can be arranged in a monotonously decreasing manner.

$$SI_1=\max[SI_m], \quad SI_J=\min[SI_m],$$

$$SI_1 \geqslant SI_2 \geqslant \cdots\cdots \geqslant SI_m \geqslant \cdots\cdots \geqslant SI_J, \ m=1, 2, \cdots\cdots, J$$

which is how the confidence of similarity is defined.

When the mean index of similarity of a new series is $\overline{SI}$,

$$\overline{SI} = \frac{\sum_{m=1}^{J} SI_m}{J}$$

6.3 *Site-specific and quantitative forecast of TC rainfall*

When the kth observation station in the nth similar historical sample measures a rainfall of $R_n(k, t)$ at the time of t, the site-specific and quantitative forecasts of the TC rainfall are the composites of similar historical samples that take as weights the similarity index SI_n,

$$\hat{R}(k,t) = \frac{\sum_{n=1}^{N} SI_n \cdot R_n(k,t)}{\sum_{n=1}^{N} SI_n}, \ n=1, 2, \cdots\cdots, N,$$

$$k=1, 2, \cdots\cdots, K, \ t=0, 24, 48, 72,$$

where n is the order of similar samples, N the total amount of similar samples, k the order of forecast stations, K the total number of forecast stations, and t the forecast validity.

6.4 *Forecast experiments*

For the forecast experiment, two TCs that have significant differences are selected. One is TC Matsa (0509), which has large rain rates and a long duration of rain for seven days (with landfall in Zhejiang), and the other is TC Dujuan (0313), which has small rain rates and a short duration of rain for only three days. They were used in five and three forecast experiments respectively. Table 1 presents the scores of the forecast results based on 697 national observation stations for Matsa and 408 stations for Dujian, respectively.

As shown in the results of the forecast experiments, the composite forecast yields smaller mean absolute errors but higher critical success index (CSI) than the similarity forecast for the storm rainfall of both TCs. It suggests that the composite forecast—weighted with similarity indexes with a number of similar samples—reduces miss rates and smoothes forecast errors, making the overall accuracy higher than that of individual members of the forecast.

As Matsa has a larger rainfall than Dujian does, its forecast error is also larger.

Table 1. Errors and scores of rainfall forecast in the experiments in the unit of mm.

Typhoon		Mean errors/mm	Scores of storm rainfall⩾50 mm	
			Hit rate/%	CSI/%
Matsa(0509)	1st analogy forecast	35	83	37
	Ensemble forecast	29	64	44
Dujuan(0313)	1st analogy forecast	28	86	41
	Ensemble forecast	23	72	47

Besides, according to statistics based on the results of forecasts for observation stations with storm rainfall ⩾50 mm, the hit rate is much higher with the first similarity forecast than with the composite forecast. As the latter smoothes the results from individual forecast members, forecasts of extreme situations are less satisfactory than those by individual forecast members.

The first five similar samples for the 5th forecast of Matsa (2000 Beijing Standard Time, or BST, 5 August 2005) and the 3rd forecast of Dujuan (2000 BST 2 September 2003) are listed in Table 2, their situation fields of similar samples Figures 2-3 and rainfall forecasts Figures 4-5.

Table 2. Main similar samples.

Typhoon	Order of similarity *m*	1	2	3	4	5
Matsa(0509)	Index of comprehensive similarity	187	170	146	138	132
	Similar typhoon	Winnie	Rananim	Jeff	Fred	Abe
Dujuan(0313)	Index of comprehensive similarity	169	159	158	148	146
	Similar typhoon	York	Wanda	Ellen	Gorden	Lois

As shown in Figure 2, the similar samples are similar to the forecast samples in terms of 700 hPa stream fields in the middle and lower latitudes: they both have anti-cyclonic circulation of the subtropical high in the western North Pacific in the TC circulation and its western and northwestern sides. Within 24 hours, the anti-cyclonic circulation weakens to some extent.

As shown in Figure 3, the similar samples are much like the forecast samples as far as the 925 hPa water vapor fluxes are concerned. Water vapor is transported from low-latitude tropical oceans to the TC circulation, with the latter having an area of high water vapor flux to its east and a low-value center to its west that is surrounded by a high-value area. Within 24 hours, the band of water vapor transport and the source of low-latitude water vapor from the tropical ocean interrupt, substantially reducing the transport of water vapor.

As shown in Figure 4, the forecast method has forecasts of Matsa that are close to the observation, especially with regard to the rainfall distribution and order of magnitude by the first similarity forecast.

As shown in Figure 5, the forecast method has forecasts of Dujuan that are basically consistent with the observation, especially with the distribution of rainfall and order of magnitude over the continent by the first similarity and composite forecast.

7 CONCLUSIONS

(1) Our forecast method applies the technique of similarity identification to interpret NWP products to perform site-specific and quantitative forecasts of storm rainfall of the TC, avoiding the secondary approximation error resulting from the use of interpolation.

(2) For a forecast method, the key to solving the storm rainfall forecasting lies in the introduction of similar changes of future environmental fields. By using the outputs from NWP as the criteria for determining the similarity of future environmental fields and taking into account the continuous and dynamic similarity of the initial and future environmental fields, similarity is built on the basis of dynamics, which is therefore relatively reasonable and reliable and thus helpful in increasing the forecast accuracy. One should note that forecast results obtained this way will be inevitably subject to the errors of NWP.

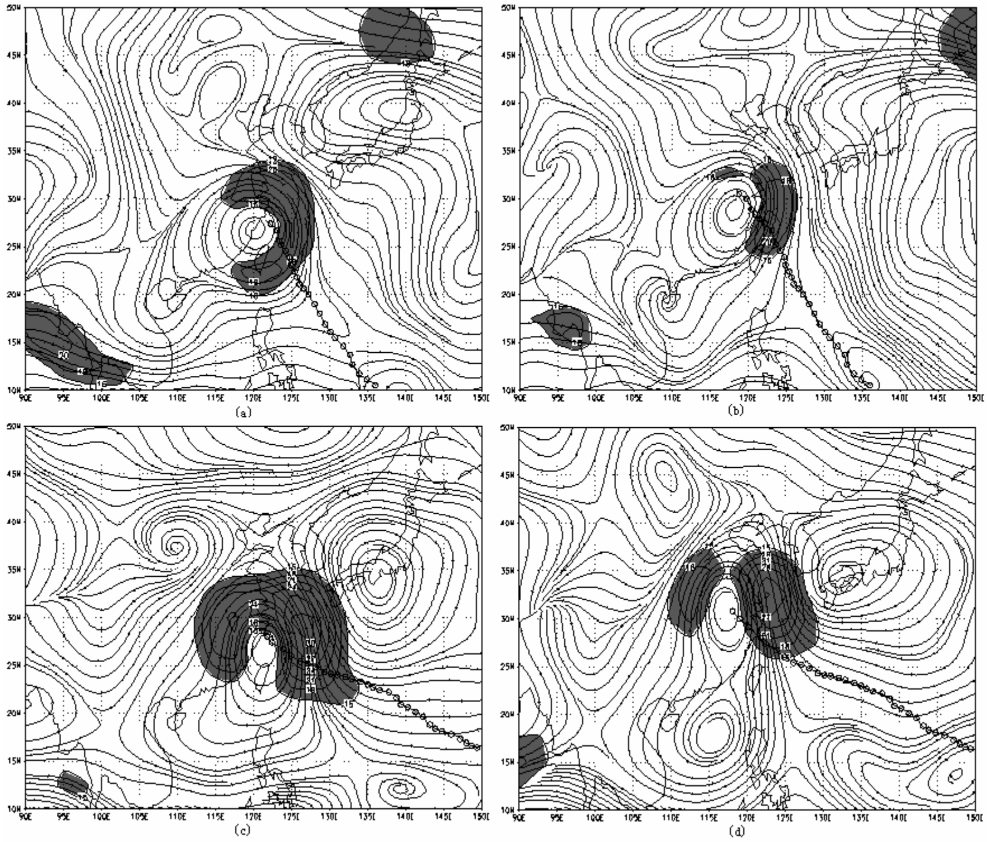

Figure 2. 700 hPa stream fields of forecast sample (Matsa) and similar sample (Winnie). (a): forecast sample for 0200 BST 6 August 2005; (b): forecast sample for 0200 BST 7 August 2005; (c): similar sample for 1400 BST 18 August 1997; (d): similar sample for 1400 BST 19 August 1997. The shades stand for regions where wind speed ≥15m/s, with the numerals for wind speed and unit in m/s. The small circles are the positions of TC centers every 6 hours.

(3) The similarity index defined with the method introduced here in this work is consistent for various criteria. As it is non-linearly distributed, the weights of similar samples are increased while those of dissimilar samples are decreased, thus making it easier to separate the historical samples in terms of similarity and selecting optimal similar samples.

(4) As the criteria for determining similarity directly affects the effect of forecasts, their choice should focus on factors that play roles in or have effects on the storm rainfall of TCs. For NWP, future environmental fields are the criteria playing a key part in the similarity forecast. Our method introduced a domain that moves with the TC for similarity identification. As it conforms to the characteristics of TC motion, the criteria for similarity are more reasonable.

(5) The forecast method succeeds in presenting comprehensive assessment of the effect of interactions between the TC and its environment on the rain brought about by the TC, thus being favorable for the determination of optimal cases of similar historical TCs. It is of course possible that such comprehensive representation could reduce the unique roles played by a particular criterion. As shown in the forecast experiments, this method has some skills in giving site-specific and quantitative forecasts of TC rainfall.

(6) In the composite forecast conducted with a number of similar samples using weights as the similarity index, miss rates have been reduced and overall forecasting accuracy has been improved. However, it smoothes the forecasts for a small number of samples, performing more poorly in extreme situations as compared to individual forecast members

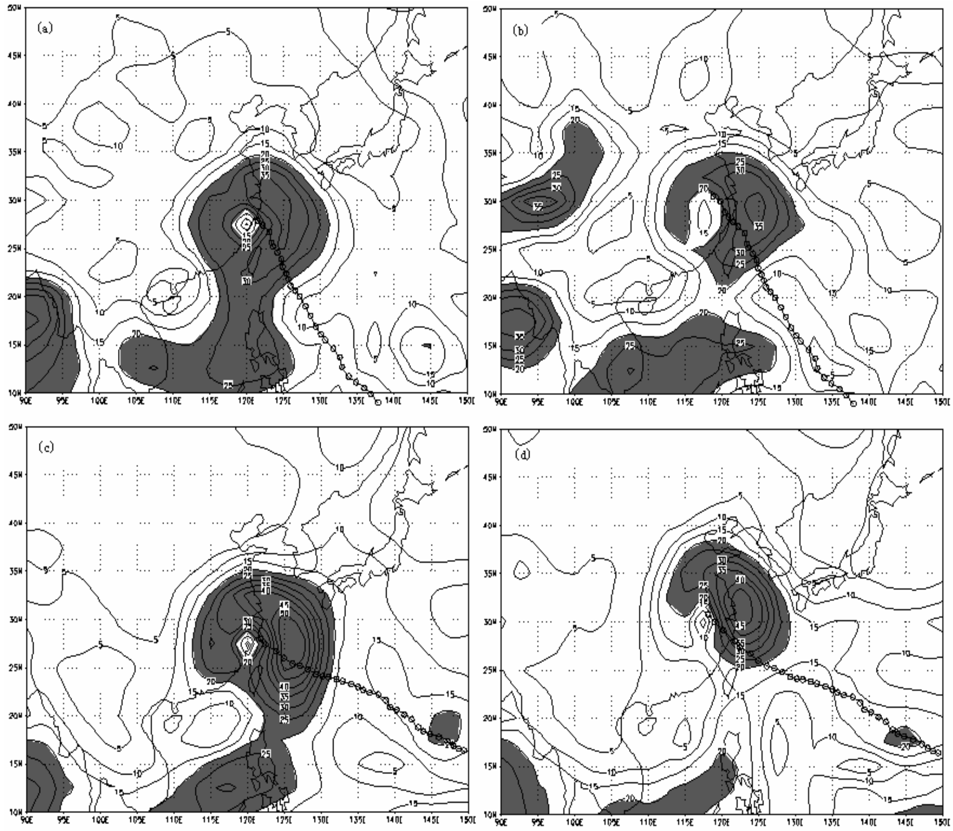

Figure 3. 925 hPa water vapor fluxes for the forecast sample Matsa and similar sample Winnie. Units: g/(s• cm²• hPa). The shades stand for the region where water vapor fluxes ≥20 g/(s• cm• hPa). Other captions are the same as Figure 1.

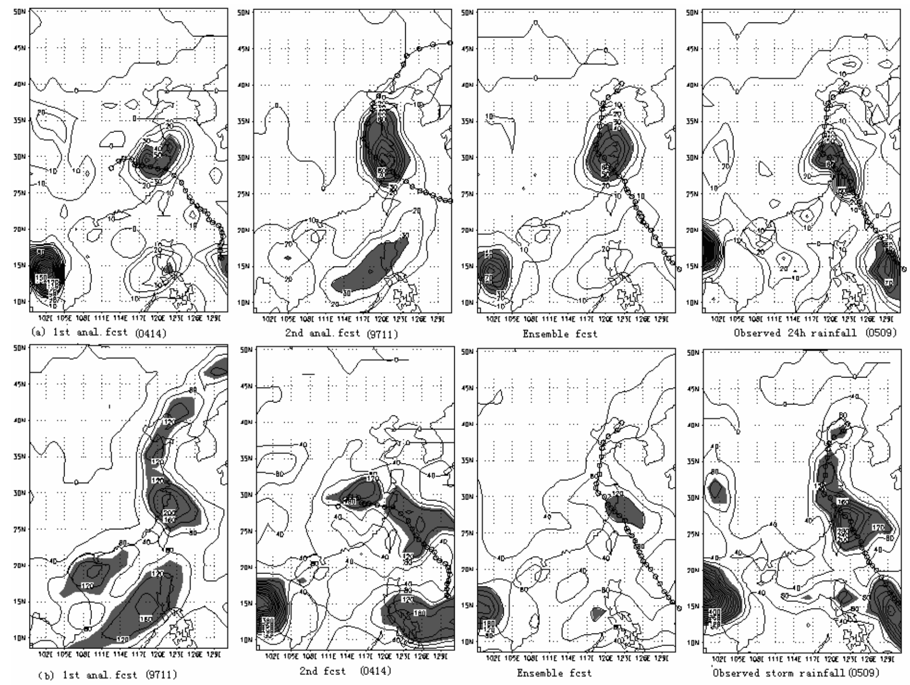

Figure 4. Forecast results of Matsa by the forecast method and observations. (a): 24-hour rainfall with the shades indicating where the regions with rainfall ≥30 mm are; (b): storm rainfall with the shades indicating where the regions with rainfall ≥100 mm are. Small circles indicate the 6-hourly positions of the TC center.

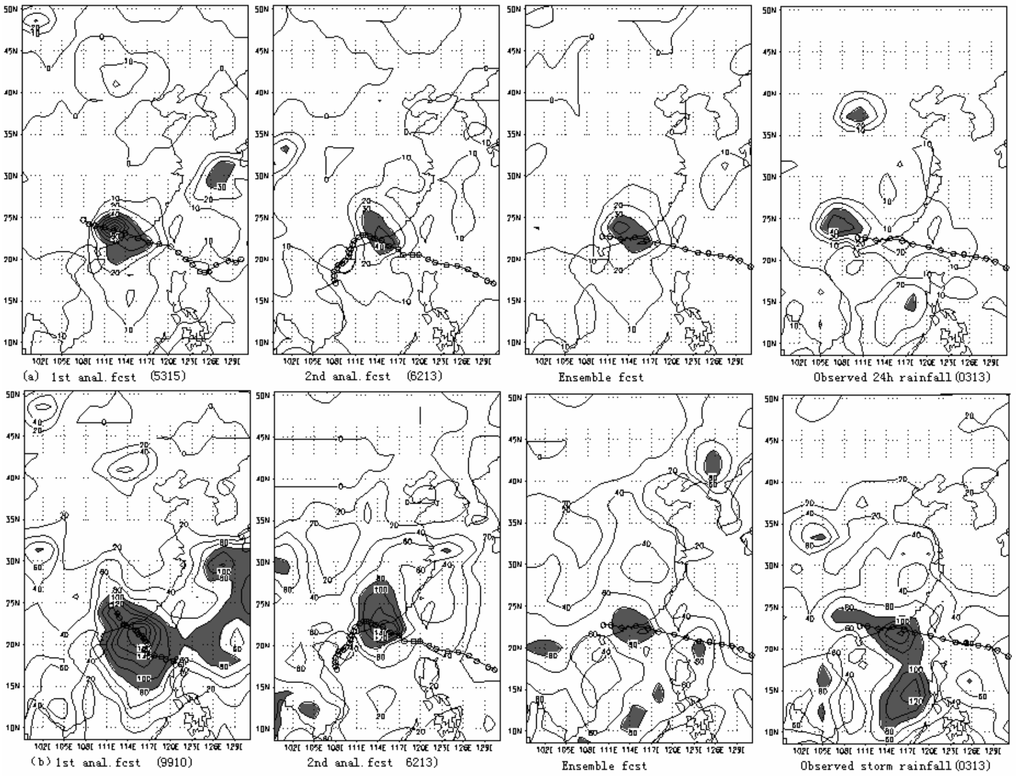

Figure 5. Forecast results of Matsa by the forecast method and observations. (a): 24-hour rainfall with the shades indicating where the regions with rainfall $\geqslant$20 mm are; (b): storm rainfall with the shades indicating where the regions with rainfall $\geqslant$80 mm are. Small circles indicate the 6-hourly positions of the TC center.

REFERENCES:

[1] ANTHES R A. Tropical Cyclones (Their Evolution, Structure and Effects) [M]. LI Yu-fang et al. (translators), Beijing: China Meteorological Press, 1987, 21-56.

[2] ZHU Hong-yan, CHEN Lian-shou, XU Xiang-de. A numerical study of the interactions between typhoon and mid-latitude circulation and its rainfall characteristics [J]. Chin. J. Atmos. Sci., 2000, 24(5): 669-675.

[3] LI Jiang-nan, WANG An-yu, HOU Er-bin, et al. A numerical prediction experiment of track and heavy rainfall round about typhoon Fitow [J]. J. Trop. Oceanogr., 2004, 23(1): 16-24.

[4] LI Zhi-nan, ZHENG Xin-jiang, ZHAO Ya-min, et al. The mesoscale torrential rain cloud cluster from the low pressure periphery area of the typhoon No. 9608 [J]. J. Trop. Meteor., 2000, 16(4): 316-326.

[5] ZHENG Qing-lin, WU Jun, JIANG Ping. Numerical study on the effect of the distribution of the southeast sealine of China on the amplifying of the torrential rain of the landing typhoon 9216 [J]. J. Trop. Meteor., 1996, 12(4): 304-312.

[6] NGUYEN H V, CHEN Y L. High-resolution initialization and simulations of Typhoon Morakot (2009) [J]. J. Trop. Meteor., 2001, 11(4): 452-461.

[7] ZHAO Yu, WU Zeng-mao, LIU Shi-jun, et al. Potential vorticity analysis of an torrential rain triggered by a neutercane in Shandong province [J]. J. Trop. Meteor., 2005, 21(1): 33-43.

[8] HOU Jian-zhong, WANG Chuan, LU Yuan-ping, et al. Activity of typhoons and extreme rainstorms in Shanxi province [J]. J. Trop. Meteor., 2006, 22(3): 229-236.

[9] XU Ai-hua, YE Cheng-zhi, OUYANG Li-cheng, et al. The diagnostic analysis of the track and precipitation of typhoon "Rananim" after landfall [J]. J. Trop. Meteor., 2006, 22(2): 203-208.

[10] CHEN Lian-shou, MENG Zhi-yong. An overview on tropical cyclone research progress in China during the past ten years [J]. Chin. J. Atmos. Sci., 2001, 25(3): 420-432.

[11] CHEN Lian-shou. Research advances on tropical cyclone landfall process [J]. Acta Meteor. Sinica, 2004, 62(5): 541-549.

[12] ZHONG Yuan, JIN Yi-ming. A complex analogue scheme to forecast the rainfall and wind of tropical cyclone [J]. Acta Meteor. Sinica, 1995, 53(Suppl.): 569-577.

[13] KITADE T. Numerical experiments of tropical cyclones on a plane with variable Coriolis parameter [J]. J. Meteor. Soc. Japan, 1980, 58(4): 471-488.

[14] ANTHES R A, HOKE J E. The effect of horizontal divergence and latitudinal variation of the Coriolis parameter on the drift of a model hurricane [J]. Mon. Wea. Rev., 1975, 103(5): 757-763.

[15] HOPE J R, NEUMANN C J. An operational technique for relating the movement of existing tropical cyclones to past tracks [J]. Mon. Wea. Rev., 1970, 98(7): 925-933.

[16] CHAN J C L, GRAY W M. Tropical cyclone movement and surrounding flow relationship [J]. Mon. Wea. Rev., 1982, 110(9): 1354-1374.

[17] ZHONG Yuan, YU Hui, WANG Dong-fa. Effect of environmental field on tropical cyclone track after landfall from East China Sea [J]. J. Trop. Meteor., 2006, 22(4): 314-320.

[18] ZHONG Yuan, YU Hui, WANG Dong-fa, et al. Factors affecting the track of tropical cyclones after landfall in eastern China [J]. J. Trop. Meteor., 2007, 23(3): 209-218.

Citation: ZHONG Yuan, PAN Jin-song, ZHU Hong et al. A similarity scheme for quantitative forecast of precipitation of typhoons. *J. Trop. Meteor.*, 2012, 18(3): 322-331.