Analogue-Dynamical Prediction of Monsoon Precipitation in Northeast China Based on Dynamic and Optimal Configuration of Multiple Predictors*

XIONG Kaiguo[1](熊开国), FENG Guolin[2,3†](封国林), HUANG Jianping[1](黄建平), and CHOU Jifan[1](丑纪范)

1 *Key Laboratory for Semi-Arid Climate Change of the Ministry of Education, College of Atmospheric Sciences, Lanzhou University, Lanzhou* 730000

2 *Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing* 100081

3 *Key Laboratory of Regional Climate–Environment Research for Temperate East Asia (RCE–TEA), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029

(Submitted June 1, 2010; in final form March 29, 2011)

ABSTRACT

Based on the National Climate Center (NCC) of China operational seasonal prediction model results for the period 1983–2009 and the US National Weather Service Climate Prediction Center merged analysis of precipitation in the same period, together with the 74 circulation indices of NCC Climate System Diagnostic Division and 40 climate indices of NOAA of US during 1951–2009, an analogue-dynamical technique for objective and quantitative prediction of monsoon precipitation in Northeast China is proposed and implemented. Useful information is extracted from the historical data to estimate the model forecast errors. Dominant predictors and the predictors that exhibit evolving analogues are identified through cross validating the anomaly correlation coefficients (ACC) among single predictors, meanwhile with reference of the results from the dynamic analogue bias correction using four analogue samples. Next, an optimal configuration of multiple predictors is set up and compared with historical optimal multi-predictor configurations and then dynamically adjusted. Finally, the model errors are evaluated and utilized to correct the NCC operational seasonal prediction model results, and the forecast of monsoon precipitation is obtained at last. The independent sample validation shows that this technique has effectively improved the monsoon precipitation prediction skill during 2005–2009. This study demonstrates that the analogue-dynamical approach is feasible in operational prediction of monsoon precipitation.

Key words: analogue-dynamical prediction, monsoon precipitation, correction of errors, dynamic and optimal configuration

Citation: Xiong Kaiguo, Feng Guolin, Huang Jianping, et al., 2011: Analogue-dynamical prediction of monsoon precipitation in Northeast China based on dynamic and optimal configuration of multiple predictors. *Acta Meteor. Sinica.*, **25**(3), 316–326, doi: 10.1007/s13351-011-0307-1.

1. Introduction

In recent decades, due to frequent occurences of various weather disasters, short-term climate prediction becomes an important issue that has attracted increasing attention worldwide. China is located in East Asia with complicated natural conditions, where both the amount of precipitation and the distribution of the rain belt are controlled by the East Asian summer monsoon (Zhu, 1934). Monsoon precipitation anomalies can lead to large-scale droughts and floods, and thus significantly impact on the socio-economic development and people's daily life. Seasonal prediction of precipitation has been the top priority of disaster prevention and reduction in China. Under the background of global warming, increasing economic loss is

*Supported by the Special Public Welfare Research Fund for Meteorological Profession of China Meteorological Administration (GYHY200806005), National Natural Science Foundation of China (40875040 and 40930952), and National Science and Technology Support Program of China (2007BAC29B01 and 2009BAC51B04).

†Corresponding author: fenggl@cma.gov.cn.

(Chinese version to be published)

caused by climate anomalies and extreme weather events, so improving the accuracy of seasonal climate prediction is imperative.

Northeast China lies in the dry and warm continental monsoon climate zone. Large amounts of data reveal that the summer floods and droughts and extreme weather events are fairly active in recent years. There are signs indicating that there is an increasing trend of frequency for droughts and floods. With regard to the causes of monsoon precipitation in Northeast China, plenty of investigations have been done. The relevant studies show that geographical locality of the subtropical high ridge, the Northeast China cold vortex, and the Arctic dipole anomaly has significant correlations with summer rainfall in Northeast China (Sha and Guo, 1998; He et al., 2006; Wu et al., 2008). Liu et al. (2003) analyzed the relationship between seasonal variations of 500-hPa geopotential height in the North Pacific Oscillation (NPO) region (25°–70°N, 140°W–150°E) and precipitation in Northeast China. They found that summer precipitation over Northeast China is above normal when the 500-hPa geopotential height in the NPO region in the previous winter is in the negative phase, and vice versa. Sun and An (2003) found when the equatorial central-eastern Pacific SST has a positive (or negative) departure distribution and the SST over the west wind drift region has a negative (or positive) departure distribution from last summer to the preceding spring, then Northeast China tends to have more (or less) precipitation than normal in summer as a whole. Bai (2001) pointed out when the winter SST in North Atlantic turns warm in the south and cold in the north, it will bring about more precipitation in summer over Northeast China, and the reverse is also true. In addition to the atmospheric circulation and the SST anomalies that affect summer precipitation in Northeast China, many studies show that the snow cover over Eurasia, the Qinghai-Tibetan Plateau and East Asia in pre-winter also affects summer rainfall in Northeast China (Zhai and Zhou, 1997; Chen and Song, 2000).

On the whole, the factors and mechanisms affecting the monsoon precipitation in Northeast China have been proved to be comprehensive and complicated, especially for operational prediction. On the one hand, there are some key factors that influence monsoon precipitation of Northeast China, which have not yet been identified. On the other hand, given all the impact factors under consideration, usually for a particular region, the monsoon precipitation is the result of multi-factor co-action, and in different time periods the configuration or combination of the factors is different. Therefore, it is necessary to use available historical data to explore the formation mechanism of monsoon precipitation so as to improve the monsoon precipitation prediction level in Northeast China.

Statistical and dynamic methods are two basic methods used in short-term climate prediction. Both methods have their own advantages and shortcomings. In general, dynamical seasonal prediction has a limited skill and falls behind the high expectation. The general consensus is that statistical and dynamic methods should be combined and developed together (Chou, 2003). The problem of how to effectively integrate the two methods has been extensively researched (Thomas, 1970; Mo and Straus, 2002; Tippett et al., 2005). Gu (1958) put forward the importance and the feasibility of introducing historical data into numerical prediction. Chou (1986) discussed the principle on how to combine statistical and dynamical methods in the long-term climate prediction. Various prediction methods based on different principles have been proposed, such as utilizing multiple-instant observations in numerical weather forecast models (Zheng and Du, 1973; Chou, 1974), analogue-dynamical prediction (Qiu et al., 1989; Huang and Wang, 1991; Huang et al., 1993), and the methods based on the atmospheric self-memorization (Cao, 1993; Gu, 1998; Feng et al., 2001; Bao et al., 2004). Numerical experiments proved that these methods can improve the prediction skill.

In recent years, with increased observational data and improvement in numerical models, short-term climate prediction has rapidly developed (Ren and Chou, 2005, 2007; Zheng et al., 2009), but the current operational prediction level is not high enough, so further improvement is still needed (Wang, 2001; Barnston et al., 2005; Ren and Chou, 2007).

This study aims to conduct precipitation prediction over Northeast China (40°–55°N, 110°–135°E; total 77 grids) based on the analogue-dynamical method.

Through searching for analogues to the initial field in history, dynamical model errors can be estimated by the corresponding analogous errors. A new technique called changeable/dynamic and optimal configuration of multiple predictors for monsoon precipitation analogue-dynamical prediction is proposed and experimented.

2. Data

The National Climate Centre (NCC) of China has established an operational short-term climate prediction system (Ding et al., 2002; Li et al., 2005). The model horizontal resolution is about $1.875° \times 1.875°$. In this study, hindcast and predicted precipitation data (after being interpolated to a $2.5° \times 2.5°$ grid) for a total of 27 yr (1983–2009) from the NCC operational seasonal prediction model are used. Note that only ensemble mean results for June–August from the integrations beginning at the end of February in each year are selected. The CPC (Climate Prediction Center, National Weather Service, U. S.) Merged Analysis of Precipitation (CMAP) data are taken as observation, and then the model errors, namely, the discrepancies between the model and the CMAP data, are obtained. Analogues are selected from the 74 circulation indices from the NCC Climate System Diagnostic Division and the 40 climate indices from NOAA of US.

3. Basic principle of the analogue-dynamical prediction

In essence, the numerical prediction model is an initial value problem of a set of partial differential equations, which are expressed as follows:

$$\begin{cases} \dfrac{\partial \psi}{\partial t} + L(\psi) = 0, \\ \psi(x, t_0) = \psi_0(x), \end{cases} \tag{1}$$

where $\psi(x, t)$ is the model state vector to be predicted; x and t, respectively, are the vector in spatial coordinate and time; L is the differential operator of ψ, which corresponds to a real numerical model; t_0 denotes the initial time; and ψ_0 is the initial value. The value ψ or its function $P(\psi)$ at $t > t_0$ can be obtained by numerical integration. Then the exact model that the real atmosphere satisfies can be written as

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

where E is the error operator that expresses the errors of the real numerical model. From the dynamical point of view, the historical data can be regard as a series of special solutions that satisfy Eq. (1).

Numerical model is an approximation of the behavior of the actual atmosphere. Scientists have devoted to the enhancement of climate models; however, existing models are still far from being sufficiently precise in prediction. Such efforts are generally carried out by directly improving the model's physical parameterization schemes, numerical methods, and so on, thus to reduce the model errors. Nonetheless, no matter how the model develops, the errors will remain inevitable and even substantial. In fact, although we do not know the exact solutions that control the atmospheric motion, we possess a large number of observation data that are actually a series of special solutions. Thus, we can utilize the information of historical analogues to estimate model errors, thereby to compensate the model deficiencies and reduce the model errors.

The atmosphere is a forced dissipative nonlinear system. The system adjusts to external forcing. Long-term operational prediction experiences have shown that, with alike initial and boundary conditions, the evolutions of the atmosphere are also alike. Therefore, for any initial value, we can utilize the model error information of historical analogue, assuming that the prediction field is a small perturbation superposed on the historical analogue, so the current model prediction errors can be estimated by the corresponding errors of historical analogues. In other words, the problem is transformed from improving the dynamical prediction of numerical models into the inverse problem of utilizing historical data to estimate current unknown model errors based on known historical analogical information.

4. Method

The mechanisms for monsoon precipitation are different in different regions, which means that each region has its own precipitation impact factors. Even under the same external forcing, due to lack of knowledge of the internal physical processes, defects in physical parameterization schemes, and so on, the response pattern and model prediction errors may also vary for different regions. But for a given region, under the analogous external forcing, the model errors will be analogous on certain timescales. Thus, we can utilize the analogue error correction method to correct the model results. The current study is based on the principle of analogue-dynamical prediction, aiming at improving the monsoon precipitation prediction accuracy of the NCC operational seasonal prediction model. Figure 1 shows the flowchart of the analogue-dynamical prediction of monsoon precipitation based on dynamic and optimal multi-predictor configuration.

First, according to the 1983–2009 CMAP data and precipitation data from the NCC operational seasonal prediction model, the model errors in the recent 27 years are obtained. Subsequently, 114 climate indices are taken as 1368 predictors, and 4 cases of historical analogues are selected based on the analogue-dynamical principle. For each predictor, cross-validation is carried out. After sorting these predictors according to their cross-validation ACCs and the evolving analogous scheme, the predictors that exhibit evolving analogues are identified. With these predictors, multi-factor configuration and cross-validation are carried out again, and then the optimal multi-factor configuration is obtained. Combined with the recent historical optimal multi-factor configuration, the steady optimal multi-predictor configuration in a prediction period is eventually determined. At last, based on this final optimal multi-predictor configuration and the analogue-dynamical prediction

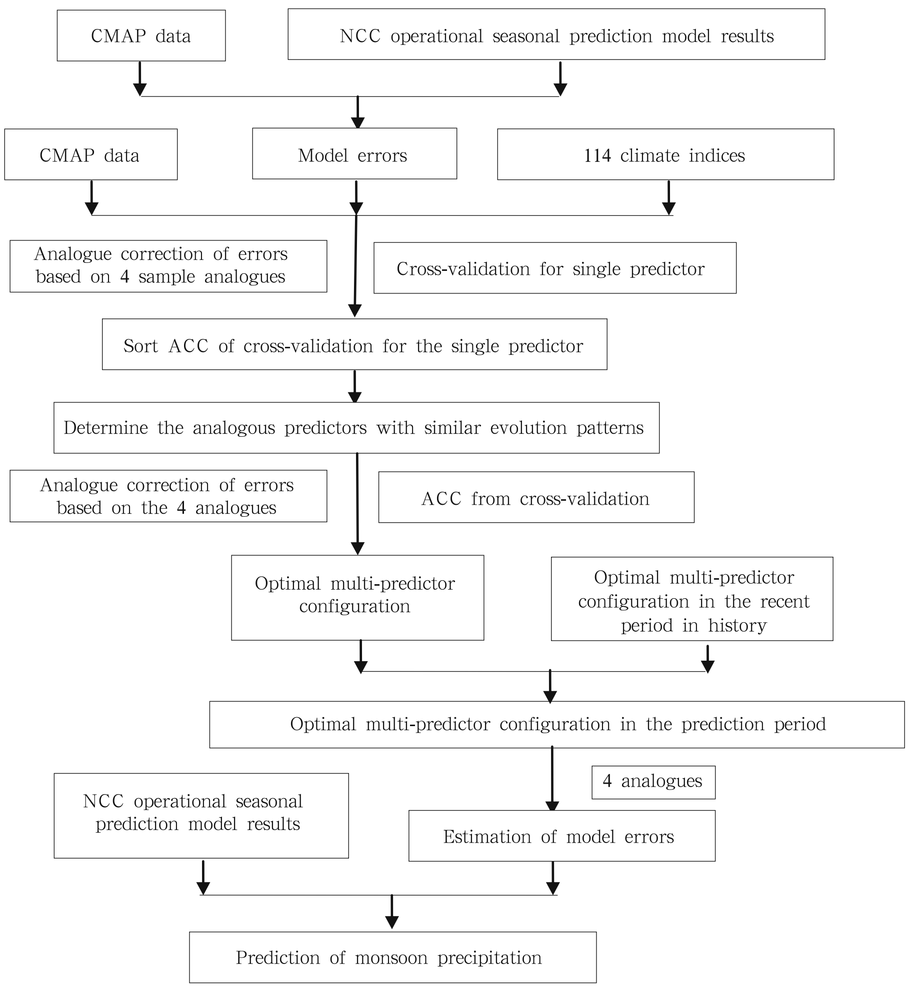

Fig. 1. Flowchart of the analogue-dynamical prediction of monsoon precipitation based on dynamic and optimal configuration of multiple predictors.

method, the model errors are estimated, by which the model prediction results are corrected and the model prediction results are obtained.

4.1 Analogue selection

The analogue-dynamical prediction results are directly influenced by the selection of analogues. Appropriate selection of historical analogues and their numbers can effectively improve the prediction skill. In general, within a certain time frame, the more analogous the initial fields to history, the more closer the model climate evolution and model errors to history and past model errors. Thus, selection of the right historical analogues are very important. The Euclidean distance is taken as the historical analogue selection criterion, and historical model errors for the four most analogous cases are averaged as the prediction model errors (Bao et al., 2004; Ren and Chou, 2007). The related algorithm is shown as follows:

$$\mathrm{AI} = \left[\frac{\sum_{i=1}^{k} w_i(\varphi'_{ij} - \varphi'_{ik})^2}{\sum_{i=1}^{k} w_i} \right]^{\frac{1}{2}}, \tag{3}$$

where AI denotes the Euclidean distance, φ'_{ij} and φ'_{ik} represent two fields at different times (j and k), and w_i is the weight coefficient. The smaller AI is, the more analogous the initial field to the historical field. In view of the basic theory of analogue-dynamical prediction, four historical analogues are selected, and then the prediction model errors can be expressed as:

$$\hat{E}(\psi_0) = \sum_{j=1}^{4} b_j \hat{E}(\widetilde{\psi}_j) \Big/ \sum_{j=1}^{4} b_j = \sum_{j=1}^{4} a_j \hat{E}(\widetilde{\psi}_j), \tag{4}$$

where $\hat{E}(\widetilde{\psi}_j)$ is the model error with respect to historical analogue $\widetilde{\psi}_j, a_j = b_j / \sum_{j=1}^{4} b_j$ is the normalized weight coefficient, and b_j is the coefficient to be determined.

4.2 Selection of regional predictors

A number of previous studies have been carried out to find out predictors that impact on the monsoon precipitation in Northeast China (Zhai and Zhou, 1997; Sha and Guo, 1998; Chen and Song, 2000; Bai, 2001; Liu et al., 2003; Sun and An, 2003; He et al., 2006; Wu et al., 2008). In the current study, 74 circulation indices from NCC and 40 climate indices from NOAA are used as predictors. The distribution of circulation indices varies with months. Therefore, the response of the short-term climate prediction model is different under different circulation conditions which are supplied as the boundary or external forcing to the model. A climate index can be considered as 12 predictors following 12 months, so 114 climate indices can be viewed as 1368 predictors (impact factors).

The selection of regional monsoon precipitation predictors should follow these basic principles: 1) these factors should have clear physical meanings; 2) they should be correlated with the model errors. Predictors are often different in different periods of time, so the prediction skills of predictors in different periods of time vary. Taking into account the above considerations, the principle and process for selection of optimal multiple predictors are given as follows (taking the prediction in 2009 as an example):

1) Selecting the factors that affect regional monsoon precipitation by sorting ACC values. By processing cross-validation on single factor in the three periods: 1983–2006, 1983–2007, and 1983–2008, the factors are sorted according to cross-validation ACCs in each period. The reason why the prediction of 2009 only considers the previous three periods is that the summer rainfall in Northeast China has about 2–4-yr periodic variations (Sun et al., 2000), and a more stable forecasting skill can be achieved for the analogue-dynamical prediction based on the optimal multi-predictor configuration from the previous three periods.

2) Determining the dominant factor. Select at least one factor that can obviously improve the ACC from sorted factors in the above three periods and nominate it the dominant factor. The single factor cross-validation shows that, for a given area, the dominant climatic factor varies from time to time and its contribution to precipitation is stable, thus this factor is chosen as the dominant factor (for example, the EPO in April).

3) Determining the factors with the evolving analogues. For different periods, factors that always play an important role in the rainfall prediction should be

a priority. From the single factor cross-validation ACC or the correlation coefficient between the factors and observed monsoon precipitation, it can be seen that there are many factors impacting monsoon precipitation in Northeast China. The questions are: what factor is dominant and which factor only shows a false correlation? To solve these problems, the concept of "evolving analogues" is brought forth. The results show that adopting the evolving analogues to select regional predictors is one of the reliable ways in improving the precipitation prediction skill. It removes those factors that have nothing to do with the precipitation in this region. Meanwhile, the process of determining the factor with the evolving analogues is essential for selecting only one factor from a climate index. This greatly reduces the degrees of freedom of prediction factors. The process of selecting the factor with the evolving analogues is as follows: ① For a given period of time, select all the factors that have larger ACCs for predictions using the analogue error correction than those using the systematic bias correction; ② find the factors that originate from the same climate index. For a climate index, if many of the factors from it (in different months) have the ability to improve the monsoon prediction skill, then we can say this climate index has the evolving analogues. For the same climate index, only the factor corresponding to the greatest ACC is considered. For simplicity, we take the factor corresponding to the highest ACC as the evolving analogous factor.

4) Determining the optimal multi-predictor configuration. From the factors with the evolving analogues in each period, the configuration of optimal factors is identified through the cross-validation of ACC. Taking the common optimal factors in each period and establishing the regional impact factor set, then the optimal multi-predictor configuration is finalized.

For selection of multi-factor configuration, as a result of large amount of data and large degree of freedom, the calculation work load is huge. Additionally, the noise in the multiple factors and the correlation between factors will lead to nonlinear growth of errors. Mo and Straus (2002) show that, by utilizing EOF decomposition, the degree of freedom can be reduced greatly while the influence of noise can be eliminated effectively; essentially the dimension of the factors is decreased.

EOF decomposition of the multi-factor field $\psi_{m\times n}$ can be expressed as:

$$\psi_{m\times n} = V_{m\times n} T_{n\times n}, \tag{5}$$

where m and n are the space and time dimensions respectively. $V_{m\times n}$ is the normalized orthogonal base field, and $T_{m\times n}$ the time coefficients field for principal components (PCs). For simplicity, only the first h ($h < m$) PCs that jointly explain about 80% of the total variance are used:

$$\psi_{m\times n} = V_{m\times n} T_{n\times n} \approx V_{m\times h} T_{h\times n}. \tag{6}$$

4.3 Sensitivity of the number of predictors

As we know, in short-term climate prediction, for a linear system, the more historical data we use, the better the prediction is. However, for a nonlinear system, the prediction skill will not depend on data quantity but on particularity of data. The atmosphere is a nonlinear complex system and it will be an effective way to utilize the dynamic optimal multi-predictor configuration to compress the degree of freedom by extracting the component predictability so as to improve short-term prediction skill.

Figure 2 shows the sensitivity of ACC to the number of predictors in three periods: 1983–2006, 1983–2007, and 1983–2008. The multi-factor cross-validation is done in each period. It is seen that, for the analogue-dynamical monsoon precipitation prediction, it is not true that the more predictors, the better the prediction. On the whole, with the increase of number of predictors, the ACC increases at first and

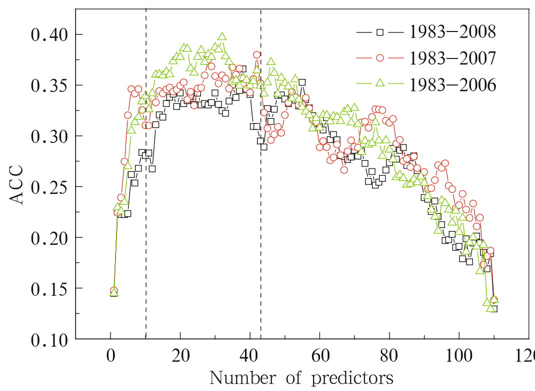

Fig. 2. ACC versus number of predictors for three time periods.

then decreases. When the number of predictors increases from 1 to about 10, the ACC in each period of time increases rapidly to a high value. When it expands to 40 or so, the ACCs for the three periods achieve relatively stable values and maintain a high prediction skill. After more impact factors are brought in, the ACC decreases rapidly. Moreover, it can be seen from Fig. 2 that the cross-validation of ACC for any single factor is below 0.2, while for the optimal multi-factor configuration it is generally above 0.35, some even reaching 0.4, implying the optimal multi-actor configuration is skillful than any single factor. Therefore, the optimal multi-factor configuration is an effective way in short-term climate prediction.

4.4 Physical properties of predictors

With the analogue-dynamical prediction method and optimal multi-predictor configuration, we made a hindcast of monsoon precipitation in Northeast China in 2009. The validation results are given in Table 1. Note that only the factors in January are taken as the factors in the same year while the factors from February to November are taken as the factors of last year. The ACCs in Table 1 are cross-validation results for 1983–2008. The correlation coefficients are between the predictors and the PC1 of observed monsoon precipitation/the NCC seasonal prediction model errors in Northeast China.

For the analogue-dynamical prediction of monsoon precipitation in Northeast China, the dominant predictor is the East Pacific Oscillation (EPO) index, also known as the East Pacific-North Pacific pattern (EP-NP) (Bell and Janowiak, 1995), in last April. The cross-validation shows that for this single predictor, the average ACCs in periods 1983–2006, 1983–2007, and 1983–2008 are 0.14, 0.15, and 0.14, respectively, ranking the second highest, and the highest correlations, respectively. This well indicates the importance of this predictor in monsoon precipitation prediction in Northeast China.

Figure 3 shows the relationship between the anomaly percentages of EPO in last April and the PC1 of monsoon precipitation in Northeast China. Figure 4 gives distributions of the correlation coefficients between the EPO in last April and observed monsoon precipitation/model errors. Especially, from Fig. 4, it

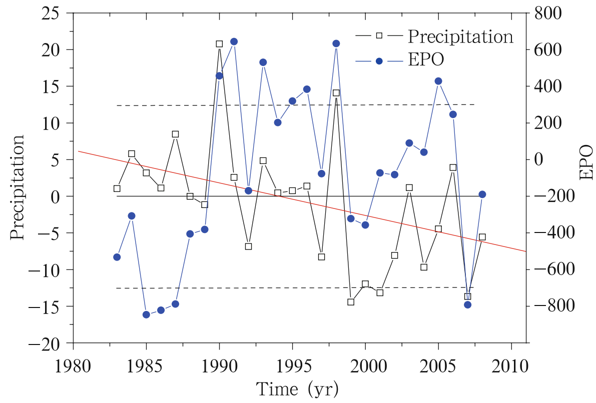

Fig. 3. The relationship between the EPO in April and the PC1 of monsoon precipitation in Northeast China. All values are anomaly percentage, and the climate value is the average from 1983 to 2002.

can be seen that monsoon precipitation and model errors in Northeast China have a strong correlation with the EPO in last April, especially in south of Northeast China, with values passing the 0.01 confidence test over parts of the region. Therefore, the EPO in last April is indeed a factor that impacts on the monsoon precipitation in Northeast China. Since the NCC operational seasonal prediction model has an accurate description of climate state, the distribution of the correlation coefficient between the predictor and model prediction errors is accordant with that between the predictor and observed monsoon precipitation. This is also true for other predictors (figure omitted).

Due to the complexity and nonlinear characteristics of the atmosphere, dominant predictor and evolving analogues are selected by cross-validation of ACCs instead of from the correlation coefficient between the factor and the observed precipitation or model prediction errors. In general, the regional monsoon precipitation is the result of joint effects of multiple factors, and the fact that the factors have a high correlation coefficient with observational precipitation or model prediction errors does not mean a good prediction skill. This may be the reason why the impact factors we selected are different from those of previous researchers (Zhai and Zhou, 1997; Chen and Song, 2000; He et al., 2006; Wu et al., 2008). It can be seen from Table 1, the correlation coefficient between the EPO in last April and PC1 of observed precipitation or model prediction errors in Northeast China are 0.33 and

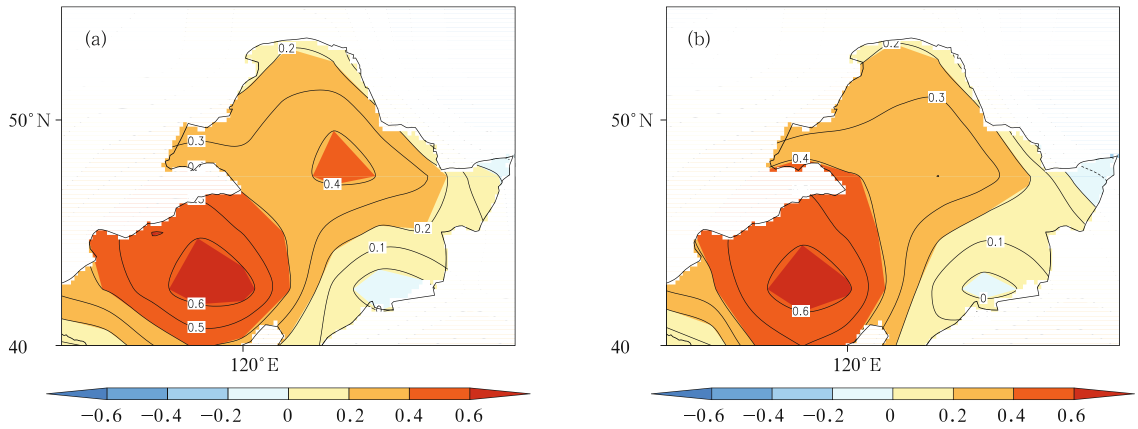

Fig. 4. Distributions of the correlation coefficient between the EPO in April and (a) observed monsoon precipitation, and (b) model errors in Northeast China.

Table 1. The optimal multi-predictor configuration for prediction of monsoon precipitation in Northeast China in 2009

Month	Predictors	Cross-validation ACC	Correlation coefficient with PC1 of observed precipitation/model errors
1	Intensity of the East Asian trough	0.08	–0.20/–0.08
2	Northern position of the subtropical high over the South China Sea	0.02	–0.18/ 0.15
4	EPO	0.14	0.33/–0.11
4	North Tropical Atlantic SST	0.05	–0.38/ 0.21
6	Pacific warm pool	0.01	–0.35/ 0.35
7	India-Burma trough	–0.02	–0.11/–0.03
10	Ridge position of the subtropical high over India	0.001	–0.42/ 0.41
*10	Area-averaged precipitation for Arizona and New Mexico	–0.005	–0.16/ 0.27

*http://www.esrl.noaa.gov/psd/data/climateindices/list/

–0.11, respectively. Although both of them could not pass the 0.05 confidence test, its cross-validation ACC is the largest; whereas the correlation coefficient between the ridge of the subtropical high over India in last October and PC1 of observed precipitation or model prediction errors in Northeast China passes the 0.05 confidence test, but the ACC is only 0.001. Thus, by use of the single-factor cross-validation ACC, selection of regional impact factors is more direct and effective.

In addition, the relationship between the PC1 of eight predictors in Table 1 and the monsoon precipitation in Northeast China is analyzed here (see Fig. 5). With EOF decomposition of the eight predictors, the first 5 PCs' cumulative explained variance already reaches 80%, with each accounting for 37%, 20%, 12%, 11%, and 9% of the total variance, respectively. The relationship between the PC1 of eight predictors and the monsoon precipitation in Northeast China is obviously different from that of the EPO in last April (see Fig. 3). The correlation coefficient for the former is

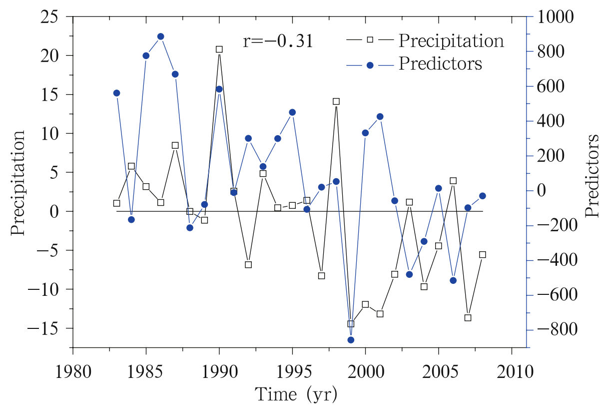

Fig. 5. The relationship between monsoon precipitation in Northeast China and the configuration of eight indices. All values are the anomaly percentage of PC1 after EOF, and the correlation coefficient $r = –0.31$.

–0.31, but for EPO is 0.31. The average cross-validation ACC from 1983 to 2008 for the 8-predictor configuration is 0.18, better than the systematic bias correction prediction and the analogue error correction prediction using the only predictor of EPO in last April, for which ACCs are –0.08 and 0.14, respectively. The improvement also shows in the correlation coefficients: for the systematic bias correction the value is –0.17, while for the analogue errors correction it reaches 0.14, showing a higher prediction skill.

5. Results of the independent sample validation

There are only 27-yr data that can be used in analogue-dynamical prediction experiments, so it is difficult to select a much better historical analogue that excludes the abnormal predictors in certain years. Taking the impact of the length of historical data on the selection of historical analogues into consideration, in this paper, independent sample validation experiments were only carried out in the recent 5 years from 2005 to 2009. Figure 6 shows both ACC and RMSE from the independent sample validation. Based on the principles of analogue selection, according to values of the Euclidean distance, cases for 5 yr are selected and ordered as first, second, third, and fourth analogues in Table 2 (the data in the brackets are the ACC of monsoon precipitation between analogue year and the year of hindcast).

It can be seen from the independent sample validation of ACCs in Fig. 6a that, compared with the systematic bias correction, the ACCs from the analogue error correction under the optimal multi-predictor configuration are mostly increased effectively, except for 2005 and 2006 in the hindcast of 2005–2009 in Northeast China. The ACCs in the rest 3 years are higher, changing from –0.24, –0.08, and 0.09 to 0.13, –0.06, and 0.32, respectively. In 2005, though the ACC from the analogue error correction is smaller, but it still reaches 0.18. As to RMSE, it can be seen from Fig. 6b that, compared with the systematic bias correction, the analogue error correction does not make the RMSE reduced significantly in these 5 years, except 2007. The RMSEs of the rest 4 years increase slightly; the reasons herein need to be further investigated. The ACCs between analogue years and prediction years are shown in Table 2. The ACC values between the first analogue year and prediction years are positive except for 2007. This clarifies that the analogue-dynamical prediction of monsoon precipitation based on the optimal multi-predictor configuration is reasonable and potentially useable.

Table 2. Analogue years of the independent sample validation

Prediction year	First analogue	Second analogue	Third analogue	Fourth analogue
2005	2003(0.06)	1991(–0.08)	2001(0.46)	1997(–0.33)
2006	2005(0.07)	2002(–0.02)	2001(–0.08)	1990(–0.34)
2007	1988(–0.06)	2003(0.20)	1984(0.50)	1989(0.29)
2008	1988(0.26)	1992(–0.07)	1989(0.05)	1984(0.05)
2009	2000(0.36)	2002(0.42)	1992(0.22)	2004(–0.12)

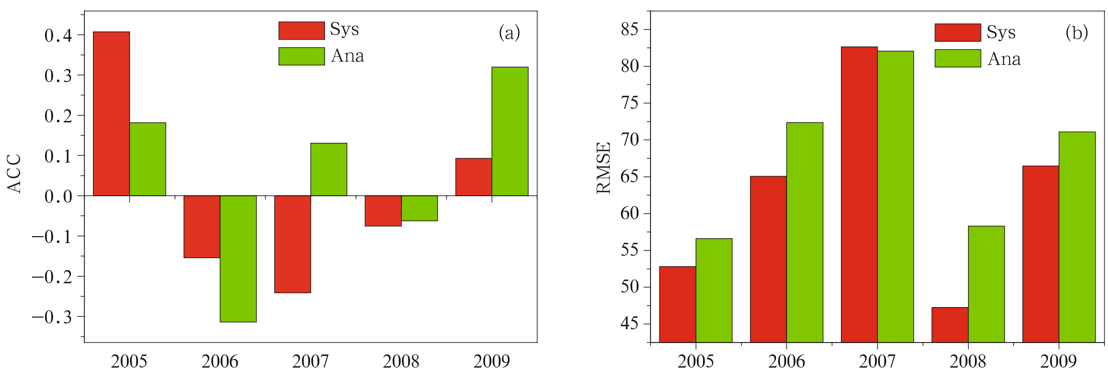

Fig. 6. ACC (a) and RMSE (b) of the independent sample validation. Red and green bars represent results from the systematic bias correction and the anologue error correction, respectively.

6. Conclusions and discussion

Aiming at improving the skill of the NCC seasonal prediction model that uses available historical data, and through searching for analogues to the initial field in the historical data, dynamical model prediction errors can be estimated according to the model prediction errors corresponding to historical analogues. By using the CMAP data from 1983 to 2009 and precipitation data in the same period from the NCC operational seasonal prediction model, we have identified predictors and found historical analogues based on 40 climate indices from NOAA and 74 circulation indices from NCC Climate System Diagnostic Division. Through sorting the predictors according to the cross-validation ACC, we found out the dominant factors and implemented the evolving analogue scheme, then the optimal multi-factor configuration is used, and a new technique called the analogue-dynamical prediction based on changeable/dynamic and optimal configuration of multiple predictors is proposed. Experimental monsoon prodictions were made over Northeast China. Sensitivity tests based on the cross-validation ACC were carried out, which showed that the multi-predictor configuration in monsoon precipitation prediction is feasible and practical. The results of the independent sample validation from 2005 to 2009 show that this method can effectively improve the monsoon precipitation prediction ability of the NCC operational seasonal prediction model in Northeast China. But even in these five years, the ACC average is only 0.05, at about the same level with that of the present operational short-term climate prediction. That is to say, it is not enough to improve the NCC operational seasonal prediction model by only using this method. Like other statistical prediction methods, the ACC in these five years is unstable, and it is shown that there is still a long way to go for improving the short-climate prediction skill. In addition, the impact mechanism of the identified factors on the precipitation needs to be investigated as well.

REFERENCES

Bai Renhai, 2001: Relations between the anomaly of sea surface temperature in the Atlantic and the precipitation in summer over Northeast China. *Marine Sci. Bull.*, **20**(1), 23–29. (in Chinese)

Bao Ming, Ni Yongqi, and Chou Jifan, 2004: The experiment of monthly mean circulation prediction using the analogy-dynamical model. *Chinese Sci. Bull.*, **49**(11), 1112–1115. (in Chinese)

Barnston, A. G., A. Kumar, L. Goddard, et al., 2005: Improving seasonal prediction practices through attribution of climate variability. *Bull. Amer. Meteor. Soc.*, **86**, 59–72.

Bell, G. D., and J. E. Janowiak, 1995: Atmospheric circulation associated with the midwest floods of 1993. *Bull. Amer. Meteor. Soc.*, **76**(5), 681–695.

Cao Hongxing, 1993: Self-memorization equation in atmospheric motion. *Sci. China (Ser. B)*, **23**(1), 104–112. (in Chinese)

Chen Xingfang and Song Wenling, 2000: Analysis of relationship between snow cover on Eurasia and Qinghai Xizang Plateau in winter and summer rainfall in China and application to prediction. *Plateau Meteorology*, **19**(2), 214–223. (in Chinese)

Chou Ge-Fen, 1974: A problem of using past data in numerical weather forecasting. *Scientia Sinica*, **17**(6), 814–825.

Chou Jifan, 1986: Why to combine dynamical and statistical methods together and how to combine. *Plateau Meteorology*, **5**(4), 367–372. (in Chinese)

—–, 2003: Short term climatic prediction: Present condition, problems and way out. *Bimonthly of Xinjiang Meteorology*, **26**(1), 1–4. (in Chinese)

Ding Yihui, Liu Yiming, Song Yongjia, et al., 2002: Research and experiments of the dynamical model system for short-term climate prediction. *Climatic and Environmental Research*, **7**(2), 236–246. (in Chinese)

Feng Guolin, Cao Hongxing, Gao Xinquan, et al., 2001: Prediction of precipitation during summer monsoon with self-memorial model. *Adv. Atmos. Sci.*, **18**(5), 701–709.

—–, —–, Wei Fengying, et al., 2001: On area rainfall ensemble prediction and its application. *Acta Meteor. Sinica*, **59**(2), 206–212. (in Chinese)

Gu Xiangqian, 1998: A spectrum model based on self-memorization principle. *Chinese Sci. Bull.*, **43**(1), 1–9. (in Chinese)

Gu Zhenchao, 1958: On the utilization of past data in numerical weather forecasting. *Acta Meteor. Sinica*, **29**(3), 176–184. (in Chinese)

He Jinhai, Wu Zhiwei, Qi Li, et al., 2006: Relationships among the Northern Hemisphere annual mode, the Northeast cold vortex and the summer rainfall in northeast China. *Journal of Meteorology and Environment*, **22**(1), 1–5. (in Chinese)

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

Huang Jianping and Wang Shaowu, 1991: The experiment of seasonal prediction using the analogy-dynamical model. *Sci. China (Ser. B)*, **2**, 216–224. (in Chinese)

Li Weijing, Zhang Peiqun, Li Qingquan, et al., 2005: Research and operational application of dynamical climate model prediction system. *J. Appl. Meteor. Sci.*, **16**(Suppl.), 1–11. (in Chinese)

Liu Zongxiu, Lian Yi, Shen Baizhu, et al., 2003: Seasonal variation features of 500-hPa height in north pacific oscillation region and its effect on precipitation in Northeast China. *J. Appl. Meteor. Sci.*, **14**(5), 553–56. (in Chinese)

Mo Ruping and D. M. Straus, 2002: Statistical-dynamical seasonal prediction based on principal component regression of GCM ensemble integrations. *Mon. Wea. Rev.*, **130**, 2167–2187.

Qiu Chongjian and Chou Jifan, 1989: An analogue dynamical method of weather forecasting. *Chinese J. Atmos. Sci.*, **13**(1), 22–28. (in Chinese)

Ren Hongli and Chou Jifan, 2005: Analogue correction method of errors by combining both statistical and dynamical methods together. *Acta Meteor. Sinica*, **63**(6), 988–993. (in Chinese)

—— and ——, 2007: Strategy and methodology of dynamical analogue prediction. *Sci. China (Ser. D)*, **37**(8), 1101–1109. (in Chinese)

—— and ——, 2007: Study progress in prediction strategy and methodology on numerical model. *Advances in Earth Science*, **22**(4), 376–385. (in Chinese)

Sha Wanying and Guo Qiyun, 1998: Variations of summer rainfall over China in relation to the geographical locality of subtropical high ridge over West Pacific. *J. Appl. Meteor. Sci.*, **9** (Suppl.), 31–38. (in Chinese)

Sun Li, An Gang, Ding Li, et al., 2000: A climatic analysis of summer precipitation features and anomaly in Northeast China. *Acta Meteor. Sinica*, **58**(1), 70–82. (in Chinese)

—— and ——, 2003: The effect of North Pacific sea surface temperature anomaly on the summer precipitation in Northeast China. *Acta Meteor. Sinica*, **61**(3), 346–353. (in Chinese)

Thomas, A. G., 1970: Statistical-dynamical prediction. *J. Appl. Meteor.*, **8**, 333–344.

Tippett, M. K., L. Goddard, and A. G. Barnston, 2005: Statistical-dynamical seasonal forecasts of central-southwest Asian winter precipitation. *J. Climate*, **18**, 1831–1843.

Wang Shaowu, 2001: *Progress in Climatologically Studies.* China Meteorological Press, Beijing, 306–311. (in Chinese)

Wu Bingyi, Zhang Renhe, and D'Arrigo Rosanne, 2008: Arctic dipole anomalies and summer rainfall in Northeast China. *Chinese Sci. Bull.*, **53**(12), 1422–1428. (in Chinese)

Zhai Panmao and Zhou Qinfang, 1997: The change of Northern Hemisphere snow cover and its impact on summer rainfalls in China. *J. Appl. Meteor. Sci.*, **8**(2), 230–235. (in Chinese)

Zheng Qinglin and Du Xingyuan, 1973: A new numerical weather prediction model utilizing multiple-instant observation. *Sci. China (Ser. A)*, **3**, 289–297. (in Chinese)

Zheng Zhihai, Ren Hongli, and Huang Jianping, 2009: Analogue correction of errors based on seasonal climatic predictable components and numerical experiments. *Acta Physica Sinica*, **58**(10), 7359–7367. (in Chinese)

Zhu Kezhen, 1934: Southeast monsoon and rainfall in China. *Acta Geographica Sinica*, **1**(1), 1–27. (in Chinese)