ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 32, DECEMBER 2015, 1583–1591

Extended Range (10–30 Days) Heavy Rain Forecasting Study Based on a Nonlinear Cross-Prediction Error Model

XIA Zhiye*[1,2,3], CHEN Hongbin[1], XU Lisheng[1,4], and WANG Yongqian[3]

[1]*Key Laboratory of Middle Atmosphere and Global Environment Observation, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029

[2]*University of Chinese Academy of Sciences, Beijing* 100049

[3]*College of Resources and Environment, Chengdu University of Information Technology, Chengdu* 610225

[4]*Atmospheric Radiation & Satellite Remote Sensing Laboratory, Chengdu University of Information Technology, Chengdu* 610225

(Received 30 November 2014; revised 17 May 2015; accepted 28 May 2015)

ABSTRACT

Extended range (10–30 d) heavy rain forecasting is difficult but performs an important function in disaster prevention and mitigation. In this paper, a nonlinear cross prediction error (NCPE) algorithm that combines nonlinear dynamics and statistical methods is proposed. The method is based on phase space reconstruction of chaotic single-variable time series of precipitable water and is tested in 100 global cases of heavy rain. First, nonlinear relative dynamic error for local attractor pairs is calculated at different stages of the heavy rain process, after which the local change characteristics of the attractors are analyzed. Second, the eigen-peak is defined as a prediction indicator based on an error threshold of about 1.5, and is then used to analyze the forecasting validity period. The results reveal that the prediction indicator features regarded as eigen-peaks for heavy rain extreme weather are all reflected consistently, without failure, based on the NCPE model; the prediction validity periods for 1–2 d, 3–9 d and 10–30 d are 4, 22 and 74 cases, respectively, without false alarm or omission. The NCPE model developed allows accurate forecasting of heavy rain over an extended range of 10–30 d and has the potential to be used to explore the mechanisms involved in the development of heavy rain according to a segmentation scale. This novel method provides new insights into extended range forecasting and atmospheric predictability, and also allows the creation of multi-variable chaotic extreme weather prediction models based on high spatiotemporal resolution data.

Key words: nonlinear cross prediction error, extended range forecasting, phase space

Citation: Xia, Z. Y., H. B. Chen, L. S. Xu, and Y. Q. Wang, 2015: Extended range (10–30 days) heavy rain forecasting study based on a nonlinear cross-prediction error model. *Adv. Atmos. Sci.*, **32**(12), 1583–1591, doi: 10.1007/s00376-015-4252-2.

1. Introduction

Heavy rain is a type of disastrous weather that affects many areas of the world, often triggering landslides, mudslides, floods, urban waterlogging, and many other secondary disasters. As is known, the accuracy of heavy rain forecasting on the 24-h time scale is currently about 20%, on average. Therefore, improving prediction accuracy is an interesting challenge in heavy rain prediction studies, especially at the extended range (10–30 d) scale. Improving accuracy is especially important for disaster prevention and mitigation.

There are three main approaches currently used in heavy rain prediction studies. Firstly, numerical weather prediction (NWP). Some of the methods developed according to this approach include the mesoscale prediction models MM5 and WRF, and the medium-and large-scale numerical models GRAPES (Huang et al., 2013) and AREMS (He et al., 2006). Zhang et al. (2010) adopted an ingredients-based methodology (Doswell et al., 1996) and performed some significant research on the mechanisms of heavy rain formation and prediction validity periods.

Secondly, considering the fact that numerical models are strongly sensitive to the initial atmospheric state and boundary conditions (external forcing), it means that any instability can generate errors during the forecast process, and even minor error from the initial conditions and model can lead to significant loss of forecast ability (Lorenz, 1963a). Epstein (1969) proposed an ensemble method for weather prediction, and demonstrated that this method is an effective way to address error from the initial state and model, compensating for the lack of a forecast validity period (Buizza et al., 2005). Currently, short- and medium-term ensemble prediction mainly applies representative disturbances to the above instabilities to obtain the probability distribution of forecasts.

Regarding the initial conditions, these are mainly ob-

* Corresponding author: XIA Zhiye
Email: xiazhiye@cuit.edu.cn

tained via initial perturbation methods, such as the breeding method (Toth and Kalnay, 1997), singular vector analysis (Molteni et al., 1996), perturbed observation (Houtekamer et al., 1996), 4D-Var assimilation (Gong et al., 1999), conditional nonlinear optimal perturbation (Mu and Jiang, 2008), initial perturbations based on ensemble transformation (Wei et al., 2006), and the nonlinear local Lyapunov vectors method (Feng et al., 2014). In terms of the uncertainty error from the model, it more often utilizes disturbed boundary conditions, or adopts different parameterization schemes according to parameters' sensitivities to prediction objects (Mu et al., 2010). Gao and Cao (2007) and Gao and Ran (2009) also developed a variety of dynamic predictors and integrated them into a heavy rain prediction algorithm using reanalysis datasets, mostly, and found that severe storm rain forecast precision improved on the synoptic and subsynoptic scales. Several studies have revealed that ensemble prediction methods can improve forecasting techniques, but many of these methods require further refinement. Despite great advances in prediction methods, the acceptable prediction validity period of numerical models is currently only about 5 d; forecasting precision decreases rapidly about 10 d later, and little improvement in precision for predictions of 10 or more days can be achieved (Chou and Ren, 2006).

Thirdly, the atmosphere is a complex nonlinear dynamical system; chaos is its inherent characteristic. The continuous accumulation of initial errors can lead to greater uncertainty in the prediction model, and to some extent cannot even be predicted. Lorenz (1963b) showed that the atmospheric predictability limit is about 2 weeks on average, but the precision of extended range (10–30 d) forecasting is beyond that limit. Because prediction accuracy is sensitive to both initial error and the boundary conditions from weather systems, the forecasting model must consider the interaction of these factors, which requires new theory and methodology.

Ding and Li (2009a) and Li and Ding (2009, 2011) introduced the concept of nonlinear error growth dynamics to the spatiotemporal distribution of atmospheric predictability in the case of 500 hPa geopotential height. A number of studies have indicated that atmospheric predictability is as high as 20 days, which is beyond the 2 weeks limit proposed by Lorenz. This work provides a theoretical basis for extended range (10–30 d) heavy rain forecasting. Besides, the low-frequency synoptic chart (LFSC) (Sun et al., 2010) was proposed based on the Madden–Julian Oscillation (MJO) method (Madden and Julian, 1971), but the study area for the extended range scale based on the MJO method is mainly around the equatorial regions. Preliminary results show that the validity period with respect to the LFSC and MJO methods for heavy rain prediction is up to 10–45 d (Sun et al., 2010); these two methods are both regarded as valuable for extended range forecasting, theoretically. However, because the speed of the MJO and the synoptic system is hard to know in advance, these methods are still difficult to apply in practice.

Given that extended range forecasting must consider the chaotic characteristics of the weather system, Krishnamurti et al. (2000) adopted an approach wherein atmospheric variables can be decomposed into two parts, i.e., a component that is less sensitive to the initial atmospheric state and a chaotic component that is sensitive to initial values. Different procedures can be applied to different components, but the chaotic component can currently only be solved in the form of a probability distribution set calculated by historic data. Chou et al. (2010) applied different strategies to address the predictability and chaotic components in 10–30 d extended range weather forecasting, and suggested that combining dynamic and statistical methods is necessary to improve extended range prediction precision.

In summary, general numerical models cannot be applied to extended range scale rainstorm prediction. As such, exploring the chaotic characteristics of the nonlinear weather system is an important endeavor.

In this paper, we develop a novel nonlinear cross-prediction error model (NCPE) based on phase space reconstruction of single-variable chaotic time series of heavy rain. Using nonlinear dynamics and statistical theory, 10–30 d prediction effects are analyzed by evaluating the local dynamic features. The rest of the paper is structured as follows: Section 2 describes the NCPE algorithm in detail. Section 3 details the preprocessing of the NCPE dataset and parameters, tests the NCPE model in 100 heavy rain cases, defines the prediction indicator and analyzes the forecasting validity period. Section 4 makes some comparisons and analysis between heavy rain chaotic systems and stable time series based on the NCPE model. Finally, a summary and further discussion is provided in section 5.

2. Model description

2.1. *Theoretical background*

Climate is a normal nonstationary system. In fact, the hierarchy feature of climate system is the cause to produce nonstationary behaviors, such as chaos system movement, and the nonstationary behaviors of climate process is just the important expression of hierarchy structure. A stranger attractor is the reflection of a chaotic system's movement trajectory projected in phase space; it is also the interaction outcome between its overall stability and local instability (Yang and Zhou, 2005). However, nearly all current theories and methods for climate prediction, including those in statistics and nonlinear sciences, are based on the assumption that the process is stationary, which is in contrast to the nature of climate processes. This contradiction is probably an important cause of the low level of climate prediction.

Chaotic dynamic systems are common in nature; most of these systems cannot be depicted explicitly by dynamic equations, and can only be understood through the available time series data (Liu, 2010). Regarding the sensitivity of a chaotic system to initial error, Eckmann and Ruelle (1985) proved that a system is chaotic if it has at least one positive Lyapunov exponent (LE), which can be used to predict the system variables, by the maximum LE, and then depict the global features of attractors. However, development of the initial error is not identical in all cases, because of the local dynamical characteristics of the trajectory of attractor movement in phase space. Farmer and Sidorowieh (1987) showed that local prediction methods of chaos are better than global ones in the same embedding dimension. Analysis of ENSO temporal evolution data based on global function fitting and LEs showed that better prediction results can be obtained using fewer data compared with other models when chaos time series analysis is adopted (Li and Li, 2007). Ding and Li (2007, 2009b) showed that the atmospheric predictability limit is as high as 20 d in the case of 500 hPa geopotential height, based on the nonlinear local LE developed, and also indicated that nonlinear local error growth is more effective in describing the local structure of the attractor. Specifically, the classic or finite-time LEs are the global LEs, which are established based on the assumption that the initial perturbations are very small that their evolution can be approximately governed by the tangent linear model of a nonlinear system, which essentially belongs to linear error growth dynamics. Clearly, as long as an uncertainty remains infinitely small within the framework of the linear error growth behavior, it cannot pose a limit to predictability. Therefore, nonlinear behavior in error growth should be considered.

In this paper, we analyze the relative dynamical error based on the NCPE model proposed, monitor the local dynamics change characteristics of heavy rain chaotic structure, and diagnose the forecasting validity period.

2.2. *NCPE model*

The general approach undertaken for a given single chaotic variable time series $z(n)$ involves estimation of a parameter $R(i)$ and analysis of its variation within a certain threshold range based on a selection of samples $z_i(n)$. However, characteristics analysis depends on the statistical probability distribution of the parameter, which often causes weakening or annihilation of sensitive features because of the averaging process, such as the run-test method (Forrer and Rotach, 1997). Different to nonlinear time series analysis, it often uses the phase space, time delay and embedding dimension theorems etc. based on the above theorems. It does not calculate the parameter $R(i)$ of $z(n)$ directly, but finds the characteristics of the global attractor $\boldsymbol{Z}$ in phase space indirectly. The trajectory of an attractor based on phase space reconstruction can be described as

$$\boldsymbol{Z} \equiv \{\boldsymbol{z}(n), n = 1, \cdots N_Z\} , \qquad (1)$$

where $\boldsymbol{Z}$ is the phase points set that consists of every phase point $\boldsymbol{z}(n)$, and N_Z is the total number of phase points on the trajectory. For a single phase point in the phase trajectory, $\boldsymbol{z}(n)$ can be denoted by

$$\boldsymbol{z}(n) = \{z(n), z(n+T), \cdots z[n+(m-1)T]\} , \qquad (2)$$

where T and m are the time delay and embedding dimension, respectively. The global attractor $\boldsymbol{Z}$ can be denoted by a "pseudo" state space that models dynamical properties of the true state qualitatively.

Prediction models for dynamical systems can be built based on the above reconstructed attractor, and a series of algorithms have been proposed. The algorithms include the correlation dimension D_{c}, information dimension D_{I}, and the traditional LEs, particularly the space time-index method (Kennel, 1997). These parameters may be used to analyze global variation trends but cannot describe the chaotic local trajectory because of the smooth processing in these methods.

Nichols et al. (2003) made predictions for the dynamics of a damaged structure using attractor-based models of a healthy structure's dynamics. The overall idea in their work was to study the evolution of local neighborhoods of trajectories on the attractor and use the evolved neighborhood for prediction; then, the increase of mean self-prediction error can be seen as the damage of the system. Atmospheric mutations are similar to the above damage structure. However, in this paper, we do not compare an attractor in the baseline structural condition to itself in a subsequent condition; rather, we study the relationships between the different local parts of the attractor, i.e., $r(i, j)$, i and j are the different local parts. The quality of this relationship is measured through a relative dynamical prediction error matrix; the error matrix serves as the abnormal or mutation indicator of heavy rain system.

Conveniently, we reconstruct two attractors, $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$. These can also be regarded as two different local parts of one entire attractor trajectory. A trajectory is randomly selected with index f on one of the reconstructed attractors and named $\boldsymbol{z}_1(f)$. We then create a set ρ including the nearest P points to this trajectory, which appear on the other reconstructed attractor $\boldsymbol{Z}_2$; namely, $\rho \subset \boldsymbol{z}_2(p_q)$ and

$$\boldsymbol{z}_2(p_q) = \{|p_q - f|\rangle w, q = 1, \cdots P\} , \qquad (3)$$

where w is often called the Theiler window. Equation (3) is a purely geometric construction, so that the time indices p_q do not have any temporal relationship to the fiducial time index f. In fact, by requiring that the chosen neighbors be no closer in time than w steps, we are purposefully removing temporal correlations. The number of neighbors to choose depends on the purpose of the model being built; while large neighborhoods tend not to depict local dynamics, and would result in reduced sensitivity to changes, very small neighborhoods may be much more sensitive to noise. An easy rule is to choose the neighborhood $P = N_Z/1000$ (Pecora and Carroll, 1996).

This local neighborhood is then used to predict some evolution time-steps s into the future of what the fiducial trajectory on $\boldsymbol{Z}_1$ will do. The choice of evolution time-steps s must follow the rule that $s < T$, where T is time delay referred to above; generally, s will be chosen as $T/2$ for convenience. Given s, the predictor may be selected as the centroid (or mean) of the time-evolved neighborhood:

$$\boldsymbol{z}_2(s) = \frac{1}{P} \sum_{q=1}^{P} z_2(p_q + s) , \qquad (4)$$

The NCPE may then be computed as

$$r(\boldsymbol{Z}_1,\boldsymbol{Z}_2)=\sqrt{\frac{1}{N_Z-m}}\|\boldsymbol{z}_2(s)-\boldsymbol{z}_1(f+s)\|\ , \tag{5}$$

where the $\|\cdot\|$ operator indicates the Euclidean norm. Figure 1 shows the NCPE procedure; $\boldsymbol{z}_1(f+s)$ and $\boldsymbol{z}_2(s)$ are location 1 and location 2 of the phase trajectory state, respectively, and $r(\boldsymbol{Z}_1,\boldsymbol{Z}_2)$ is a matrix that indicates the local relative dynamic error of attractor evolution.

By looking now at how relationships between various local pairs of attractors on the structure may change as the dynamics change, we can look at a geometric transfer function between pairs, to a certain extent.

In fact, the occurrence of heavy rain is the mutation from a certain angle of the whole weather observation process. In other words, this mutation can typically be seen as local perturbations to the phase structure, and then these relative dynamic changes can be detected by observing how an attractor transfer function model may change across the heavy rain process. This approach is similar to the concept of frequency domain transfer function methods, such as the autoregressive (AR) approach; however, the AR model assumes that the time series is a linear combination of past values of itself, or can be depicted by the following equation:

$$x(n+1)=\sum_{k=0}^{M}a_kx(n-k)\ , \tag{6}$$

where $x(n)$ represents observed measurements and a_k is the AR coefficients. NCPE may be used to detect local dynamic change characteristics of the chaotic attractor in a spatial domain; this capability is beneficial to diagnosing the chaotic process of heavy rain. In particular, when the entire attractor is split into different continuous parts (or segments) in the length scale L, $\{\boldsymbol{z}_i,\boldsymbol{z}_j,1,\cdots i,j\cdots N_L\}$, N_L is the total segments number. In general, $r(\boldsymbol{Z}_i,\boldsymbol{Z}_j)\neq r(\boldsymbol{Z}_j,\boldsymbol{Z}_i)$, or $r(\boldsymbol{z}_i,\boldsymbol{z}_j)\neq r(\boldsymbol{z}_j,\boldsymbol{z}_i)$, and it is proved that chaos is irreversible; $r(\boldsymbol{z}_i,\boldsymbol{z}_i)$ is defined as the diagonal nonlinear cross-prediction error (DNCPE) when $i=j$, which is also the main content of this paper.

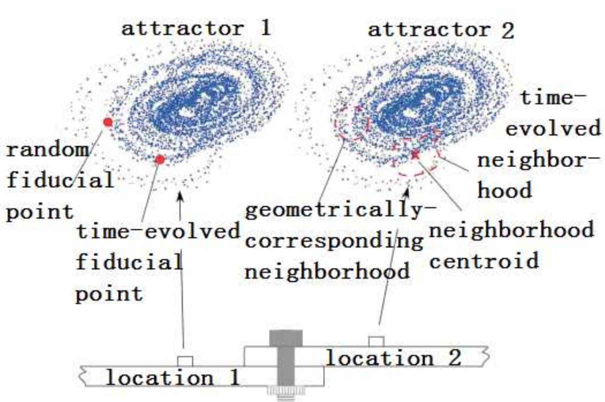

Fig. 1. The nonlinear cross-prediction error process.

3. Case study and analysis

3.1. *Data quality control and preprocessing*

One hundred heavy rain cases globally are selected as research objects in this paper, based on their reported damage caused, from websites and examples in other articles (e.g., Sun et al., 2010). The cases are mainly in China, Japan and India in Asia; America, Canada and Mexico in North America; England, France, Germany and Greece in Europe; and countries in equatorial regions. The test data for the NCPE model comprise a single-variable time series of precipitable water (PWAT), from the four-times-daily National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis dataset. PWAT is the measure of the depth of liquid water at the surface that would result after precipitating all of the water vapor in a vertical column over a given location, usually extending from the surface to 300 hPa; it is not the realistic rainfall amount, but can still depict the characteristics of rainfall to some extent.

The reanalysis grid has a global spatial coverage of 2.5×2.5, or 144×73 grid points. The rules for selecting the PWAT time series in this paper are as follows: the initial length of the PWAT series is about 31 days or more; a length of 5 days is also selected after a heavy rain case; and the PWAT data are from the single grid that covers the heavy rain location in latitude/longitude coordinates. Only in this way can we reveal the evolution mechanism of heavy rain using the NCPE in phase space.

Data preprocessing includes five steps that should be carried out before inclusion in the NCPE model. In step 1, optimal interpolation of the raw data is conducted. The length of the initial time series is so short that it cannot meet the chaos analysis rules; for example, phase space reconstruction (Wolf et al., 1985). Then, interpolation processing is required and, meanwhile, the chaotic structure of the interpolated series cannot be varied. High-order nonlinear fractal interpolation is regarded as an effective method of interpolation, which can obtain the optimal interpolation times (OIT) at the restriction of metric entropy (Xia and Xu, 2010); because metric entropy is scaled by the fractal dimension D of the series, the invariant D can prove that the chaotic structure is unchanged during the interpolation process.

Figure 2 shows the metric entropy relative error surface along with segments and interpolation times for a chaotic time series sample. It can be seen that the chaotic structure approximation of the initial series is not done by any interpolation time, but by optimal times. In this situation, the relative error of metric entropy is the least and uncorrelated with segments. So, in Fig. 2 for example, whose OIT is 22, it can be seen that the interpolated structure is most similar to the original only by the interpolation times of 22; the following steps for calculation are based on the optimal interpolated data.

In step 2, phase space reconstruction is performed based on Takens (1981) theorem. Here, the interpolated time series data are vectorized into m dimension phase space. In step 3, the time delay T is calculated using mutual information

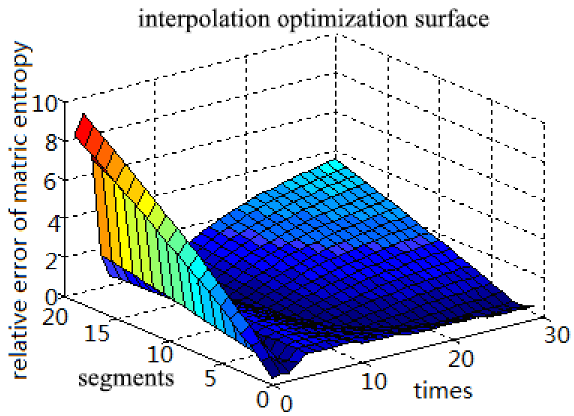

Fig. 2. The 3D metric entropy error surface, relative error versus segment and interpolation times. There is a location whose errors are less under the limit of segment number and interpolation times.

(Kantz and Schreiber, 1997). In step 4, the G-P algorithm (Ledrappier, 1981) is applied to calculate the correlation dimension D; parameter D depicts the fractal structure of the series. In step 5, the embedding dimension m is calculated based on the false nearest neighbor method (Kennel et al., 1992), and it is ensured that $m \geqslant 2D+1$.

It is noted that how to choose the embedding dimension m is also important to the phase reconstruction, and m is selected in this paper by the empirical theory (Andrew and David, 2005). The meteorological elements sensitivity analysis of different extreme weather to the NCPE model, and also the effects from the initial error, external forcing and parameter error, will be discussed in a future paper; we only give parameters here.

3.2. *Cases validation and analysis*

As an example, heavy rain occurred on 21 July 1996 in Chongqing, China. Its total precipitation was 206 mm, and the raw data length was $31 \times 4 = 124$, it's dimensionless quantity. The parameters calculated through the above steps, 1–5, are that OIT $= 43$, time delay $T = 16$, also dimensionless quantity. embedding dimension $m = 4$, maximum LE $\lambda_1 = 0.1128 > 0$; λ_1 greater than zero shows that the raw data PWAT are chaotic. For convenience, the entire attractor projected by the interpolated data in 4-dimensinonal phase space is split into 31 continuous segments, i.e., one phase trajectory segment corresponds to one day's state for the atmosphere, which reflects the phase spatial variation of the heavy rain chaotic system every day. As the DNCPE in Fig. 3 shows, there is a maximum error of DNCPE at segment 11, equal to 2.49. We consider the value 1.5 as the threshold in this paper, based on the calculation of 100 heavy rain cases globally. That is to say, the DNCPE errors greater than 1.5 are regarded as eigen-peaks for heavy rain prediction. In Fig. 3, the heavy rain occurs in segment 21, and the prediction eigen-peak is in segment 11, which is referred to as the most unstable transition zone of the total attractor about the chaotic system. We define this eigen-peak as the prediction indicator, and therefore, the forecasting validity period $t = 21 - 11 = 10$ d, approximately, in this case.

Specifically, the entire attractor projected into phase space from the interpolated data is split into 31 continuous segments, i.e., one phase trajectory segment corresponds to one day's state for the atmosphere. It can also be split into more segments, such as 62 based on some rule. These continuous segments are equivalent to $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ or location 1 and location 2; in other words, there are 31 locations corresponding to the entire attractor on this condition. Therefore, it is very meaningful to explore local dynamic characteristics by observing individual or mutual relationships of these 31 segments.

In detail, we again stress that the main point of this paper is to exploit the information contained in the relative dynamic error $r(\boldsymbol{z}_i, \boldsymbol{z}_j)$, in addition to that contained in the diagonal terms $r(\boldsymbol{z}_i, \boldsymbol{z}_i)$. The NCPE matrix of the Chongqing case described above is shown in Fig. 4. For the relative dynamic error of every phase trajectory segment (y-axis) on the prediction database of one other phase trajectory

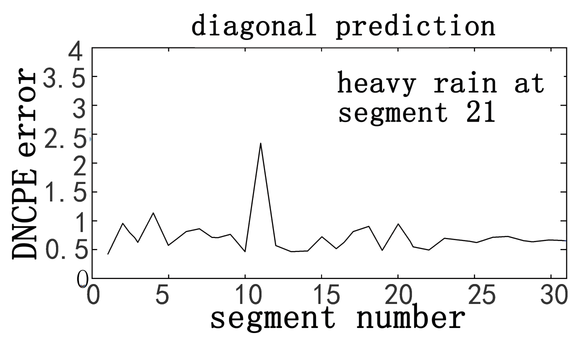

Fig. 3. The DNCPE for 21 July 1996, Chongqing, China. Parameters: OIT $= 43$; time delay $T = 16$; embedding dimension $m = 4$. Eigen-peak is in segment 11 and heavy rain is in 21; forecasting validity period is 10 d.

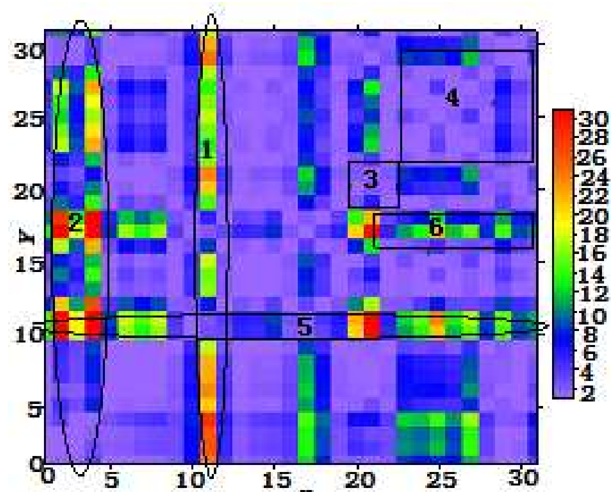

Fig. 4. The NCPE matrix for 21 July 1996, Chongqing, China.

segment (x-axis), the numbers $1, 2, 3, 4, 5$ and 6 in the matrix map represent six locations, respectively. The diagonal, whose slope is 1, represents the DNCPE, and the remaining values in the matrix map represent the NCPE. As in Fig. 3, the matrix is asymmetric and shows that the chaotic system is irreversible again.

Location 1 is the NCPE on the database of phase trajectory segment 11. The relative error to the y-axis is much greater overall, indicating that this zone is most unstable in the phase trajectory, which can explain the prediction indicator characteristics. Location 5 depicts the NCPE in segment 11 on the database of every phase segment (x-axis). The relative error is also greater overall, the same as location 1. Location 2 for the NCPE is in the database phase trajectory segment 1–5 zone. The relative error is large overall, and this result can be explained by Simmons et al. (1995) in that the effects of the initial state error are apparent in the initial stage of modern NWP models, but the role of error from models will become more important along with the prediction validity period growth. The relative dynamical error at locations 3 and 4 stand for before and after the rainstorm, respectively; the NCPE is relatively small and smooth compared to other sections, indicating that the local structure of the attractor is relatively stable in these two sections. In other words, the mutation of the local relative dynamical structure of the attractor may appear long before the heavy rain period, such as location 1. Location 6 spans before and after the heavy rain. The relative dynamical error is a gradually reducing process in the x-axis direction, which indicates that change in the local structure of the attractor in this period is relatively slow with no mutation.

By the same argument, Fig. 5 is the DNCPE of a heavy rain event in Xinyang, China. There were two heavy rain cases on 1 July and 24–25 July 2007, and the total precipitation was 226 mm in the 1 July case. The parameters calculated are OIT $= 50$, time delay $T = 19$, embedding dimension $m = 4$, and $\lambda_1 = 0.0833$. The eigen-peaks are in segments 8 and 30, respectively, and the two heavy rain cases are in segments 17 and 40, respectively, so the forecasting validity period is 9 d and 10 d, respectively. The DNCPE can predict these two cases accurately, without omission.

There is an interesting phenomenon regarding the DNCPE shown in Fig. 6, for heavy rain in Koblenz, Germany, on 1 April 2003 (total precipitation: 50 mm). The parameters calculated are OIT $= 47$, time delay $T = 26$, embedding dimension $m = 4$, and $\lambda_1 = 0.1530$. The two peaks are distributed in segments 5 and 14, respectively, and the heavy rain cases in segment 26. The peak in segment 14 is seen as the eigen-peak without considering the peak at 5 based on the above view from Simmons et al. (1995), so the forecasting validity period is 12 d.

Cases also exist that are not interpreted well by the present NWP model, i.e., false alarms or omissions. We also test this phenomenon based on the NCPE model, such as the results shown in Fig. 7, which shows a heavy rain case that occurred in Medellín, Columbia, on 20 November 2004 (total precipitation: 96 mm). The parameters are OIT $= 44$, time delay $T = 21$, embedding dimension $m = 4$, and $\lambda_1 = 0.177$. Three peaks are distributed in segments 12, 15 and 20, respectively; two or more peaks indicate that the chaotic attractor structure of this heavy rain case is complicated. In this situation, the experiential rule for selecting the prediction eigen-peak is to select the first peak that appears after segment 5. So, the peak in segment 12 is regarded as an eigen-peak, heavy rain is in segment 27, and then the fore-

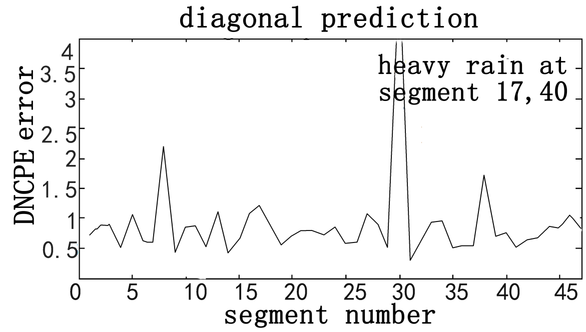

Fig. 5. The DNCPE for 1 July and 24–25 July 2007, Xinyang, China. Parameters: OIT $= 50$; time delay $T = 19$; embedding dimension $m = 4$. Eigen-peak is in segment 8 and 30, respectively, and the two heavy rain cases in segment 17 and 40; forecasting validity period is 9 d and 10 d, respectively.

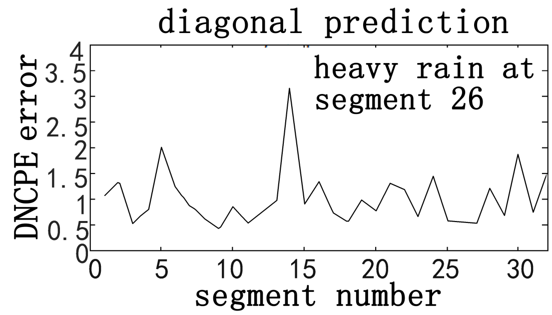

Fig. 6. The DNCPE for 1 April 2003, Koblenz, Germany. Parameters: OIT $= 47$, time delay $T = 26$; embedding dimension $m = 4$. Eigen-peak is in segment 14 and heavy rain is in 26; forecasting validity period is 12 d.

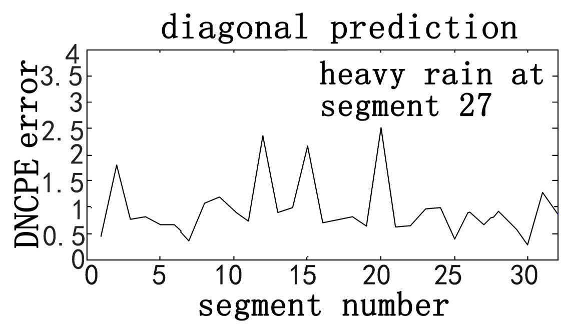

Fig. 7. The DNCPE for 20 November 2004, Medellín, Columbia. Parameters: OIT $= 44$; time delay $T = 21$; embedding dimension $m = 4$. Eigen-peak is in segment 12 and heavy rain is in 27; forecasting validity period is 15 d.

casting validity period is 15 d.

4. Stable time series comparisons and statistical analyses

In general, time series contain two types, stable and unstable series, which differ greatly in their motion characteristics (Brockwell and Davis, 2001). Unstable time series are also seen as nonlinear dynamical systems, such as heavy rain extreme weather. So, we also test two other stable time series by the NCPE model. Shown in Fig. 8 is the DNCPE of Gaussian white noise: data length of 9200; parameter time delay $T = 15$; and embedding dimension $m = 3$. $\lambda_1 = 1.2927 > 1$ indicates that Gaussian white noise may induce chaos, but it is effected by noise greatly, i.e., signal-to-noise ratio. The prediction time length for the above Gaussian white noise series is short, comparatively, and the DNCPE is smooth, which shows that Gaussian white noise does not have mutation features, even projected in phase space.

Let us convert to other extreme weather cases by using the NCPE model; drought, for example. Chuxiong is a severe drought area in China. The DNCPE based on PWAT is shown in Fig. 9, through the same processing. The parameters calculated are $T = 19$, $m = 4$; $0 < \lambda_1 = 0.0293 < 1$ shows that the drought weather system is still chaotic and can be predicted for the PWAT variable. While the DNCPE is smooth without an obvious eigen-peak, the result may be affected by the following aspects. Firstly, the NCPE model may not be more sensitive to drought forecasting when using the PWAT database; other meteorological elements such as temperature or pressure will be tested and discussed in a future paper. Secondly, it may also be related to the effect of the time scale of variables in the NCPE model.

One hundred heavy rain cases are analyzed through the above same calculation based on the NCPE model. The results show that the heavy rain cases tested are completely chaotic based on LE. The prediction validity periods for the above 100 heavy rain cases are shown statistically in Fig. 10. The short range, medium range and extended range for weather prediction are often regards as 1–2 d, 3–9 d, and 10–30 d respectively now. In this tests, the prediction validity periods for 1–2 d, 3–9 d and 10–30 d are 4, 22 and 74 cases, respectively, with no false alarms or omissions. Note that the meaning of no false alarms or omissions does not represent the prediction of the location and amount of precipitation, but represents the prediction indicator.

There are 74 cases that reach the time scale for the 10–30 d extended range, but the remaining 26 cases belong to the short- and medium-term time scale. The prediction validity period exists over a time span of 1–30 d. The phenomenon of the time span may be related to the difference from the chaotic structure of individual heavy rain cases, or can possibly be explained by the fact that the atmospheric predictability limit has a spatiotemporal distribution difference (Ding and Li, 2009b).

Therefore, the heavy rain process can be considered as a collection of unstable signals of time series data, and weather may be viewed as a complicated nonlinear system that combines the effects of both stable and unstable processes. The segment size is determined by the trade-off between the statistical stability of $r(\mathbf{z}_i, \mathbf{z}_j)$ for long segments and a finer time resolution for shorter segments. A slight advantage may be gained using overlapping segments.

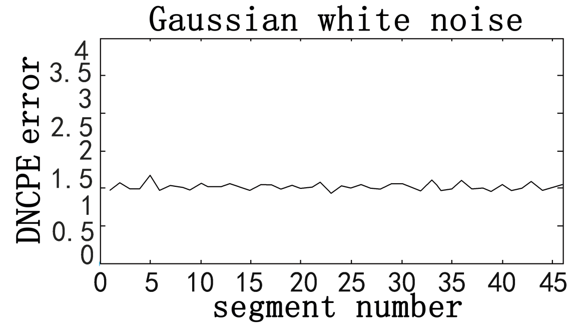

Fig. 8. The DNCPE for Gaussian white noise. Parameters: time delay $T = 15$; embedding dimension $m = 3$. $\lambda_1 = 1.2927 > 0$ means it is chaotic without local mutation on the trajectory of the attractor.

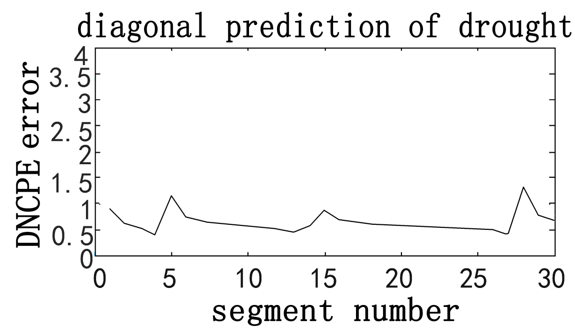

Fig. 9. The DNCPE of drought in Chuxiong, China. Parameters: $T = 19$; $m = 4$; $0 < \lambda_1 = 0.0293$; without obvious eigen-peak.

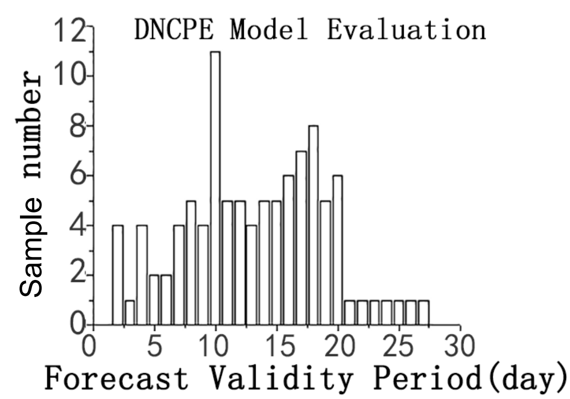

Fig. 10. Cartogram of the prediction validity period for 100 heavy rain cases globally. Prediction validity period is 1–2 d, 3–9 d and 10–30 d in 4, 22 and 74 cases, respectively, without false alarms or omissions.

5. Conclusion and discussion

An NCPE model is developed in this paper using single-variable time series of a chaotic system combined with nonlinear dynamics and statistics based on phase space. The NCPE model may be used to calculate nonlinear cross error of attractor local pairs, depict the local dynamical change features in the attractors, and evaluate the development mechanism for a heavy rain chaotic process. Prediction validity periods for 1–2 d, 3–9 d and 10–30 d occur in 4, 22 and 74 cases, respectively, without false alarms or omissions. Preliminary results based on the 100 rainstorm samples show that the NCPE model can achieve the heavy rain medium and 10–30 d extended range forecasting, which provides a basis for extended range (10–30 d) heavy rain predictability.

Besides, because the data length of variables and segment numbers are referred to in the NCPE model, the NCPE model can diagnose the chaotic movement characteristics of heavy rain at different time scales through the trade-off between the variable date length and segment numbers. This novel processing can also provide a new approach for developing multi-variable chaotic weather system prediction algorithms based on high-resolution spatiotemporal data.

Theoretically, the prediction validity period of the NCPE model will be more stable when coupled with chaotic multi-variables. Single-variable series data based on phase construction are commonly applied to nonlinear chaotic trajectory analysis, but this method presents several limitations (Li and Chou, 1996). A single variable cannot be used to analyze the nonlinear dynamical features of the atmosphere perfectly, which is affected by multiple variables. Ding and Li (2007) calculated the nonlinear local LE based on an N-dimensional chaotic system, but could not distinguish the contribution from every error vector because of computation precision limitation. This finding reveals that calculating the variation rate in m-dimensional space is difficult, although the Gram–Schmidt Orthogonalization method is believed to be theoretically able to do so.

How, then, can multivariable chaotic elements be coupled into the NCPE model to improve extended range prediction precision? Sensitivity analysis of the effects of meteorological elements of different extreme weather conditions to the NCPE model, as well as effects from initial error, external forcing and other parameter errors, will be discussed in a future report.

Acknowledgements. Funding for this research was provided by the National Natural Science Foundation of China (Grant Nos. 41275039 and 41471305) and the Preeminence Youth Cultivation Project of Sichuan (Grant No. 2015JQ0037).

REFERENCES

Andrew, H., and E. David, 2005: Nonlinear dynamical analysis of noisy time series. *Nonlinear Dynamics, Psychology and Life Sciences*, **9**, 399–433.

Brockwell, P. J., and R. A. Davis, 2001: *Time Series: Theory and Methods*. Springer Press, 596 pp.

Buizza, R., P. L. Houtekamer, G. Pellerin, Z. Toth, Y. J. Zhu, and M. Z. Wei, 2005: A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**(5), 1076–1097.

Chou, J. F., and H. L. Ren, 2006: Numerical weather prediction–necessity and feasibility of an alternative methodology. *Journal of Applied Meteorological Science*, **17**(2), 240–244. (in Chinese)

Chou, J. F., Z. H. Zheng, and S. P. Sun, 2010: The think about—10-30 d extended-range numerical weather prediction strategy-facing the atmosphere chaos. *Scientia Meteorologica Sinica*, **30**(5), 569–573. (in Chinese)

Ding, R. Q., and J. P. Li, 2007: Nonlinear finite-time Lyapunov exponent and predictability. *Phys. Lett. A.*, **364**, 396–400.

Ding, R. Q., and J. P. Li, 2009a: Application of nonlinear error growth dynamics in studies of atmospheric predictability. *Acta Meteorologica Sinica*, **67**(2), 241–249. (in Chinese)

Ding, R. Q., and J. P. Li, 2009b: The temporal spatial distributions of weather predictability of different variables. *Acta Meterologica Sinica*, **67**(3), 343–354. (in Chinese)

Doswell, C. A., H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, **11**, 560–581.

Eckmann, J. P., and D. Ruelle, 1985: Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, **57**, 617–656.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, **21**, 739–759.

Farmer, J. D., and J. J. Sidorowieh, 1987: Predicting chaotic time series. *Phys. Rev. Lett*., **59**(8), 845–848.

Feng, J., R. Q. Ding, D. Q. Liu, and J. P. Li, 2014: The application of nonlinear local Lyapunov vectors to ensemble predictions in the Lorenz systems. *J. Atmos. Sci.*, **71**(9), 3554–3567.

Forrer, J., and M. W. Rotach, 1997: On the turbulence structure in the stable boundary layer over the Greenland ice sheet. *Bound.-Layer Meteor.*, **85**, 111–136.

Gao, S. T., and J. Cao, 2007: Physical basis of generalized potential temperature and its application to cyclone tracks in nonuniformly saturated atmosphere. *J. Geophys. Res.*, **112**, D18101, doi: 10.1029/ 2007JD008701.

Gao, S. T., and L. K. Ran, 2009: Diagnosis of wave activity in a heavy-rainfall event. *J. Geophys. Res.*, **114**, D08119, doi: 10.1029/2008JD010172.

Gong, J. D., W. J. Li, and J. F. Chou, 1999: Forming proper ensemble forecast initial members with four-dimensional variational data assimilation method. *Chinese Science Bulletin*, **44**(16), 1527–1531.

He, G. B., J. Chen, Y. H. Xiao, Q. Y. Gu, and C. Li, 2006: Precipitation prediction with AREM numerical model in Sichuan flood season in 2005. *Meteorological Monthly*, **32**(7), 64–71. (in Chinese)

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.

Huang, Y. Y., J. S. Xue, Q. L. Wan, Z. T. Chen, W. Y. Ding, and C. Z. Zhang, 2013: Improvement of the surface pressure operator in GRAPES and its application in precipitation forecasting in south China. *Adv. Atmos. Sci.*, **30**(2), 354–366, doi: 10.1007/s00376-012-1270-1.

Kantz, H., and T. Schreiber, 1997: *Nonlinear Time Series Analysis.* Cambridge University Press, 388 pp.

Kennel, M. B., 1997: Statistical test for dynamical nonstationarity in observed time-series data. *Phys. Rev. E*, **56**(1), 316–321.

Kennel, M. B., R. Brown, and H. D. I. Abarbanel, 1992: Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A*, **45**, 3403–3411.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196–4216.

Ledrappier, F., 1981: Some relations between dimension and Lyapounov exponents. *Commun. Math. Phys.*, **81**, 229–238.

Li, J. P., and J. F. Chou, 1996: Some problems exited in estimating fractal dimension of attractor with one dimensional time series. *Acta Meteorologica Sinica*, **54**(3), 312–323. (in Chinese)

Li, J. P., and R. Q. Ding, 2009: Studies of predictability of single variable from multi-dimensional chaotic dynamical system. *Chinese J. Atmos. Sci.*, **33**(3), 551–556. (in Chinese)

Li, J. P., and R. Q. Ding, 2011: Temporal–spatial distribution of atmospheric predictability limit by local dynamical analogs. *Mon. Wea. Rev.*, **139**, 3265–3283.

Li, K. Y., and X. D. Li, 2007: Nonlinear prediction of ENSO. *Acta Scientiarum Naturalium Universitatis Pekinensis*, **43**(1), 30–34. (in Chinese)

Liu, Z., 2010: Chaotic time series analysis. *Mathematical Problems in Engineering*, Vol. 2010, Article ID 720190, 31 pp.

Lorenz, E. N., 1963a: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–148.

Lorenz, E. N., 1963b: Section of planetary sciences: The predictability of hydrodynamic flow. *Transactions of the New York Academy of Sciences*, **25**, 409–432.

Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci*., **28**(5), 702–708.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.

Mu, M., and Z. N. Jiang, 2008: A new approach to the generation of initial perturbations for ensemble prediction: Conditional nonlinear optimal perturbation. *Chinese Science Bulletin*, **53**(13), 2062–2068.

Mu, M., W. Duan, Q. Wang, and R. Zhang. 2010: An extension of conditional nonlinear optimal perturbation approach and its applications. *Nonlin. Processes Geophys.*, **17**, 211–220.

Nichols, J. M., M. D. Todd, M. Seaver, and L. N. Virgin, 2003: Use of chaotic excitation and attractor property analysis in structural health monitoring. *Phys. Rev. E*, **67**, 016209.

Pecora, L. M., and T. L. Carroll, 1996: Discontinuous and non-differentiable functions and dimension increase induced by filtering chaotic data, *Chaos*, **6**, 432–439.

Simmons, A. J., R. Mureau, and T. Petroliagis, 1995: Error growth and estimates of predictability from the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **121**, 1739–1771.

Sun, G. W., F. Xin, C. Y. Kong, B. M. Chen, and J. H. He, 2010: Atmospheric low-frequency oscillation and extended range forecast. *Plateau Meteorology*, **29**(5), 1142–1147. (in Chinese)

Takens, F. 1981: Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*. *Lecture Notes in Mathematics*, D. Rand and L. S. Young, Eds., Springer, 366–381.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.

Wei, M., Z. Toth, R. Wobus, Y. J. Zhu, C. H. Bishop, and X. G. Wang, 2006: Ensemble transform kalman filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus A*, **58**, 28–44.

Wolf, A., J. B. Swift, H. L. Swinney, and J. A. Vastano, 1985: Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, **16**, 285–317.

Xia, Z. Y., and L. S. Xu, 2010: Research on optimal interpolation times of nonlinear time-series using metric entropy and fractal interpolation. *International Workshop on IWCFTA*, Kunming, Yunnan, 411–415.

Yang, P. C., and X. J. Zhou, 2005: On nonstationary behaviors and prediction theory of climate systems. *Acta Meterologica Sinica*, **63**(5), 556–570. (in Chinese)

Zhang, X. L., S. Y. Tao, and J. H. Sun, 2010: Ingredients-based heavy rainfall forecasting. *Chinese J. Atmos. Sci.*, **34**(4), 754–766. (in Chinese)