Science in China Series D: Earth Sciences

Strategy and methodology of dynamical analogue prediction

REN HongLi[1,2†] & CHOU JiFan[1]

[1] College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China;
[2] Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing 100081, China

In order to effectively improve numerical prediction level by using current models and data, the strategy and methodology of dynamical analogue prediction (DAP) is deeply studied in the present paper. A new idea to predict the prediction errors of dynamical model on the basis of historical analogue information is put forward so as to transform the dynamical prediction problem into the estimation problem of prediction errors. In terms of such an idea, a new prediction method of final analogue correction of errors (FACE) is developed. Furthermore, the FACE is applied to extra-seasonal prediction experiments on an operational atmosphere-ocean coupled general circulation model. Prediction results of summer mean circulation and total precipitation show that the FACE can to some extent reduce prediction errors, recover prediction variances, and improve prediction skills. Besides, sensitive experiments also show that predictions based on the FACE are evidently influenced by the number of analogues, analogue-selected variables and analogy metric.

dynamical analogue prediction, prediction strategy, analogue correction of errors, extra-seasonal prediction

1 Introduction

Numerical weather forecasting and short-term climate prediction are quickly developed by the following serial improvements of observation data and models. But their capability for practical application is yet unsatisfactory and predictive performance still needs to be promoted[1−3]. In the past ten years, prediction strategy and methodology based on numerical model have been increasingly developed and become an important approach for improving prediction, in which many related researches were carried out in domestic and overseas works[4]. Thus, under the current conditions of models and data, it will have important values to deeply develop and innovate in prediction strategy and methodology.

Actually, the essential idea in prediction strategy and methodology research is to seek the combination of dynamical and statistical methods. At present, meteorologists have come to an agreement on combining the numerical model with statistical experience. But most important, how to effectively combine them needs further study. Early in the 1950s, Gu put forward the importance and feasibility of introducing historical data into numerical prediction[5,6]. In fact, error correction techniques employed in short-term climate prediction are just based on statistical features between hindcasts and observations[7−9]. In order to realize the combination of dynamical and statistical methods in nature, Chinese meteorologists have proposed many innovative prediction methods, such as the methods using multi-time historical data[10,11], the analogue-dynamical methods[12−15], and the methods based on atmospheric self-memorization principle[16,17], and so on. Although these methods can provide new approaches for improving numerical prediction level, further works will be still necessary for

Received April 11, 2007; accepted June 4, 2007
doi: 10.1007/s11430-007-0109-6
†Corresponding author (email: renhl@cma.gov.cn)
Supported jointly by the National Natural Science Foundation of China (Grant Nos. 40675039 and 40575036) and the National Science and Technology Support Program of China (Grant No. 2006BAC02B04)

their operational applications. In current days, numerical models are being gradually perfected and historical data are being abundantly accumulated. Then, how can the above-mentioned prediction ideas with theoretical advantages be applied to these abundant data and advanced models? This is the basic motivation of related studies in the present paper, which has significant values and meanings.

In recent works, the concept and methodology of dynamical analogue prediction (DAP) has been put forward, and its validity has been also documented by primary experiments[18,19]. The strategy and methodology of the DAP will be further studied in this paper. A new prediction method will be developed by introducing the new idea to predict the prediction errors of dynamical model based on historical analogue information, and the extra-seasonal prediction experiments on an operational model will be conducted.

2 Strategy of dynamical analogue prediction

Generally, numerical prediction is mathematically put forward in terms of the initial value problem of partial differential equations and can be expressed as

$$\frac{\partial \psi}{\partial t} + L(\psi) = 0, \qquad (1)$$
$$\psi(r, t_0) = \psi_0(r),$$

where $\psi(r,t)$ is the model state vector to be predicted, r is the vector in the spatial coordinate, t is time, and L is the differential operator of ψ, which is corresponding to real numerical model and usually nonlinear. t_0 is initial time and ψ_0 is initial value. When $t > t_0$, ψ or their functions $P(\psi)$ may be obtained by numerically integrating initial values. Similarly, the exact model that real atmosphere satisfies can be written as

$$\frac{\partial \psi}{\partial t} + L(\psi) = E(\psi), \qquad (2)$$

in which E, as the functions of ψ, stands for the errors between real atmospheric process and the dynamical process described by model, and just reflects the unknown total errors of numerical model, namely model errors. Then from point of view of dynamics, historical observed data may be regarded as a series of special solutions $\breve{\psi}$ or their functions $\breve{P}(\psi)$ of eq. (2).

In general, based on theoretical and experimental researches, a great many complicated numerical models are continually developed in order to reduce model errors $E(\psi)$ by improving dynamical frameworks, physical processes, and so on. But anyway, there always objectively exists considerable errors in models. Consequently, according to the idea of studying prediction strategy, model errors $E(\psi)$ can be estimated and reduced by utilizing the information of historical data in terms of inverse problem under the condition of existing models.

For any initial value ψ, in order to pertinently select observed data suitable for estimating model errors, one may consider to use the model error information provided by historical analogue $\tilde{\psi}$ similar to ψ. Compared with the traditional statistical analogue prediction[20], the concept of dynamical analogue prediction (DAP) has been introduced in order to effectively utilize historical analogy information in dynamical prediction and realize the adequate combination of dynamical and statistical methods[4,19]. Hence, substituting $\tilde{\psi}$ into eq. (2) yields

$$\frac{\partial \tilde{\psi}}{\partial t} + L(\tilde{\psi}) = E(\tilde{\psi}). \qquad (3)$$

Note that ψ is quite close to $\tilde{\psi}$, so $E(\psi)$ can be Taylor expanded to the first order in terms of ψ about $\tilde{\psi}$:

$$E(\psi) = E(\tilde{\psi}) + (\psi - \tilde{\psi}) D\big|_{\tilde{\psi}},$$

where D stands for the sum of the partial differentials of E with respect to every component of ψ. As we can see, when $D|_{\tilde{\psi}}$ is bounded and $\|\psi - \tilde{\psi}\|$ is small enough, one may estimate $E(\psi)$ on the right-hand side of eq. (2) by using error term $E(\tilde{\psi})$ on the right side of eq. (3), and obtain the analogue-correction equation of errors (ACEE)

$$\frac{\partial \psi}{\partial t} + L(\psi) = \frac{\partial \tilde{\psi}}{\partial t} + L(\tilde{\psi}). \qquad (4)$$

It can be clearly seen that the first term on the right hand of eq. (4) is known and the second term can be calculated by the numerical model. Thus, eq. (4) may be considered to append an analogue-correction term of errors into eq. (1) in order to be closer to eq. (2) that represents the exact model satisfied by real atmosphere.

For current initial value ψ_0, prediction of eq. (1) is denoted as $P(\psi_0)$ and that of eq. (2) is denoted as $\breve{P}(\psi_0)$ (namely observed data, unknown). Then, under the condition of taking no account of observation errors, subtracting eq. (1) from eq. (2) after they are time-integrated respectively yields

$$\hat{E}(\psi_0) \equiv \int_{t_0}^{t_0+\delta t} E(\psi)\mathrm{d}t = \breve{P}(\psi_0) - P(\psi_0).$$

where δt is integrated time period and $\breve{P}(\psi_0)$ is the observed data corresponding to $P(\psi_0)$. It can be obviously seen that $\hat{E}(\psi_0)$ is the contribution of model error term $E(\psi)$ to prediction result. If one estimates $\hat{E}(\psi_0)$ aforehand, prediction can be expressed as $P(\psi_0) + \hat{E}(\psi_0)$. At this time, the above prediction problem has been transformed into solving $\hat{E}(\psi_0)$, which indicates the error correction for dynamical prediction $P(\psi_0)$.

It follows from the above analyses that for improving the prediction of dynamical model, one may propose a new idea to predict the prediction errors of dynamical model and to transform the dynamical prediction problem into the estimation problem of prediction errors, which can be carried out by statistical methodology. The information of prediction errors is actually involved in a series of special solutions of the exact model that real atmosphere or climate system satisfies, which need to be extracted in virtue of existing numerical models. Thus it will be a key to discussing estimating methods based on the information of these prediction errors.

3 Prediction of prediction errors (POPE) in dynamical model

If historical observed ψ_i is regarded as initial value, the prediction of eq. (1) can be denoted as $P(\psi_i)$ and that of eq. (3) can be denoted as $\breve{P}(\psi_i)$ (namely historical observed data, here $i = 1, 2 \ldots N$, where N may be taken as the number of all of samples in historical data). Therefore, one can obtain N known prediction errors $\hat{E}(\psi_i) = \breve{P}(\psi_i) - P(\psi_i)$. Based on their arithmetic average, systematic errors $\hat{E}(\psi_0)$ can be estimated. However, such the estimation of prediction errors has no pertinence for various initial values with time even though historical information is also used to improve model prediction in systematic error correction.

For different ψ_0, one may pertinently select historical analogue $\tilde{\psi}_j$ suitable for estimating current errors (here $j = 1, 2 \ldots m$, where m is the number of selected analogues). Similar to $\hat{E}(\psi_0)$, here the information of prediction errors can be solved as

$$\hat{E}(\tilde{\psi}_j) \equiv \int_{t_\mathrm{h}}^{t_\mathrm{h}+\delta t} E\left(\tilde{\psi}\right)\mathrm{d}t = \breve{P}(\tilde{\psi}_j) - P(\tilde{\psi}_j).$$

This just is prediction errors presented by analogue $\tilde{\psi}_j$, where t_h is historical time, and $\breve{P}(\tilde{\psi}_j)$ is historical observed data corresponding to $P(\tilde{\psi}_j)$ and both of them are known. Thus, the new idea to predict the prediction errors of dynamical model on the basis of historical analogue information is put forward so as to transform the dynamical prediction problem into the estimation problem of prediction errors: $\hat{E}(\tilde{\psi}_j) \to \hat{E}(\psi_0)$.

Furthermore, time-integrating the differential equation (4) yields

$$\int_{t_0}^{t_0+\delta t} \frac{\partial \psi}{\partial t}\mathrm{d}t + \int_{t_0}^{t_0+\delta t} L\left(\psi\right)\mathrm{d}t = \int_{t_\mathrm{h}}^{t_\mathrm{h}+\delta t} \frac{\partial \tilde{\psi}}{\partial t}\mathrm{d}t + \int_{t_\mathrm{h}}^{t_\mathrm{h}+\delta t} L\left(\tilde{\psi}\right)\mathrm{d}t. \quad (5)$$

And then, from current initial value and historical data, one can respectively obtain

$$\int_{t_0}^{t_0+\delta t} \frac{\partial \psi}{\partial t}\mathrm{d}t = \hat{P}\left(\psi_0\right) - \psi_0, \text{ and}$$

$$\int_{t_\mathrm{h}}^{t_\mathrm{h}+\delta t} \frac{\partial \tilde{\psi}}{\partial t}\mathrm{d}t = \breve{P}(\tilde{\psi}_j) - \tilde{\psi}_j.$$

Here, $\hat{P}(\psi_0)$ represents the prediction results solved under the condition that the error term on the right-hand side of eq. (5) is estimated by using analogy information, which will become $P(\psi_0)$ if the right term of eq. (4) is equal to 0, and which will be $\breve{P}(\psi_0)$ if the right term is $E\left(\psi\right)$. For ψ and $\tilde{\psi}_j$, based on the prediction model represented by eq. (1), one can respectively obtain

$$\int_{t_0}^{t_0+\delta t} L\left(\psi\right)\mathrm{d}t = \psi_0 - P(\psi_0), \text{ and}$$

$$\int_{t_\mathrm{h}}^{t_\mathrm{h}+\delta t} L\left(\tilde{\psi}\right)\mathrm{d}t = \tilde{\psi}_j - P(\tilde{\psi}_j).$$

Substitute them into eq. (5) and derive the ultimate prediction at current time $t_0+\delta t$ as follows:

$$\hat{P}(\psi_0) = P(\psi_0) + \breve{P}(\tilde{\psi}_j) - P(\tilde{\psi}_j). \quad (6)$$

This is called as the DAP equation (DAPE), which is significantly deferent from the statistical analogue prediction equation (SAPE) $\hat{P}(\psi_0)=\breve{P}(\tilde{\psi}_j)$. The difference between them is the DAP increment (DAPI) $P(\psi_0)-P(\tilde{\psi}_j)$, which reflects the implication of "dynamical analogue". Clearly, on the right side of eq. (6), $\hat{E}(\psi_0)$ is estimated by utilizing prediction errors $\hat{E}(\tilde{\psi}_j)$ corresponding to historical analogues[18]. Thus, eq. (6) shows to append an analogue-correction term of errors on the basis of model prediction so as to realize the prediction of current prediction errors.

4 A new method of dynamical analogue prediction

Evidently, the DAPE can be suitable for prediction on every timescale. Especially, for monthly and seasonal mean predictions interested in short-term climate prediction, one can further develop a new method of analogue correction of errors suitable for mean prediction based on eq. (6).

4.1 Analogue correction of errors (ACE)

Consider to write eq. (6) at time $t_0+\delta t$, $t_0+2\delta t$ … $t_0+k\delta t$, where k may vary with various prediction objectives. Time-integrating eq. (6) during the period of $[t_1, t_2]$ after initial time t_0 yields

$$\int_{t_1}^{t_2}\hat{P}(\psi_0)=\int_{t_1}^{t_2}P(\psi_0)+\int_{t_1}^{t_2}\breve{P}(\tilde{\psi}_j)-\int_{t_1}^{t_2}P(\tilde{\psi}_j). \quad (7)$$

This is the DAPE in the time-integration sense. If prediction objective is any monthly or seasonal mean in future, then we conduct monthly or seasonal mean for eq. (7) during $[t_1, t_2]$ and obtain

$$\hat{P}_{\mathrm{MM}}(\psi_0)=P_{\mathrm{MM}}(\psi_0)+\breve{P}_{\mathrm{MM}}(\tilde{\psi}_j)-P_{\mathrm{MM}}(\tilde{\psi}_j), \quad (8)$$

$$\hat{P}_{\mathrm{SM}}(\psi_0)=P_{\mathrm{SM}}(\psi_0)+\breve{P}_{\mathrm{SM}}(\tilde{\psi}_j)-P_{\mathrm{SM}}(\tilde{\psi}_j). \quad (9)$$

They are named monthly and seasonal mean DAPEs, respectively. The corresponding DAPIs, defined as the difference between two predictions based on current initial value and historical analogue, are $P_{\mathrm{MM}}(\psi_0)-P_{\mathrm{MM}}(\tilde{\psi}_j)$ and $P_{\mathrm{SM}}(\psi_0)-P_{\mathrm{SM}}(\tilde{\psi}_j)$ respectively. If let DAPI = 0, eqs. (8) and (9) will get monthly and seasonal mean SAPEs: $\hat{P}_{\mathrm{MM}}(\psi_0)=\breve{P}_{\mathrm{MM}}(\tilde{\psi}_j)$ and $\hat{P}_{\mathrm{SM}}(\psi_0)=\breve{P}_{\mathrm{SM}}(\tilde{\psi}_j)$. Here, mean observed data corresponding to historical analogical initial value are directly regarded as the current mean prediction.

It can be seen that a new method of analogue correction of errors (ACE) suitable for short-term climate prediction has been proposed in eqs. (8) and (9). Different from the universal definition of the ACE[18] in eq. (6), here is the final analogue correction of errors (denoted as FACE). Besides, it should be pointed out that in the derivation process with respect to the FACE, prediction objectives are not expressed as prediction variables ψ but their functions $P(\psi)$, which indicates that the FACE can be directly applied to the correction of prediction variables (e.g., geopotential height) or their functions (e.g., precipitation).

4.2 Estimation of prediction errors

As we know, in the process of estimating current prediction errors by using historical analogy information, there usually exist a good many analogues in history. If one only utilizes an analogue for error estimation in terms of the DAPE, then a great deal of useful information will be lost, which could influence predictive effect. Hence, it will be very important to study the methodology of estimating prediction errors based on multi-analogue information, which will be discussed theoretically and practically in the following.

(i) Hyperplane approximation method (HAM). After obtaining m prediction errors $\hat{E}(\tilde{\psi}_j)$ from historical analogues, how can current prediction errors $\hat{E}(\psi_0)$ be estimated? Set $\psi\in R^n$, the error estimation problem may be mathematically expressed as

$$\hat{E}(\psi_0)=C(\hat{E}(\tilde{\psi}_1),\hat{E}(\tilde{\psi}_2),\cdots,\hat{E}(\tilde{\psi}_m)), \quad (10)$$

where ψ_0 is current initial value, $\tilde{\psi}_j$ is analogical initial value (j = 1, 2 … m, where m is the number of selected analogues), C stands for the algorithms estimating current errors based on errors from historical analogues, the special form of which can be decided in terms of given situation. It can be seen from Figure 1 that one may look on the function $\hat{E}(\psi)$ satisfied by $\hat{E}(\psi_0)$ and $\hat{E}(\psi_j)$ with respect to ψ as a curve plane S. Similar to the curve-plane fitting method in objective analysis of meteorological data, one can determine coefficients in the analyzed expression of the curve plane S by using observed station data and calculate the value of any spot on the curve plane. Indeed, for n-dimensional problem here, it will be very difficult to *a priori* assume an expression of S.

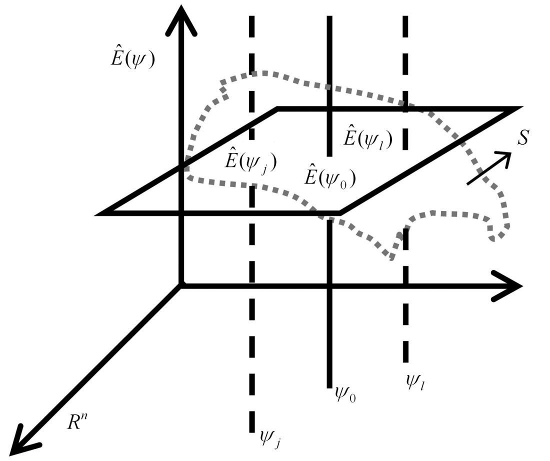

Figure 1 Schematic illustration of functional curve-plane and its tangent plane approximation satisfied by $\hat{E}(\psi)$ in n-dimensional space.

Here, a tangent linear approximation of S can be used where there is a tangent plane passing through $\hat{E}(\psi_0)$ (also see Figure 1). Then, take the component form $\psi=(\varphi_1,\varphi_2,\cdots,\varphi_i,\cdots,\varphi_n)$, and let $\psi_j=\psi_0+\psi'_j$, and suppose that the selected analogue ψ_j is enough close to ψ_0. So on this plane, Taylor-expand $\hat{E}(\psi_j)$ with respect to ψ_0 and only retain the first order approximation to have

$$\hat{E}(\psi_j)=\hat{E}(\psi_0)+\left.\frac{\partial\hat{E}}{\partial\varphi_1}\right|_{\psi_0}\varphi'_1(j)+\left.\frac{\partial\hat{E}}{\partial\varphi_2}\right|_{\psi_0}\varphi'_2(j)+\cdots+\left.\frac{\partial\hat{E}}{\partial\varphi_n}\right|_{\psi_0}\varphi'_n(j), \tag{11}$$

where $\varphi'_i(j)=\varphi_i(j)-\varphi_i(0)$. Let $\hat{E}_j=\hat{E}(\psi_j)=\hat{E}(\psi_0+\psi'_j)$, $\hat{E}_0=\hat{E}(\psi_0)$, $d_i=\partial\hat{E}/\partial\varphi_i|_{\psi_0}$, $\varphi'_{ji}=\varphi'_i(j)$. At this time, eq. (11) can be rewritten as a matrix form:

$$A\vec{u}=\vec{v}, \tag{12}$$

in which

$$A=\begin{pmatrix}1&\varphi'_{11}&\varphi'_{12}&\cdots&\varphi'_{1n}\\1&\varphi'_{21}&\varphi'_{22}&\cdots&\varphi'_{2n}\\\vdots&\vdots&\vdots&\vdots&\vdots\\1&\varphi'_{m1}&\varphi'_{m2}&\cdots&\varphi'_{mn}\end{pmatrix},\ \vec{u}=\begin{pmatrix}\hat{E}_0\\d_1\\\vdots\\d_n\end{pmatrix},\ \vec{v}=\begin{pmatrix}\hat{E}_1\\\hat{E}_2\\\vdots\\\hat{E}_m\end{pmatrix}.$$

For some j, if $\varphi'_{j1}=\varphi'_{j2}=\cdots=\varphi'_{jn}=0$, then easily get $\hat{E}(\psi_0)=\hat{E}(\psi_j)$ from eq. (12). When $m=n+1$ and $|A|\neq 0$, then eq. (12) has a unique solution $\vec{u}=A^{-1}\vec{v}$. In fact, $\hat{E}_0,\hat{E}_j\in R^n$, so only obtain a component of $\hat{E}_0$ ahead. Further, its n components $(\vec{u}_k, k=1,2\cdots n)$ can also be obtained after solving all the components $(\vec{v}_k, k=1,2\cdots n)$ of $\hat{E}_j$ in eq. (12) again and again for n times. Ultimately, $\hat{E}(\psi_0)$ is obtained.

Accordingly, based on eq. (12), error estimation problem has been transformed into retrieval problem of linear algebraic equation group. Thus a so-called hyper-plane approximation method (HAM) has been proposed for estimating current prediction errors based on errors from historical analogues. However, the HAM could be only suitable for low-dimensional theoretical studies. In practical applications, it is necessary to design simplified schemes in terms of given situations.

(ii) Simple linear estimation method (SLEM) in the least-square sense. If $m\ll n$, that is to say, the number of suited historical analogues that can be selected is far smaller than the dimension of prediction variable, then it will be almost impossible to solve the retrieval problem of linear algebraic equation group presented by eq. (12). In practice, it can be seen that the process of solving $\hat{E}(\psi_0)$ from $\hat{E}(\psi_j)$ is quite analogous to the process that station data are interpolated into grid-point data in objective data analysis. The differences between them are that the interpolation is conducted on 2-dimensional plane but error estimation problem is done on n-dimensional plane. Therefore, the interpolation problem of multi-dimensional function on hyper-plane comes into being and C in eq. (10) may be taken as the simple linear estimation method (SLEM) in the least-square sense. Here, consider a simple optimization problem. In order to minimize the total distance between $\hat{E}_0$ and all of $\hat{E}_j$, an objective function can be defined as

$$J(\hat{E})=\frac{1}{2}\sum_{j=1}^{m}b_j(\hat{E}-\hat{E}_j)^2.$$

When J comes to minimum, it is not difficult to obtain the optimal $\hat{E}_0=\hat{E}^*$ according to the necessary condition as follows:

$$\frac{\partial J}{\partial\hat{E}}=\sum_{j=1}^{m}b_j(\hat{E}^*-\hat{E}_j)=0.$$

At this time, eq. (10) becomes

$$\hat{E}(\psi_0) = \sum_{j=1}^{m} b_j \hat{E}(\tilde{\psi}_j) \Big/ \sum_{j=1}^{m} b_j = \sum_{j=1}^{m} a_j \hat{E}(\tilde{\psi}_j), \tag{13}$$

where $a_j = b_j \Big/ \sum_{j=1}^{m} b_j$ stands for the jth normalized weighted coefficient and b_j is undetermined coefficient. The SLEM may be regarded as the considerable simplified version of the HAM. In other words, let $d_i = 0$ in eq. (12) and $\hat{E}(\psi_j)$ are weightedly averaged.

5 Extra-seasonal prediction experiment based on the FACE

In the process of the FACE, historical analogues are first selected based on current initial value, and then error information in history is extracted by employing current model. Further, one can estimate current prediction errors and correct original model prediction. In order to examine the performance of the FACE in practical prediction, monthly and seasonal prediction experiments have been already conducted. Here, only give the extra-seasonal prediction results of summer circulation and precipitation based on the FACE.

5.1 Atmospheric rhythm phenomena and analogue selection

First of all, the selection of historical analogue plays a key role in the DAP, and the pertinent scheme for selecting analogue needs to be introduced for prediction problems on different scales. For the extra-seasonal prediction of summer circulation and precipitation, the general initial values, viz. the corresponding variable fields in early winter, are utilized for analogue selection, the rationality of which is that there exist the atmospheric rhythm phenomena with a timescale of 3—6 months in the long-range weather process. Wang's investigation[21] of the evolution of analogical circulation anomalies has shown that if the circulation anomalies in Januaries of any two years are analogical to each other, they will be also analogical in Junes or Augusts of the same two years. Moreover, the dynamical mechanism associated with such analogy or rhythm phenomena may be discussed by employing an air-ocean coupled model in the analogue-deviation form[22]. Theoretical analyses and numerical simulations have showed that the generation of such analogy or rhythm phenomena is because under the forcing of seasonal variation of monthly mean circulation, the asymmetrical oscillations of analogue-deviation perturbation are induced by the nonlinear interactions of air-ocean coupled system.

The above-mentioned atmospheric rhythm phenomena indicate that when two monthly deviation fields in different years are analogical in some initial month, such analogy will generally get weak with time later and get strong again after about quite a few months. As the rhythm behavior of circulation evolution, it has been documented that there widely exists the analogy rhythm with a timescale of about half a year in the long-range evolution of weather anomalies[21]. Here, the scheme for analogue selection by using variable fields in early winter just reflects such analogy rhythm. Actually, in current flooding-season prediction of China, it is one of the most familiar methods to predict summer anomalies by utilizing anomalous signals in early winter fields.

5.2 Model, data and scheme

In the following experiments, a 23-year hindcast dataset during 1983—2005 produced by the operational NCC/IAP T63 atmosphere-ocean coupled general circulation model (CGCM) of CMA/National Climate Center (NCC). Here, annual June—August ensemble-mean data predicted at the end of February are used. Moreover, the initial values of atmospheric model are generated at 00Z UTC on the last 8 days of February from the NCEP/NCAR reanalysis dataset (NNRA). The initial values of oceanic model are from the global ocean data assimilation system of NCC. These initial values are perturbed and combined into 48 ensemble members of dynamical seasonal prediction. The CGCM have been detailedly introduced in existing literatures[23,24].

Experiment scheme of the FACE for predicting summer circulation and precipitation is designed as: (1) Choose summer mean 500 hPa geopotential height (GH500) and total precipitation from the 23-year hindcast dataset; (2) choose 23-winter seasonal mean GH500, sea level pressure (SLP), 1000—500 hPa thickness (THICK) from the NNRA, and extended reconstructed sea surface temperature (ERSST) from NOAA/NCDC during 1982/83—2004/05 to select historical analogues; (3) choose summer mean GH500 from the NNRA and total precipitation of CMAP provided by the NOAA/OAR/ESRL/PSD for verifying prediction; (4) the FACE is based on eq. (9) and the SLEM on eq. (13) is employed to estimate prediction errors, where for b_j, the anomaly correlation coefficient (ACC) and the root mean square error (RMSE) are taken as distances to measure analogy, and the number of selected analogues is undetermined; (5) cross-validation is used for prediction experiments, in which any year is selected as objective year every time and the known information in residual years is utilized to predict summer circulation and precipitation in objective year; and (6) verification scores consist of time correction coefficient (TCC) and pattern correction coefficient (PCC) or ACC with the spatial sense.

5.3 Prediction results of summer mean circulation

As a comparison, Figure 2 first gives the CGCM predictions of summer GH500 based on systematic prediction error correction (SPEC).

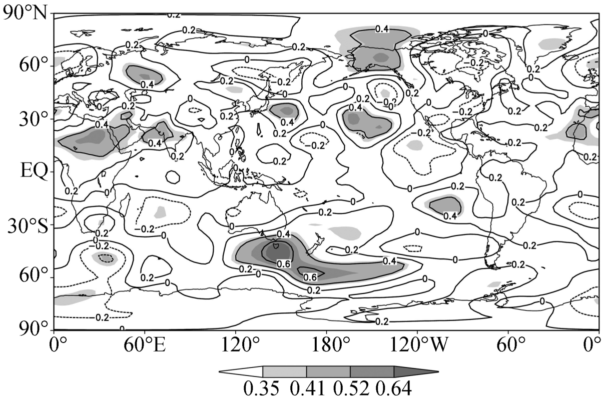

Figure 2 The distribution of TCCs between predictions and verifications of summer mean GH500 based on the SPEC, where numbers 0.35, 0.41, 0.52 and 0.64 stand for 10%, 5%, 1% and 0.1% significance levels based on Student's t-test, respectively.

It can be seen from Figure 2 that TCCs between prediction and verification of summer circulation are relatively small as a whole. The color shade shows areas that exceed a 10% significance level based on Student's t-test. Obviously, TCCs in only a few areas are significant, which indicates that the predictive effects of summer circulation based on the SPEC are unsatisfactory. Furthermore, as the available hindcast dataset only has a very short length of 23 years, the most number of selected analogues based on early winter fields is 22 for every predicted objective year. Here, only the first 4 best analogues are used for the FACE experiments in order to guarantee considerable analogy. Figure 3 presents the FACE predictions of summer mean GH500 by selecting analogues based on early winter GH500.

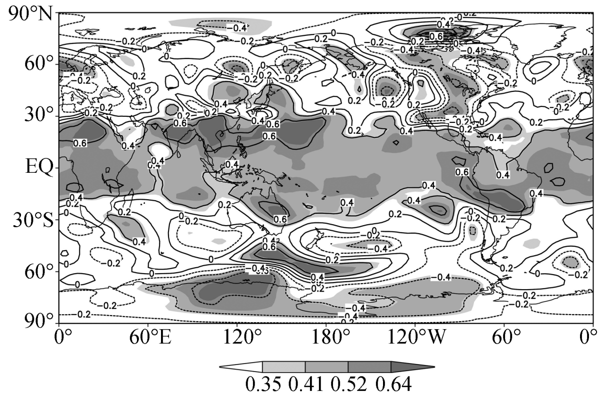

Figure 3 The same as Figure 2, but for the FACE by taking ACC as analogy metric and using the first 4 best analogues.

Compared with Figure 2, the whole low-latitude areas are almost covered by the high positive correlation beyond a 5% significance level in Figure 3. Especially, there exist several extremal centers of high correlation at the 0.1% significance level over the Asian, African and South American monsoon areas, which may provide some valuable references for seasonal prediction in these monsoon areas. Clearly, the unique significant positive correlation over the Northern Hemispheric middle-latitude zones lies in the East Asian monsoon area with high climate variability and low predictability, where a high-correlation center exists over the areas from western Northeast China to northern North China and more significant positive correlation area is located south of the Yangtze River.

It is well known that dynamical prediction is inevitably characterized by damped wave amplitudes and gradually tend to model climate mean with integrated time. Although the SPEC can help to eliminate the climatic drift, the variances of interannual variability of dynamical prediction still need be amplified again. Figure 4 further gives the comparisons between predicted and verified standard deviations of interannual variability of summer mean GH500.

In contrast to the standard deviations of verifications in Figure 4(a), those corresponding to the SPEC are so small that there only are few contours in Figure 4(b). Comparatively speaking, the FACE exhibits better corrected effects for standard deviations, where those main high-variability centers are well replicated, particularly over the southern and northern Pacific as well as middle- and high-latitude areas, even though only about 50% of standard deviations of verifications is recovered in quantity based on the FACE. Actually, the error correction of variances associated with amplitude damping can be implemented by multiplying "inflation factor" on the damped prediction for recovering their amplitudes[25], but which could cause singular values when the standard deviations of predicted interannual variability are close to 0. Surprisingly, the FACE, without special procedure for such inflation, can effectively restore amplitudes for damped prediction, which is evidently linked with error-estimation process in the FACE. This is because the prediction errors presented by historical analogues contain the information of significant differences between predicted and verified standard deviations. By the error-estimation process, such information is passed into new estimated errors and further into prediction results.

5.4 Prediction results of summer total precipitation

In the following, the FACE is applied to prediction experiments of summer total precipitation. Note that rainfall generally happens on a small spatial scale with strong locality. Here, 4 regions are chosen to calculate PCCs between prediction and verification, and Table 1 gives 23-year mean results. We can see that in the 4 regions, the SPEC exhibits positive but small skill scores, whereas PCCs based on the FACE with the first 4 best selected analogues by using early winter GH500 all exceed 0.1, which displays that prediction scores are significantly heightened.

Figure 5 presents the interannual variability of PCCs in the 4 regions. It can be seen that the skill-score curves of the SPEC nearly surround 0 lines and oscillate, whereas those of the FACE mostly lies above 0 lines. Especially, there are 17-year PCCs more than 0 in China region, which reflects the robust performance of the FACE. However, the latest 3-year PCCs are all less than 0, which implies the complexity of summer precipitation prediction. Moreover, for either the SPEC or the FACE, there is a good coherence between global and tropical curves, which indicates that the prevailing contributions to global summer precipitation pattern are from tropics.

5.5 Sensitive experiments

(i) Impact of the number of analogues on the FACE. Based on the hindcast dataset of the CGCM, the impact of the number of selected analogues on the FACE is first

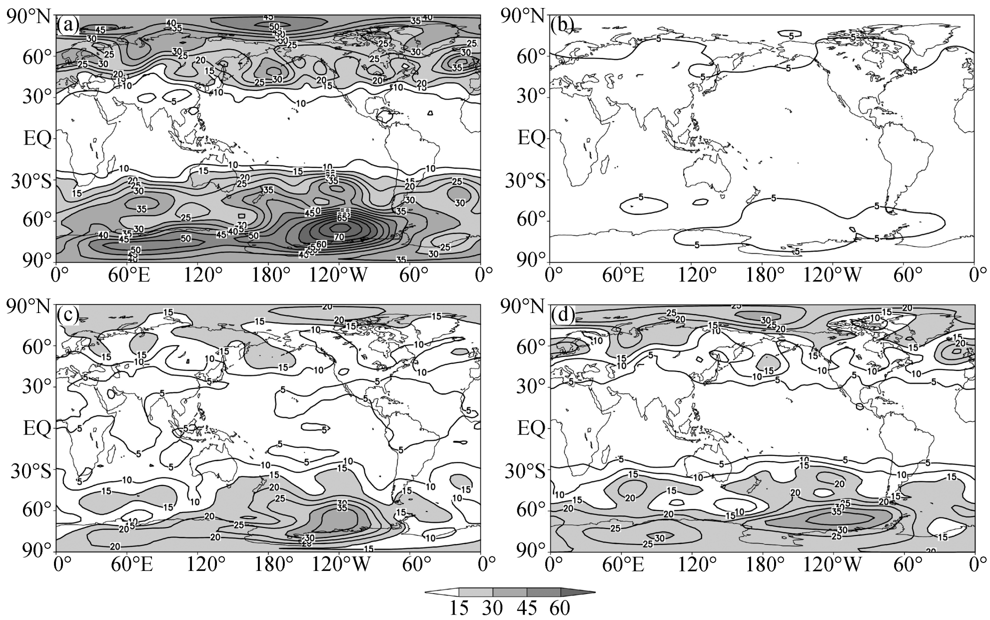

Figure 4 Standard deviations of the predicted and verified interannual variability of summer mean GH500 (Unit: gpm), where (a)−(d) are corresponding to the VERI, SPEC, FACE and VERI-FACE, respectively.

Table 1 The 23-year mean PCCs between predictions and verifications of summer precipitation in different regions

Spatial regions	Globe	Tropics	East Asia	China
	0°−360°E, 60°S−70°N	0°−360°E, 30°S−30°N	100°E−140°E, 10°N−40°N	72°E−136°E, 21°N−54°N
SPEC	0.009	0.010	0.003	0.052
FACE	0.101	0.168	0.127	0.110

ACC and early winter GH500 are used to select the first 4 best analogues in the FACE.

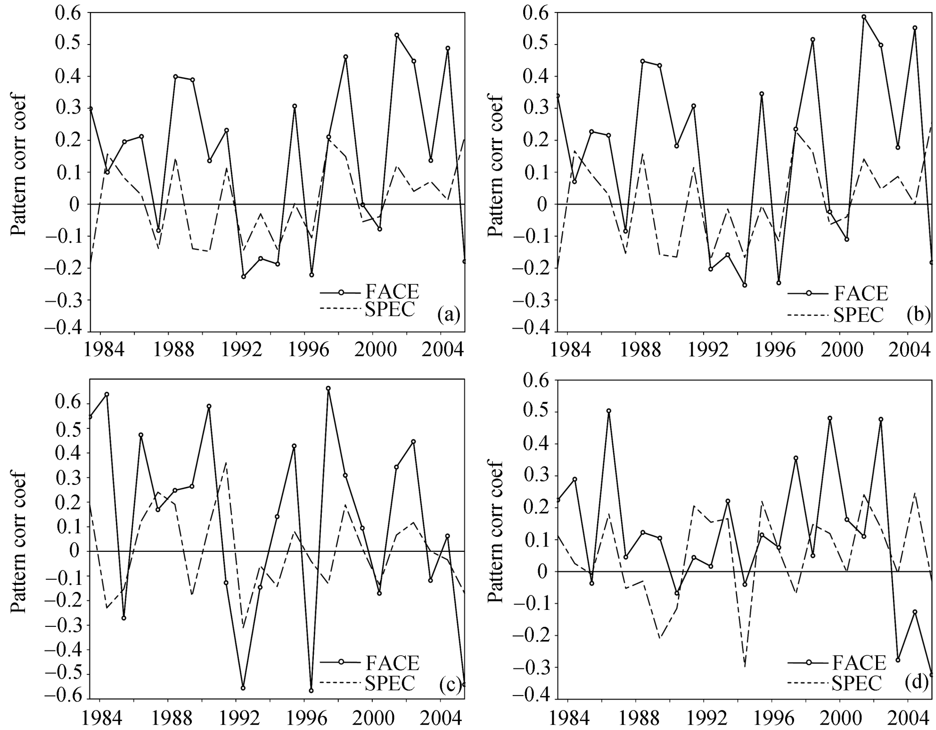

Figure 5 PCCs between predictions and verifications of summer precipitation in the 4 regions based on the FACE by taking ACC as analogy metric and using the first 4 best analogues. (a)—(d) are corresponding to the globe, tropics, East Asia and China, respectively.

examined. It may be seen from Figure 6 that the number of analogues clearly influences the predictive effects of summer circulation and precipitation. Along with increasing the number of analogues, PCCs come to maxima when 3 analogues are introduced, and to minima when 5 analogues. After that, PCCs corresponding to GH500 gradually become large till to reach a stable maximum, whereas those to precipitation come to an appreciably decreased stable value after a few oscillations. It is also shown from Figure 6(b) that the optimal number of analogues could exist. Thus, the predictive effects based on 4 analogues as above are not the best, which still may be improved by modifying the number of analogues used in the FACE.

(ii) Impact of analogue-selected variables on the FACE.

To determine variables for selecting analogues is a quite important problem, which depends on that whether the analogue-selected variables can effectively reflect physical analogy between objective fields. Table 2 lists experiment results of the FACE based on different variables. In the 23-year mean sense, regarding early winter GH500 as analogue-selected variable has more advantages, and the corresponding ACC scores are all the uppermost. Comparatively, prediction skill scores based on early winter ERSST are appreciably decreased and those on thickness field are far better for circulation prediction than precipitation prediction, whereas those corresponding to SLP are somewhat low. Thus, early winter GH500 can be regarded as analogue-selected variable for predicting summer circulation and precipitation, which is close correlated with the atmospheric rhythm phenomena and roots in the mechanism that under the forcing of seasonal variation of monthly mean circulation, the asymmetrical oscillations of analogue-deviation perturbation are induced by the nonlinear interactions of air-ocean coupled system[22]. In fact, the prediction experiments of the FACE just are based on the hindcast data generated by an atmosphere-ocean coupled model.

(iii) Impact of analogy metric on the FACE. Analogy metric may be generally expressed as the distance measured by ACC or RMSE between variable fields. Then, are they suitable for seasonal prediction? As ACC has been used to select analogues in previous experiments, Table 3 further gives prediction scores by taking RMSE as analogy metric. It can be seen that the verified statistics based on the FACE are evidently different by taking between ACC and RMSE as analogy metrics respectively. Here, for precipitation, the skill score corresponding to the former is about 0.1 whereas that corresponding to the latter is negative. For GH500, the skill

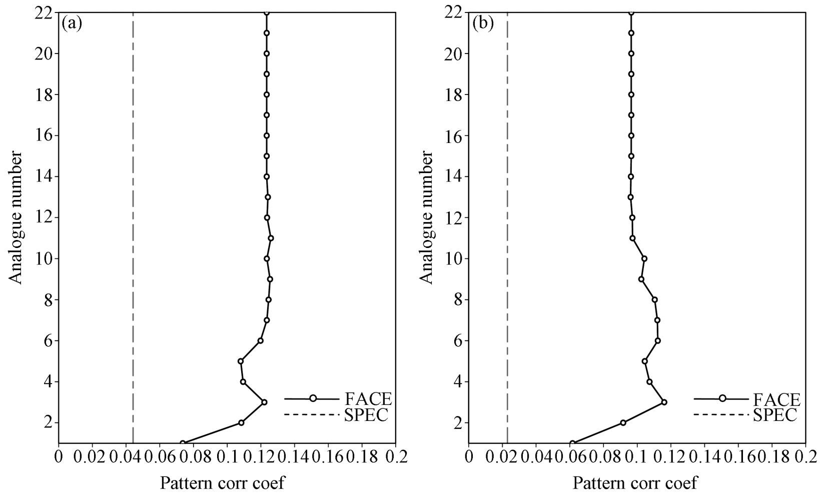

Figure 6 The global PCCs as the function of the number of analogues in summer GH500 (a) and precipitation (b) predictions based on the FACE by taking ACC as analogy metric and regarding early winter GH500 as analogue-selected variable.

Table 2 23-year mean global scores of summer GH500 and precipitation predictions based on the FACE by taking different analogue-selected variables

Prediction objectives	Summer GH500		Summer precipitation	
Prediction scores	RMSE (gpm)	ACC	RMSE (gpm)	ACC
Select analogues by early winter GH500	18.18	0.108	49.91	0.101
Select analogues by early winter ERSST	17.87	0.099	52.03	0.062
Select analogues by early winter THICK	17.87	0.107	51.66	0.050
Select analogues by early winter SLP	17.89	0.067	51.39	0.067

ACC is used to select the first 4 best analogues.

score corresponding to the former is superior to that to the latter but the RMSE scores are close to each other. In a word, the predictions by taking ACC as analogy metric show better effects, which could be related with that the signals from anomalous climatic pattern indicate better physical analogy than those from anomalous amplitude.

Table 3 23-year mean global scores of summer GH500 and precipitation predictions based on the FACE by regarding ACC as analogy metric

Prediction objectives	Summer GH500		Summer precipitation	
Prediction scores	ACC	RMSE (gpm)	ACC	RMSE (gpm)
Select analogues by ACC	0.108	18.18	0.101	49.91
Select analogues by RMSE	0.005	17.92	–0.034	53.30

Early winter GH500 is used to select the first 4 best analogues.

Generally speaking, the predictability of seasonal prediction is primarily originated from the outer forcing of tropical oceans. So it is not difficult to understand that the FACE exhibits more significant improvements for predictions over low-latitude areas than extratropical areas. By using early winter fields for analogue selection, the FACE predictions can reflect atmospheric rhythm phenomena generated from the air-ocean interaction which is the most active over low-latitude areas. Under such physical background, the extra-seasonal predictability implied by early analogy is rooted in low-latitude air-ocean interaction. Moreover, for the middle- and high-latitude prediction based on the FACE, it is the key of improving prediction how to find the analogy indices that can physically stand for predictability information, which will be further studied in the future work.

6 Summary and discussion

Under the conditions of current models and data, it is quite important for improving numerical prediction to deep investigate the prediction strategy and methodology by combining dynamical and statistical methods. In the present paper, based on previous studies of dynamical analogue prediction strategy, a new idea to predict the prediction errors of dynamical model by extracting historical analogue information is put forward in order to transform the dynamical prediction problem into the estimation problem of prediction errors. In terms of such an idea, a new prediction method of final analogue correction of errors (FACE) is further developed. The FACE can effectively use analogy information in historical data and is suitable for predicting each component of climate system models besides atmospheric model. In order to effectively utilize analogical error information for improving model prediction, the new prediction method need not rebuild new model but depend on developments of numerical models and data.

Furthermore, the FACE is applied to extra-seasonal prediction experiments on operational atmosphere-ocean coupled general circulation model of CMA/NCC by utilizing early winter variable fields for analogue selection. Prediction results of summer mean circulation and total precipitation show that the FACE can effectively improve the prediction skills of circulation over low-latitude areas and regional precipitation patterns, which exhibits more evident improvement than simplex systematic error correction. The FACE has the considerable capability of recovering variances of interannual variability of model predictions. Moreover, sensitive experiments also show that many factors including the number of analogues, analogue-selected variables, and analogy metric have significant effects on predictions based on the FACE. In conclusion, the FACE can to some extent reduce prediction errors, recover prediction variances, and improve prediction skills.

At present, there still exist many inherent shortages in short-term climate prediction by employing numerical model. Thus it is a feasible approach for improving prediction levels to develop the prediction strategy of combining dynamical and statistical methods and extracting information from historical analogical data[26]. Indeed, the predictive effects over middle-latitude areas based on the FACE are still unsatisfactory and are under improvement from the selection of analogue. In further works, we will aim at more effective analogue-selected schemes that are suitable for extra-seasonal predictions of circulation and precipitation, and design more representative analogy indices and methods of selecting analogue.

The authors would like to thank Profs. Zhang Peiqun, Li Weijing, Feng Guolin, Huang Jianping, and two anonymous referees for their valuable suggestions.

1 Kalnay E. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge: Cambridge University Press, 2003. 1－341

2 Li Z C, Bi B G, Zhu T, et al. Advances in operational weather forecast in China. Meteor Mon (in Chinese), 2004, 30(12): 4－10

3 Wang S W. Advance in Modern Climatological Studies (in Chinese). Beijing: China Meteorological Press, 2001. 306－311

4 Ren H L, Chou J F. Study Progress in Prediction Strategy and Methodology on Numerical Model. Adv Earth Sci (in Chinese), 2007, 22(4): 376－385

5 Gu Z C. On the equivalency of formulations of weather forecasting as an initial value problem and as an “evolution” problem. Acta Meteorol Sin (in Chinese), 1958, 29(2): 93－98

6 Gu Z C. The use of past data in numerical weather forecast. Acta Meteorol Sin (in Chinese), 1958, 29(3): 176－184

7 Zeng Q C, Zhang B L, Yuan C G, et al. A note on some methods suitable for verifying and correcting the prediction of climate anomaly. Adv Atmos Sci, 1994, 11(2): 121－127

8 Wang H J, Zhou G Q, Zhao Y. An effective method for correcting the seasonal-interannual prediction of summer climate anomaly. Adv Atmos Sci, 2000, 17(2): 234－240

9 Chen H, Lin Z H. A correction method suitable for dynamical seasonal prediction. Adv Atmos Sci, 2006, 23(3): 425－430

10 Chou J F. A problem of using past data in numerical weather forecasting. Sci Sin, 1974, 17(6): 814－825

11 Zheng Q L, Du X Y. A new model of numerical weather prediction by using multi-time observed data. Sci Sin (in Chinese), 1973, 16(2): 289－297

12 Qiu C J, Chou J F. An analogue-dynamical method of weather forecasting. Chin J Atmos Sci (in Chinese), 1989, 13(1): 22－28

13 Huang J P, Wang S W. The experiments of seasonal prediction using the analogy-dynamical model. Sci China Ser B, 1992, 35(2): 207－216

14 Huang J P, Yi Y H, Wang S W, et al. An analogue-dynamical long-range numerical weather prediction system incorporating historical evolution. Quart J Roy Meteor Soc, 1993, 119: 547－565

15 Bao M, Ni Y Q, Chou J F. The experiment of monthly mean circulation prediction using the analogy-dynamical model. Chin Sci Bull, 2004, 49(12), 1296－1300

16 Cao H X. Self-memorization equation in atmospheric motion. Sci China Ser B, 1993, 36(7): 845－855

17 Gu X Q. A spectral model based on atmospheric self-memorization principle. Chin Sci Bull, 1998, 43(20): 1692－1702

18 Ren H L, Chou J F. Analogue correction method of errors by combining both statistical and dynamical methods. Acta Meteorol Sin, 2006, 20(3): 367－373

19 Ren H L, Zhang P Q, Li W J, et al. A new method of dynamical analogue prediction based on multi-reference-state updating and its application. Acta Phys Sin (in Chinese), 2006, 55(8): 4388－4396

20 Livezey R E, Barnston A G, Gruza G V, et al. Comparative skill of two analog seasonal temperature prediction systems: Objective selection of predictors. J Climate, 1994, 7: 608－615

21 Wang S W. The rhythm in the atmosphere and ocean in application to long-range weather forecasting. Adv Atmos Sci, 1984, 1: 7－18

22 Huang J P, Chou J F. Studies on the analogous rhythm phenomenon in coupled ocean atmosphere system. Sci China Ser B, 1990, 33(7): 851－860

23 Ding Y H, Liu Y M, Song Y J, et al. Research and experiments of the dynamical model system for short-term climate prediction. Clim Environ Res (in Chinese), 2002, 7(2): 236－246

24 Li W J, Zhang P Q, Li Q Q, et al. Research and operational application of dynamical climate model prediction system. Quart J Appl Meteor (in Chinese), 2005, 16(Suppl.): 1－11

25 Feddersen H, Navarra A, Ward M N. Reduction of model systematic error by statistical correction for dynamical seasonal predictions. J Climate, 1999, 12: 1974－1989

26 Ren H L. Strategy and methodology of dynamical analogue prediction. Doctor Dissertation. Lanzhou: Lanzhou University, 2006