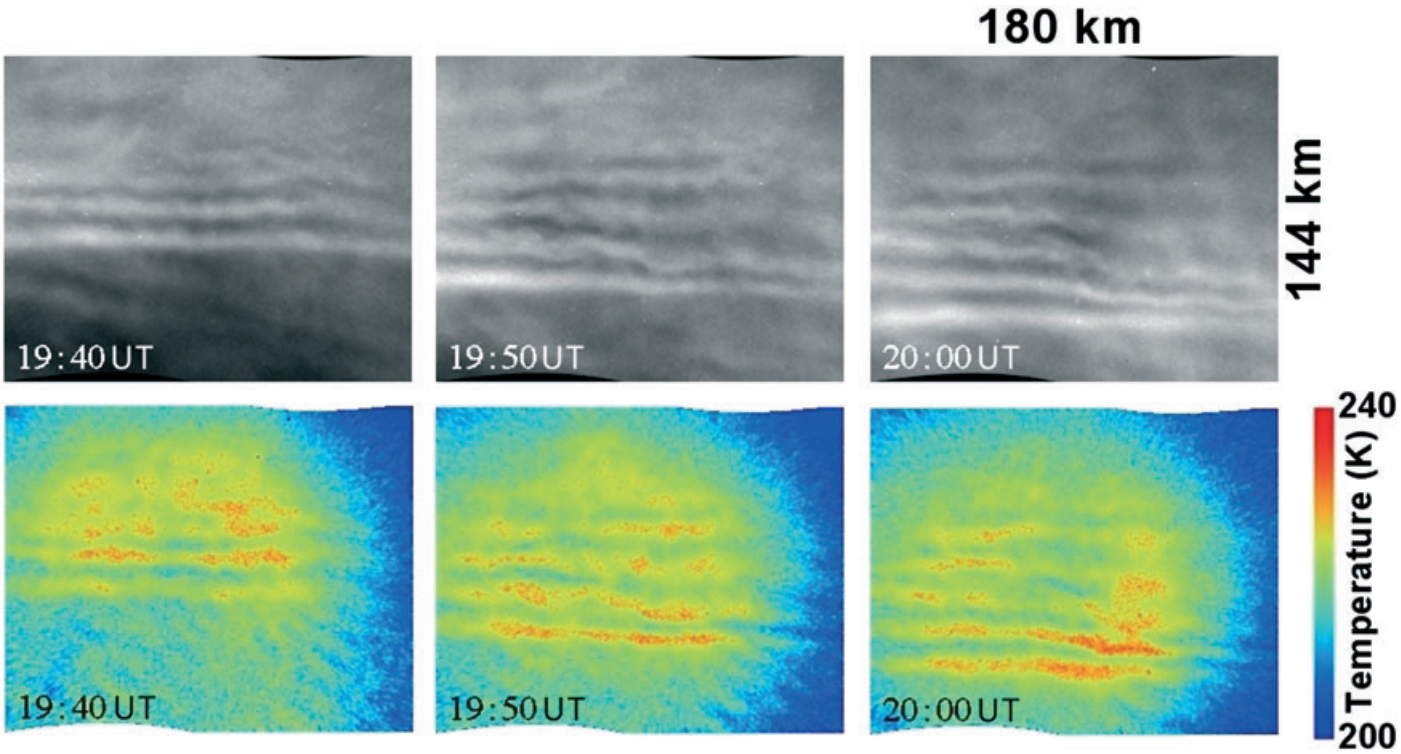
JOURNAL OF GEOPHYSICAL RESEARCH
Atmospheres
AN AGU JOURNAL

JGR

Volume 123 • Issue 1 • 16 January 2018 • Pages 1–622



AGU PUBLICATIONS

WILEY

JOURNAL OF GEOPHYSICAL RESEARCH

Atmospheres

AN AGU JOURNAL

Aims and Scope. *JGR: Atmospheres* publishes articles that advance and improve understanding of atmospheric properties and processes, including the interaction of the atmosphere with other components of the Earth system.

For submission instructions or to submit a manuscript visit: http://jgr-atmospheres-submit.agu.org.

The journal to which you are submitting your manuscript employs a plagiarism detection system. By submitting your manuscript to this journal you accept that your manuscript may be screened for plagiarism against previously published works.

JGR: Atmospheres accepts articles for Open Access publication. Please visit http://olabout.wiley.com/WileyCDA/Section/id-406241.html for further information about OnlineOpen.

Publication Charges. The publication charge income received for *JGR: Atmospheres* helps support rapid publication, allows more articles per volume, makes possible the low subscription rates, and supports many of AGU's scientific and outreach activities. Publication charge information can be found here: http://publications.agu.org/author-resource-center/publication-fees/.

To encourage papers to be written in a concise fashion, there is an excess length fee. *For JGR: Atmospheres* the fee is assessed only on the equivalent of more than 25 publication units. The excess length fee does not apply to review articles, and the editor may waive the fee on a limited number of concisely written papers that merit being longer. There is no charge for color in any format.

Individual Subscriptions. Member subscriptions are available through members.agu.org or by contacting the AGU Member Service Center. The Service Center is open from 8:00 a.m. to 8:30 p.m. Eastern time: +1 202 462 6900, +1 800 966 2481; Fax: +1 202 777 7393; e-mail: service@agu.org. Questions about meetings or membership will be referred to the appropriate staff.

Publisher. *JGR: Atmospheres* is published on behalf of the American Geophysical Union by Wiley Periodicals, Inc., 111 River St., Hoboken, NJ, 07030-5774, +1 201 748 6000.

Journal Customer Services. For institutional subscription information, claims and any enquiry concerning your journal subscription please go to www.wileycustomerhelp.com/ask or contact your nearest office.

Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or +1 800 835 6770 (toll free in the USA & Canada).
Europe, Middle East and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315.
Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000.
Japan: For Japanese speaking support, Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.

Visit our Online Customer Help available in 7 languages at www.wileycustomerhelp.com/ask.

Production Editor. For assistance with post-acceptance articles and other production issues please contact JGRDprod@wiley.com.

Access to this journal is available free online within institutions in the developing world through the AGORA initiative with the FAO, the HINARI initiative with the WHO, the OARE initiative with UNEP, and the ARDI initiative with WIPO. For information, visit www.aginternetwork.org, www.who.int/hinari/en/, www.oaresciences.org, or www.wipo.int/ardi/en.

ISSN 2169-8996 (Online)

View this journal online at http://jgr-atm.agu.org.

Cover: In Pautet et al. (https://doi.org/10.1002/2017JD027046): Times series of frontal events observed on 4 May 2013. (top) OH(3,1) band intensity. (bottom) OH(3,1) rotational temperature. See pp. 160–173.

Journal of Geophysical Research Atmospheres

Volume 123, Issue 1, 16 January 2018
Pages 1– 622
ISSN 2169–897X (print); ISSN 2169–8996 (online)

Climate and Dynamics

Aerosol and Clouds

Composition and Chemistry

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2017JD027281

Key Points:
- An extended-range real-time forecast of the low-frequency rainfall over the LYRV is conducted
- A time-varying high-order ECAR model is set up based on data-driven modeling
- Predictability of the low-frequency rainfall is estimated with the ECAR models for the 20 to 30 day oscillations

Correspondence to:
Q. Yang,
yqm0305@263.net

Citation:
Yang, Q. (2018). Predictability and prediction of low-frequency rainfall over the lower reaches of the Yangtze River valley on the time scale of 20 to 30 days. *Journal of Geophysical Research: Atmospheres*, *123*, 211–233. https://doi.org/10.1002/2017JD027281

Received 13 JUN 2017
Accepted 7 DEC 2017
Accepted article online 18 DEC 2017
Published online 10 JAN 2018

Predictability and Prediction of Low-Frequency Rainfall Over the Lower Reaches of the Yangtze River Valley on the Time Scale of 20 to 30 days

Qiuming Yang[1]

[1]Jiangsu Meteorological Institute, Nanjing, China

Abstract This paper presents a predictability study of the 20–30-day low-frequency rainfall over the lower reaches of the Yangtze River valley (LYRV). This study relies on an extended complex autoregressive (ECAR) model method, which is based on the principal components of the global 850 hPa low-frequency meridional wind. ECAR is a recently advanced climate forecast method, based on data-driven models. It not only reflects the lagged variations information between the leading low-frequency components of the global circulation and rainfall in a complex space, but also displays the ability to describe the synergy variations of low-frequency components of a climate system in a low dimensional space. A 6-year forecast experiment is conducted on the low-frequency rainfall over the LYRV for the extended-range daily forecasts during 2009–2014, based on the time-varying high-order ECAR. These experimental results demonstrate that the useful skills of the real-time forecasts are achieved for an extended lead-time up to 28 days with a fifth-order model, and are also shown to be 27-day lead for forecasts which are initiated from weak intraseasonal oscillation (ISO). This high-order ECAR displays the ability to significantly improve the predictions of the ISO. The analysis of the 20–30-day ISO predictability reveals a predictability limit of about 28–40 days. Therefore, the forecast framework used in this study is determined to have the potential to assist in improving the real-time forecasts for the 20–30-day oscillations related to the heavy rainfall over the LYRV in summer.

Plain Language Summary In recent years, the study of extended-range forecasting has become not only a hot topic of global meteorological research but also a difficulty. It is necessary to study the forecast methods and models, as well as the predictability, from multiple perspectives and angles. Using a climate forecast method that is based on data-driven modeling is one of the main ways to extend the lead time in the extended range. This paper presents that the forecast skill specific to 20 to 30 day oscillation affected the heavy rainfall process in the lower reaches of Yangtze river valley for a suite of the extended complex autoregressive model (ECAR) models, which had a good forecast skill at the lead time of approximately 28 days. It has a forecast ability far superior to the traditional autoregressive model. These ECAR models are based on the major lagged correlations with multiple different laggings hidden in a large number of observation data, along with being completely driven by the dynamic data. It is an important way to significantly improve the extreme weather forecasting accuracy for the 10 to 30 day extended-range weather forecast.

1. Introduction

The 10 to 30 day extended-range weather forecast is found to be simultaneously influenced by the initial conditions and atmospheric external force factors. Therefore, a combination of the initial meteorological conditions, along with the oceanic, atmospheric, and climate influencing factors, is required. The observational data had complex, comprehensive, and global characteristics. These large amounts of scientific data are used to reflect and characterize the complex natural phenomena and relationships, with high data correlation and multiple data attributes. This results in a very complicated forecasting process (Hoskins, 2013; Waliser et al., 2003; Zhang et al., 2013). However, a single classical data analysis method cannot fully and effectively complete such a large amount of data analyses. Hence, by crossing and integrating multiple data analysis methods and techniques, the partial valid data could be extracted from the very large amounts of total data, in order to more effectively obtain comprehensive low-frequency variation information than those that have been obtained by previous sampling analysis techniques. It is found that these obtained partial valid data show the warning and insight of an extreme weather message. This message could then be seen as a new signal for extended-range weather changes.

The understanding of the Earth's climate systems is being continuously improved, especially in regard to the extended-range forecast information sources, such as the intraseasonal oscillation (ISO) (Lee et al., 2013; Madden & Julian, 1971; Waliser et al., 2003), the North Atlantic Oscillation (Cassou, 2008; Lin & Brunet, 2011), the interactions between the stratosphere and troposphere circulation (Baldwin et al., 2003), and so on. These methods provide a development foundation for extended-range forecast. When the aforementioned particular atmospheric flow pattern signals appear continuously, the 10 to 30 day extended-range weather processes often have very good predictability. Also, this will often result in extremely abnormal weather events in the related areas, such as heavy rainfall events, as well as severe rises or drops in temperatures. Various principal oscillation patterns can be directly extracted from the observational data with a multivariate, ultra-high dimension, super long sequence, highly coupled correlation, and high spatiotemporal attributes, which have been found to be closely related to the extreme weather in a certain time scale and a certain area (such as the Yangtze River Basin). Therefore, based on the previous results regarding the interannual change laws of the different propagation characteristics and strengths of these oscillation patterns, the dynamic data have been used to understand the complex low-frequency changes and data-driven building system, reducing the system's complexity, and establishing a simplified forecasting model (Yang, 2015). This forecasting model can significantly extend the lead time of the low-frequency component that corresponds to the extreme weather processes over the specific areas.

The study of extended-range forecast has become a hot topic of global meteorological research in recent years. However, these forecasting processes have been proven to be difficult (Brunet et al., 2010; Hoskins, 2013). It is necessary to study the forecasting methods and models, as well as the predictability, from multiple perspectives and angles. Generally speaking, the basic methods of extended-range weather forecasting have mainly included two types: the dynamical methods (Fu et al., 2013; Jones et al., 2012; Miyakoda et al., 1983; Vitart & Molten, 2010) and the statistical methods (Waliser, 2012; Waliser et al., 2003). The former methods include atmospheric circulation models and ensemble numerical prediction models (Fu et al., 2013; Hudson et al., 2011; Kim & Webster, 2010; Martin et al., 2010; Miura et al., 2007; Rashid et al., 2011; Sabeerali et al., 2017), along with predictable component extracting models (Chou et al., 2010; Feng et al., 2013; Zheng et al., 2013), and so on. Meanwhile, with the increased data, as well as the improvements in the performances of the models, the lead times of the Madden-Julian Oscillation (MJO) (Madden & Julian, 1971) have been gradually extended to approximately 10–15 days. Over the past decades, dynamical models have assisted greatly in the improvements of the forecasting techniques of the tropical MJO, in which the hindcast experiment in the European Centre for Medium-Range Weather Forecasts model showed that the useful skills are up to 25 to 30 day lead in the years with significant MJO (Neena et al., 2014). Also, the useful skill is up to about 10–15 days for the global patterns of the dominant quasi-biweekly oscillation for the hindcast experiment in the National Centers for Environmental Prediction (NCEP) Climate Forecast System version 2 (Jia et al., 2013). However, the current dynamical models have difficulty in predicting the ISO signals in midlatitude regions for the time scale of 20–80 days. Due to the fact that among all of the types of uncertainties in the weather and climate numerical forecasting models, the extended-range forecasting errors for the low-frequency systems of the global atmosphere are still observed to be distinctly growing, especially over the extratropics. The sources of these uncertainties, including the atmospheric chaotic nature and complex coupling processes, continue to be examined. The latter refers to the statistical models based on the atmospheric low-frequency signal evolution, which mainly includes the following: the principal oscillation pattern analysis (POP) (Hasselmann, 1988; von Storch & Xu, 1990; Yang, 1998, 2011; Yang, Li, et al., 2012), singular value decomposition (Hsu et al., 2015; Waliser et al., 1999; Zhu et al., 2015), singular spectrum analysis (SSA) (Mo, 2001; Vautard & Ghil, 1989), analog forecasting (Alexander et al., 2017; Xavier & Goswami, 2007), regression analysis (Jiang et al., 2008; Kang & Kim, 2010; Wheeler & Hendon, 2004), linear inverse models (Cavanaugh et al., 2014), multilevel regression models (Kondrashov et al., 2013), and physics-constrained models (Chen et al., 2014). Since 1990, these experimental studies have rapidly increased in numbers. Some of the aforementioned methods have been used for real-time forecasting. The general average lead time is now up to 20 days or higher for the MJO. More recently, based on the kernel analog forecasting, the forecast skill was of up to about 50 days for the hindcast of the MJO and boreal summer intraseasonal oscillation (BSISO) (Wang & Xie, 1997) in the tropics on a time scale of 30–60 days using the nonlinear Laplacian spectral analysis indices (Alexander et al., 2017). In addition, there are the neural network models (Borah et al., 2013; Love & Matthews, 2009), empirical mode decomposition (Love et al., 2008), and physical decomposition principle

of regional-scale atmospheric transient anomaly (Qian, 2012), and the comprehensive prediction methods of these various models have displayed certain forecasting skills.

Although the prediction skill has been slowly increasing in recent years, the majority of the current operational dynamical models still have poor skills in predicting the ISO over the subtropical and midlatitude areas. Given the incapability of dynamical models in reproducing the observed ISO, empirical statistical models have been constructed to further improve for the practical ISO predictions (Waliser, 2012). It has been found that at present, the statistical methods that have been established on the basis of the research achievements of the ISO for the rainfall variations have displayed better forecasting abilities on extratropical ISO than those achieved using dynamical model (Yang, 2015; Zhu & Li, 2017). Therefore, the statistical model remains an effective way to study the ISO predictability over the extratropical regions.

As an important background of the high-frequency weather changes, the ISO with the cycles of 30–60 (MJO, Madden & Julian, 1971; BSISO1, Wang & Xie, 1997; Lee et al., 2013), 20–30 (Yang, 2009), and 10–20 (Kikuchi & Wang, 2009) and 10–30 days (BSISO2, Lee et al., 2013) has direct links between the weather and the climate. Also, the nature of these ISO modulations for the occurrence, strength, and spatial patterns of the regional climate events is observed to vary with background climate states. Therefore, the ISO has been considered to fill the "forecasting gap" between the weather forecast and the seasonal forecast. The ISO can be used as an important forecasting source for the extended-range forecasts. During the past 20 years, a great deal of research has also been conducted in this field in China, and important progress has been made (Yang, Song, et al., 2012). The results of these studies have shown that the floods in eastern China are largely correlated to the ISOs, particularly the rainfall in the middle and lower regions of the Yangtze River. These are closely related to the multiple low-frequency oscillations and their propagations over extratropics of Eastern Asia in the time scales of 20–30, 30–50, and 50–80 days. The previous studies have also shown that a single oscillation influence relationship exists between the rainfall over the lower reaches of the Yangtze River valley (LYRV: 30.5°–32.0°N, 118.0°–122.5°E; represented by the purple rectangle in Figure 1a) and the 20 to 30 day or 30 to 50 day or 50 to 80 day ISO strengths. These ISO changes have the more significant positive correlation to the heavy rainfall in the LYRV (Yang, 2009), in particular the 20 to 30 day ISO, while the influence of the MJO has been found to be relatively weak. These ISOs of the rainfall over the LYRV are not single-frequency sinusoidal oscillations. Instead, these are known to be broadband quasiperiodic oscillations (20–80 days) and have considerable event-to-event variability. Also, there is a chance that the associated high-frequency variabilities have modulated the variabilities in the observed data. Hence, the real predictability would be much lower. Figures 1b–1d show the time series of intensity (standard deviation of daily filtered rainfall during the period from 1 June to 31 August, unit: mm) of the 20 to 30 day, 30 to 50 day, and 50 to 80 day oscillations for rainfall and the number of the heavy precipitation process (from the average of the daily precipitation with greater than or equal to 25 mm) over the LYRV in June–August from 1979 to 2014, respectively. It is found that correlations of intensity of these ISOs to the number of the heavy precipitation process are 0.564, 0.369, and 0.391, respectively, and these represent the statistical significance of the correlation coefficient at the 99.9, 95, and 98% leveld, respectively. Therefore, among these oscillations, the intensity of the 20 to 30 day oscillation has the strongest positive correlation with the heavy rainfall processes over the LYRV (Figure 1b). In the cases of significant positive phases of the 20 to 30 day low-frequency rainfall, or positive-negative phase transformations, it is found that the probability of continuous heavy rainfall would be significantly increased. Thereby, the evolution of the 20 to 30 day oscillation would potentially provide a better indicative significance for the extended-range forecast of heavy rainfall over the LYRV. Consequently, continued examinations and improvements of the extended-range forecast method of the 20 to 30 day rainfall ISO that described these ISOs to emphasize the diversity of their behavior, as well as the extension of the lead time, will effectively further improve the real-time forecast accuracy of heavy rainfall in the LYRV during the 10–30 days.

In regard to the extended-range forecast of the low-frequency rainfall, the majority of the statistical forecast research in China and internationally is mainly based on a statistical analysis method in a real number domain, such as linear regression analysis and similarity analysis (Kang & Kim, 2010; Seo et al., 2009; Zhu et al., 2015). Less consideration has been given to the variation law of low-frequency variables in a complex domain, particularly the generating complex sequence by a data transformation. Therefore, some important low-frequency change information contained in the observation data has been lost. In order to more effectively analyze the spatiotemporal variation characteristics of a climate field, the empirical orthogonal function

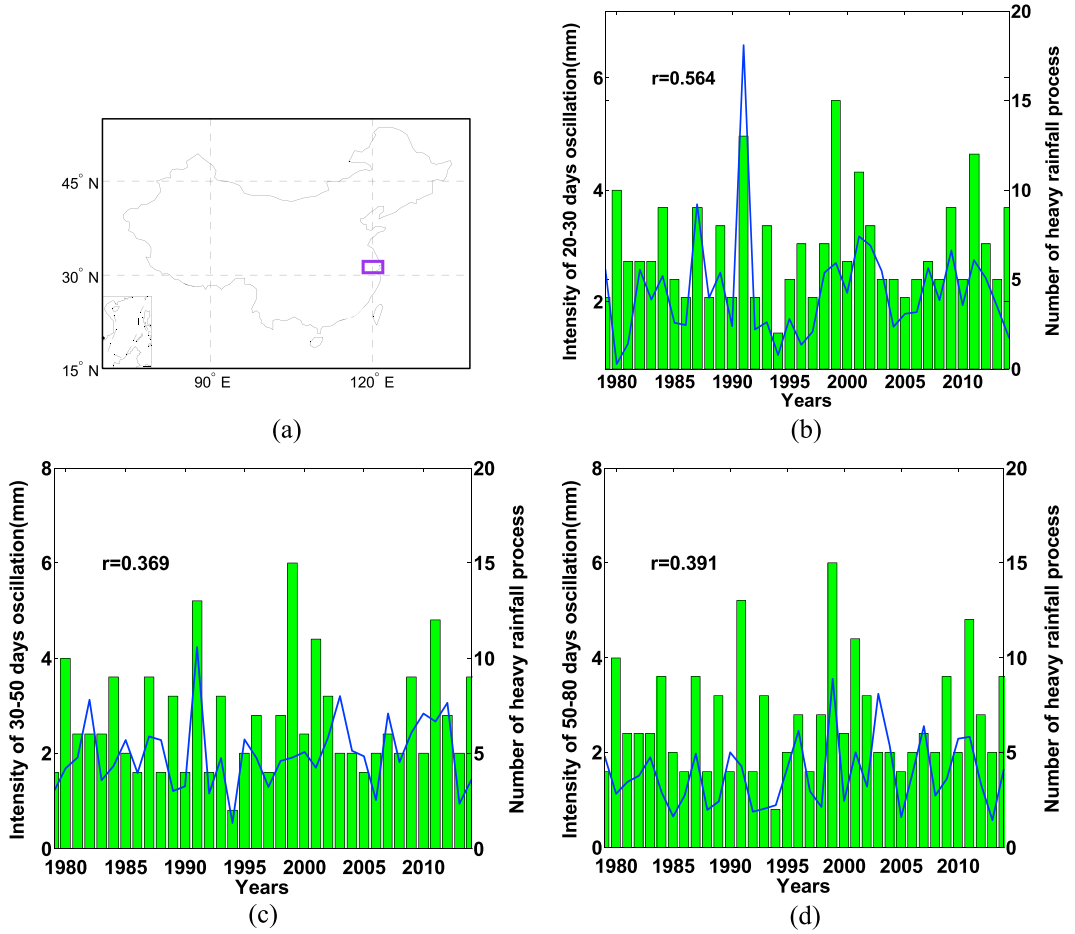


Figure 1. (a) Lower reaches of the Yangtze River valley (LYRV): purple rectangle; Time series of intensities of the (b) 20 to 30 day and (c) 30 to 50 day and (d) 50 to 80 day oscillations for the rainfall (solid line, standard deviation of daily filtered rainfall during the period ranging from 1 June to 31 August, unit: mm) and the number of the heavy precipitation processes (bar) over the LYRV in June–August from 1979 to 2014, respectively; *r* represents their correlation coefficient.

(EOF) analysis in a real domain has been promoted to complex domains. Also, complex empirical orthogonal functions (Barnett, 1983), along with complex principal oscillation pattern analysis (CPOP) (Bürger, 1993) and other methods, have been proposed. These methods can be used to obtain complex time series using the Hilbert transform. They can also be used to construct a complex matrix for decomposition, which has the ability to better identify the main spatial and temporal variations of a standing wave and traveling wave oscillation systems. It can also overcome the weaknesses of a real EOF, which has been found to be unable to reveal the changes in traveling waves. Meanwhile, these complex principal oscillation patterns can also be used for low-frequency change forecasting research, which improves the analysis and forecasting skills to a certain extent. However, the real-time extended-range forecasting of the ISO has been found to still be dominated by a real-domain statistical analysis method for the original sequence (Kang & Kim, 2010; Zhu et al., 2015). Therefore, some important low-frequency change information has continued to be ignored for the generating complex sequence by a data transformation.

Consequently, in extended-range forecasting research, the low-frequency variables can be converted from real spaces into complex sequences on Fourier space, which will then correspond to the characteristics of the entire physical field characterized by the Fourier coefficients (amplitude). This conversion is required in order to isolate more low-frequency information than in real spaces and set up a complex-domain statistical model for forecasting. In the future, this will become one of the important ways in which to improve the predictabilities and extend the lead times. Through the constructing an extended complex matrix (ECM), an extended complex autoregressive (ECAR) model was established (Yang, 2014), which changed over time.

This low-frequency component forecast model (LFCF) based on ECAR is referred to as LFCF2.0 (Yang, 2015) (http://www.lcjrerf30.org/english/). In 2013, this ECAR performed well in the hindcast experiments of the low-frequency rainfall in the LYRV for the 1 January 1 to 31 December 2013 interval in a year with the stronger 20 to 30 day oscillation. The lead time of a 20 to 30 day low-frequency component of rainfall over the LYRV is extended to 43 days. However, it is not clear whether or not the predictabilities for this ECAR model are associated with the 20 to 30 day low-frequency component for the real-time forecast on data from multiple years.

Therefore, this study focuses on the predictability of the low-frequency rainfall of the 20 to 30 day oscillations using the ECAR model. The long sequence (1 January 1979 to 31 December 2008) low-frequency principal component of the global 850 hPa circulation, as well as the 20 to 30 day low-frequency rainfall over the LYRV (closely related to the heavy rainfall process) (Yang, 2009), is used in this study. The dynamic data are adopted to drive the complex low-frequency change process and system construction. Then, a series of the time-varying *p*-order ECAR models is established. Moreover, the intraseasonal forecast skill is investigated using observed 20 to 30 day low-frequency rainfall over the LYRV for the period of 2009–2014, and the interannual variations of predictability, along with the possible reasons, are also discussed.

The structure of this paper is as follows. The ECAR model and the forecasting method of this study are described in section 2. Section 3 introduces the data. The prediction skill of the 20 to 30 day low-frequency rainfall over the LYRV and the predictability are given by forecasts of the time-varying high-order ECAR in section 4, and a summary and discussion are presented in section 5.

2. Methodology

2.1. Extended Complex Autoregressive Model (ECAR)

The filtering data array with M observation samples and N grids in the meteorological field are as follows: $\mathbf{S} = (S_1, \ldots, S_N) = (s_{i,j})$, $i = 1, 2, \ldots, M$, and $j = 1, 2, \ldots, N$. The principal component analysis (PCA) is carried out on $\mathbf{S}$, dimensioned $M \times N$.

$$\boldsymbol{S} = \boldsymbol{T}\boldsymbol{V}^T \tag{1}$$

where $\mathbf{T} = (T_1, \ldots, T_N)$ and $\mathbf{V} = (V_1, \ldots, V_N)$ are the time coefficient and matrix composed of the eigenvectors, respectively. Then, the leading L principal spatial modes $\mathbf{V} = (V_1, \ldots, V_L)$ and the time changes $\mathbf{T} = (T_1, \ldots, T_L) = (t_{i,j_1})$ (t_{i,j_1} is the principal component of the low-frequency circulation) can be obtained. The extended data array is constructed as follows:

$$\mathbf{F} = (T_1, \ldots, T_L, r_{lcj}) = (f_{i,j_2}) \tag{2}$$

where $j_1 = 1, 2, \ldots, L$, r_{lcj} is low-frequency rainfall series, and $j_2 = 1, 2, \ldots, L + 1$. A one-dimensional Fourier transformation is carried out on the $L + 1$ time series of $\mathbf{F} = (F_1, \ldots, F_{L+1}) = (f_{i,j_2})$ as follows:

$$\widetilde{f}_{i,l} = \sum_{j_2=1}^{L+1} \left[f_{i,j_2} \exp\left(-\boldsymbol{I} \frac{2\pi}{L+1} (j_2 - 1)(l-1) \right) \right] \tag{3}$$

where $\exp\left(-\boldsymbol{I}\frac{2\pi}{L+1}(j_2-1)(l-1)\right) = \cos\left(\frac{2\pi}{L+1}(j_2-1)(l-1)\right) - \boldsymbol{I}\sin\left(\frac{2\pi}{L+1}(j_2-1)(l-1)\right)$, $\boldsymbol{I} = \sqrt{-1}$, $i = 1, 2, \ldots, M$; $j_2 = 1, 2, \ldots, L + 1$, and $l = 1, 2, \ldots, L + 1$.

Therefore, the $L + 1$ complex time series $\widetilde{f}_{i,l} = a_{i,l} + b_{i,l}\boldsymbol{I}$ could be obtained to construct the extended complex matrix (ECM) as follows:

$$\widetilde{\boldsymbol{F}} = \left(\widetilde{f}_{i,l}\right) = \left(\widetilde{F}_1, \ldots, \widetilde{F}_{L+1}\right) \tag{4}$$

where $l = 1, 2, \ldots, L + 1$, each complex component $\widetilde{f}_{i,l}$ represents different waves, and these Fourier coefficients denote the main characteristics of the entire extended physical field (phase and amplitude). For each complex component $\widetilde{f}_{i,l}$, a *p*-order complex autoregressive model is constructed.

$$\widetilde{f}_{i+1,l} = B_0 + \sum_{k=1}^{p} B_k \widetilde{f}_{i-k+1,l} \tag{5}$$

A least-square method in a complex number domain is used to calculate parameter estimation B_k, $k = 0, 1, \ldots, p$ (in which $p = 1, 2, 3, \ldots$).

According to equation (5), the forecast value $\hat{\tilde{f}}_{M+1,l} = \hat{a}_{M+1,l} + \hat{b}_{M+1,l}\boldsymbol{I}$ at the moment of $M + 1$ could be obtained. Then, using a one-dimensional Fourier inverse transformation, the forecast value $\hat{f}_{M+1,j_2}$ of each low-frequency component is achieved. Equation (5) could be used for a K-step recursion, in order to obtain the forecast value $\hat{f}_{M+K,j_2}$ on the Kth day, where $j_2 = 1, 2, \ldots, L + 1$, in which $\mathrm{Re}\left(\hat{f}_{M+K,L+1}\right) = \hat{r}_{\mathrm{lcj}}(M+K)$ is the forecast value of the low-frequency rainfall component. Equation (5) is referred to as an extended complex autoregressive model (ECAR) (Yang, 2014), in which the parameter estimation is conducted by the least-square method in a complex number domain as an extension of that in the real number domain. This ECAR method differs from a wavelet-based prediction (Kang & Kim, 2010) for the sum of the predicted real parts of each spectral band using a real linear autoregression method and is instead performed by using a data transformation and complex linear autoregression method for a given spectral band. This method has the ability to describe the time variation characteristics of each low-frequency variable on complex plane for the generating complex sequence. Therefore, more comprehensive variation information can be obtained than that in real space, and the evolutions of the relationships between the observed low-frequency variables can be relatively steadily revealed.

2.2. Forecasting Methodology

In this study, the global 850 hPa meridional wind field data during the period ranging from 1 January 1979 to 31 December 2008 (10,958 days) are used to determine the global low-frequency principal components of the circulations, so as to establish the ECAR to conduct on the forecast experiments of the daily low-frequency rainfall of the LYRV from 1 January 2009 to 31 December 2014 (2,191 days). Singular spectrum analysis (SSA) (Vautard & Ghil, 1989; Mo, 2001) has a function of self-adaptive filtering signal reconstruction with precise positioning of the various main signal changes and can minimize the boundary effects of the traditional filtering methods with a T-EOF extension. Through these corrections inside SSA, it is well suited for the real-time monitoring of the ISO. Hence, the SSA with a T-EOF extension is carried out on the real-time principal components (PCs) of the circulations, as well as the observed daily rainfall over the LYRV without detrending or temporal filtering in this study. Then, the respective corresponding T-EOF component is reconstructed in order to obtain the component sequence of the 20 to 30 day oscillation signal of circulation and rainfall, and this is used as the basic data of the forecasting test. Meanwhile, an ECAR model is established for the forecasting of the daily extended-range changes of the 20 to 30 day rainfall component over the LYRV, for each day from 1 January 2009 to 31 December 2014.

As shown in the previous research, a significant influence relationship exists between the heavy rainfall in the LYRV and the ISO intensity in the different time scales of 20–30 days, 30–50 days, and over 50 days (Yang, 2009; Yang, Song, et al., 2012). For example, the ISO intensity of the 20 to 30 day rainfall is found to have the most significant positive correlation with the interannual variabilities of the heavy rainfall frequency in the LYRV from May to August. These interannual variations of the heavy rainfall frequency also show significant quasi-biennial oscillations. Through the analysis of the principal oscillation pattern (POP) (Hasselmann, 1988), it is found that from May to August in the time scale of 20–30 days, two 20 to 30 day principal oscillation patterns (POP1 and POP2) exist in the global 850 hPa circulation. One is a southern circumglobal teleconnection wave train (SCGT) of the middle latitudes in the Southern Hemisphere with an eastward propagation. The other is a tropical western Pacific pattern, with a southward propagation. These two POP patterns are closely related to the 20 to 30 day low-frequency rainfall, as well as the heavy rainfall processes in the LYRV. Thus, the principal components of the 850 hPa low-frequency meridional wind fields with the time scale of 20–30 days, with larger variance worldwide contributions (90°N–90°S, 0°–360°), are used in this study to establish the forecasting model.

Then, based on the PCA, the leading four EOFs of the global 850 hPa low-frequency meridional wind (V_1, …, V_4) are identified (time scale: 20 to 30 days, data period: 1979 to 2008, sequence length: 10,958 days, and the Butterworth filtering was used to obtain the global 850 hPa low-frequency meridional wind field), with the explained variances of 3.83, 3.55, 2.94, and 2.88%, respectively. Figures 2a–2d illustrate the structure of these leading four EOFs (V_1,…, V_4) of the low-frequency meridional wind. Modes 1–4 are expressed as a wave train structure with a zonal propagation (Yang, 2009), among which modes 1 and 2 and modes 3 and 4 correspond to the two different types of SCGT propagations, respectively. For example, modes 1

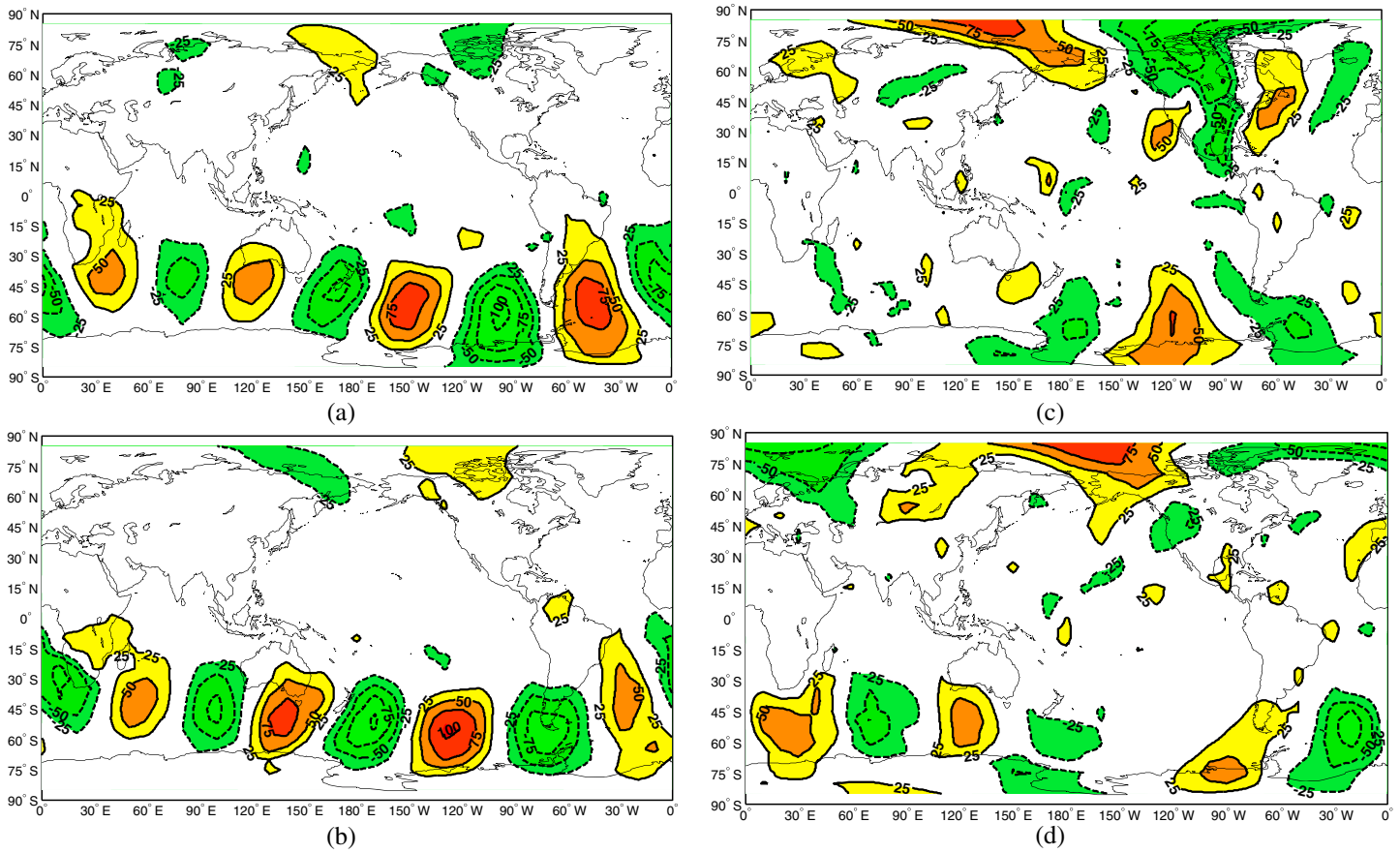


Figure 2. Principal spatial modes of the global 850 hPa 20 to 30 day low-frequency meridional wind field from 1979 to 2008. (a–d) Modes 1–4; the values in the figure have been multiplied by 1,000, and the dashed lines represent the negative values.

and 2 correspond to a wave train that propagated in the subtropical and midlatitudes of the Southern Hemisphere (Figures 2a and 2b), while modes 3 and 4 correspond to the other wave trains that propagated in the middle and high latitudes of the Southern Hemisphere (Figures 2c and 2d). The corresponding time coefficients also display significant interannual variabilities (figure omitted), which reflect the diversity of the SCGT propagation modes in the extratropics of the Southern Hemisphere. Furthermore, the stability of the eigenvectors (leading modes of the global 850 hPa low-frequency meridional wind) was examined by using the absolute value of cosine on the angle of intersection between the eigenvectors for the different sample sizes (M = 10,958 days (1979–2008, 30 years), M_1 = 9,131 days (1979–2003, 25 years)). It is found that these first four eigenvectors associated with SCGT are more stable, with the larger absolute value of cosine on the angle of intersection between the eigenvectors of 0.928, 0.890, 0.861, and 0.889 for the first four eigenvectors. It is worth noting that such a wave train propagation similar to the SCGT (modes 1–4) will mainly affect the rainfall changes in the midlatitudes of the Southern Hemisphere. In addition, these 20 to 30 day oscillations are also performed in the changes of the eddy kinetic energy at a hemispheric scale. An example would be the baroclinic annular mode, which is related to the two-way feedback between the baroclinity and the eddy heat flux (Thompson & Barnes, 2014; Thompson & Woodworth, 2014). Moreover, from the viewpoint of the correlation analysis in a sliding 300 day window among these time coefficients PC1 to PC4, which correspond to the leading four modes of the global low-frequency circulations and the low-frequency rainfall over the LYRV, the sliding correlation coefficients are found to display significant seasonal and interannual changes (figure omitted). Furthermore, definite seasonal and irregular changes are found to exist in the time intervals with strong positive and negative correlations. Therefore, through an atmospheric teleconnection, they are also able to in different ways indirectly affect the low-frequency component changes in the 20 to 30 day rainfall of the LYRV, as well as the formation of the heavy rainfall

processes, which are associated with the heating processes in the tropical Indian Ocean, and the interactions of the circulations in both hemispheres (Yang, 2009). It is therefore concluded that the variations of these leading modes are used as the important forecasting sources of the extended-range heavy rainfall forecast over the LYRV.

In each of the independent forecasting experiments conducted for the 20 to 30 day rainfall of the LYRV, the abovementioned leading four modes of the global low-frequency circulations are used. Also, the daily observed global meridional wind occurrences from 2009 to 2014 are projected onto these leading four EOFs (V_1,…, V_4) from the global low-frequency meridional wind during the period ranging from 1 January 1979 to 31 December 2008 (10,958 days), in order to obtain the observations of the first four principal components PC1 to PC4 (including the daily high-frequency disturbances). Then, these four observed PCs, along with the daily rainfall in the LYRV from 2009 to 2014, are projected onto the associated T-EOFs, which corresponded to the 20 to 30 day oscillation after the SSA, in order to obtain the 20 to 30 day reconstructed component of PC1, …, PC4, along with the daily rainfall in the LYRV, T_1, T_2, T_3, T_4, r_{lcj} (T-EOF was calculated using the daily data from 1 January 1979 to 31 December 2008). This process is completed in order to build the extended data array $\mathbf{F} = (F_1, \ldots, F_{L+1}) = (T_1, T_2, T_3, T_4, r_{lcj})$ ($L = 4$) for the period from 1 January 2009 to 31 December 2014 (2,191 days). At this point, the ECAR model is set up for the extended-range experiments on the daily low-frequency rainfall over the LYRV.

In these forecast experiments, based on the abovementioned extended data array $\boldsymbol{F}$, the correlation skills of a 1 to 30 day forecast with the ECAR model are calculated for the low-frequency component r_{lcj} of rainfall over the LYRV for the period ranging from 2009 to 2014 (2,191 forecasts). A finite memory method is used to keep the subsequence M_0 constant. The forecasting of the ECAR with the initial date t_0 is carried out for the independent sample forecasting experiment in a sliding M_0 day window with the order of these ECAR p, the forecast lead time of the complex autoregressive model $\tau = 30$ days, and the length of the subsequence $M_0 = 90$ days. Also, for the forecasting initial date t_0, we use a T-EOF extension method to minimize the end effects of the SSA reconstruction, in which an extended time series is given in order to extend the record of the end point of t_0 into record of $t_0 + 1$, $t_0 + 2$,…, $t_0 + l$ ($l = 75$ days) for the real-time PC1, PC2, …, PC4, along with the daily rainfall over the LYRV. This study's modeling scheme, in which the time-varying model is set up, is found to be helpful for adapting to the variations of the correlations between the principal low-frequency components of oscillation system (from January to December). Furthermore, a series of these time-varying ECAR models also reflect the changing relationships between each low-frequency component within the complex space during the different periods to a certain extent. For the complex initial values $\widetilde{f}_{t_0,j_2}$ ($t = t_0$), an extended complex autoregressive model, $\widetilde{f}_{i+1,l} = B_0 + \sum_{k=1}^{p} B_k \widetilde{f}_{i-k+1,l}$, could be used for the K-step recursion, in order to obtain the forecast values $\widehat{\widetilde{f}}_{t_0+K,j_2}$ (complex number) on the Kth day, $j_2 = 1, 2, \ldots, L+1$. Then, through a one-dimensional Fourier inverse transformation, the forecast value $\hat{f}_{t_0+K,j_2}$ of each low-frequency component is achieved, in which $\mathrm{Re}\left(\hat{f}_{t_0+K,L+1}\right) = \hat{r}_{\mathrm{lcj}}(t_0 + K)$ is the forecast value of the low-frequency rainfall component. The initial values $\widetilde{f}_{t_0,j_2}$ (complex number) are derived from the one-dimensional Fourier transformation. In the time-varying complex autoregressive model, the parameter $B_k = B_k(t_0)$, $k = 0, 1, \ldots, p$, also varies with the initial time t_0 in a sliding M_0 day window. Thus, the initial values $\widetilde{f}_{t_0,j_2}$ and complex parameter $B_k = B_k(t_0)$ play a key role in predictions.

In this forecasting scheme, the initial dates t_0 are from 31 December 2008 to 30 December 2014 (2,191 days) for forecasting 30 days into the future ($t_0 + 1$, $t_0 + 2$, $\cdots$, $t_0 + 30$). This does not require any future data and is therefore valid as a daily prediction. In addition, the forecasting skill of the ECAR model depends on the order (p) with multiple time steps being considered. Increasing the order increased the number of independent variables, which could potentially lead to over-fitting, whereas the higher orders may help provide the essential time evolution information that could be potentially useful. This study also considers the range of orders p from 1 to P, which represented temporal memories ranging from 1 day to P days prior to the forecast day. Finally, a time-varying high-order ECAR model is constructed with optimal parameter p. The schematic representation is shown in Figure 3.

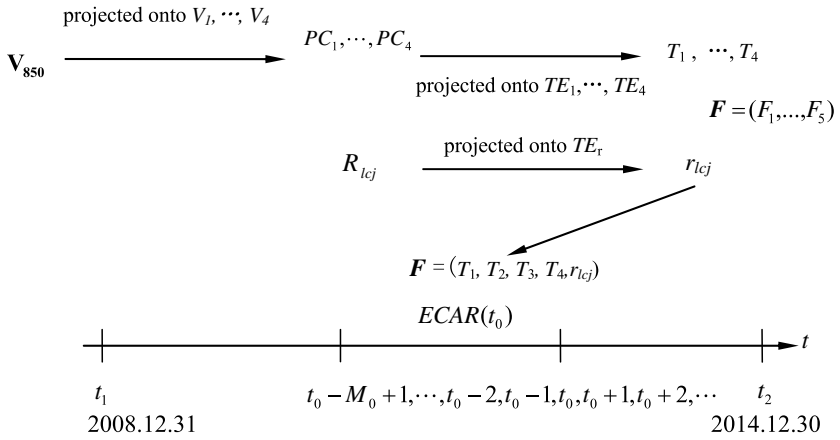


Figure 3. Schematic representation of the time-varying ECAR forecasting model. V_{850}: real-time global 850 hPa meridional wind; T_1, T_2, T_3, T_4: real-time low-frequency principal components of the global 850 hPa meridional wind; r_{lcj}: real-time low-frequency rainfall; t_0: initial time; M_0: length of the subsequence; TE_1,…, TE_4: The respective T-EOFs of the principal components of the 20 to 30 day global 850 hPa meridional wind for the first four modes V_1,…, V_4 during the period ranging from 1 January 1979 to 31 December 2008 on the time scale of 20 to 30 days; TE_r: T-EOFs of the daily rainfall over the LYRV during the period from 1 January 1979 to 31 December 2008 on the time scale of 20 to 30 days.

3. Data

Global 850 hPa meridional wind field data are sourced from the NCEP/National Center for Atmospheric Research (NCAR) daily reanalysis global wind field data, with a grid of 2.5 × 2.5° (Kalnay et al., 1996). Also, the daily rainfall data are selected from the average of 25 stations in the LYRV (30.5°–32.0°N, 118.0°–122.5°E, purple rectangle, Figure 1a). The dates used are from 1 January 1979 to 31 December 2014. The SSA is used for the original sequence of the daily rainfall over the LYRV, for the purpose of identifying the T-EOFs, and reconstructing the component sequence corresponding to the 20 to 30 day oscillation signals for the period from 1 January 1979 to 31 December 2008 (30 years). Based on an SSA window of 90 days, it is found that T-EOF5 and T-EOF6 form a doubly degenerate pair of 90° out-of-phase amplitude-modulated waves, with a spectral peak observed in the 1/(20 day) − 1/(30 day) frequency band (Figure 4a). Also, the reconstructed component RC5 + RC6 (Figure 4b) is found to display a dominant peak at 28 days in the noninteger power spectrum analysis (Schickedanz & Bowen, 1977), which explains about 26.0% of the variance in the 10 to 90 day band. This RC5 + RC6 is defined as the 20 to 30 day low-frequency rainfall component over the LYRV. Then, the daily rainfall in the LYRV is projected onto the abovementioned T-EOFs that correspond to the 20 to 30 day oscillations from the SSA for the interval ranging from 1 January 2009 to 31 December 2014. Thereby, a real-time low-frequency component sequence of the observed LYRV rainfall (Figure 4c, blue line) is obtained during the period of 2009–2014 on the time scale of 20–30 days. It is then described by RC5 + RC6 (projected onto the T-EOF5 and T-EOF6), in which a T-EOF extension method is used to minimize the end effects of the SSA reconstruction. In addition, a final 3 day running mean (i.e., from day −2 to day 0) is performed for the abovementioned real-time low-frequency rainfall in the LYRV (RC5 + RC6), in which the high-frequency disturbed signals (less than 5 days) are removed. It should be noted that this real-time RC5 + RC6 has a significantly negative correlation with the daily PC3 corresponding to the BSISO2 (BSISO2 is defined by the PC3 and PC4 of multivariate empirical orthogonal function analysis of daily anomalies of outgoing longwave radiation and zonal wind at 850 hPa (U850) in the region 10°S–40°N, 40°–160°E, which, together, mainly capture the northward/northwestward propagating variability with periods of 10–30 days during primarily the Asian summer monsoon, premonsoon, and monsoon-onset season. The PC3 has a period of about 30 days) (Lee et al., 2013), in which the simultaneous correlation coefficient is −0.092 (a significance of a 95% confidence level) based on the data from 1 January 1981 to 31 December 2012 (32 years). Then, an ECAR forecast model is constructed and the forecasts of these low-frequency rainfalls in the time scales of 20–30 days are made for the future 30 days, with each day beginning from 1 January 2009 to 31 December 2014 (6 years), in which initial dates are set from 31 December 2008 to 30 December 2014.

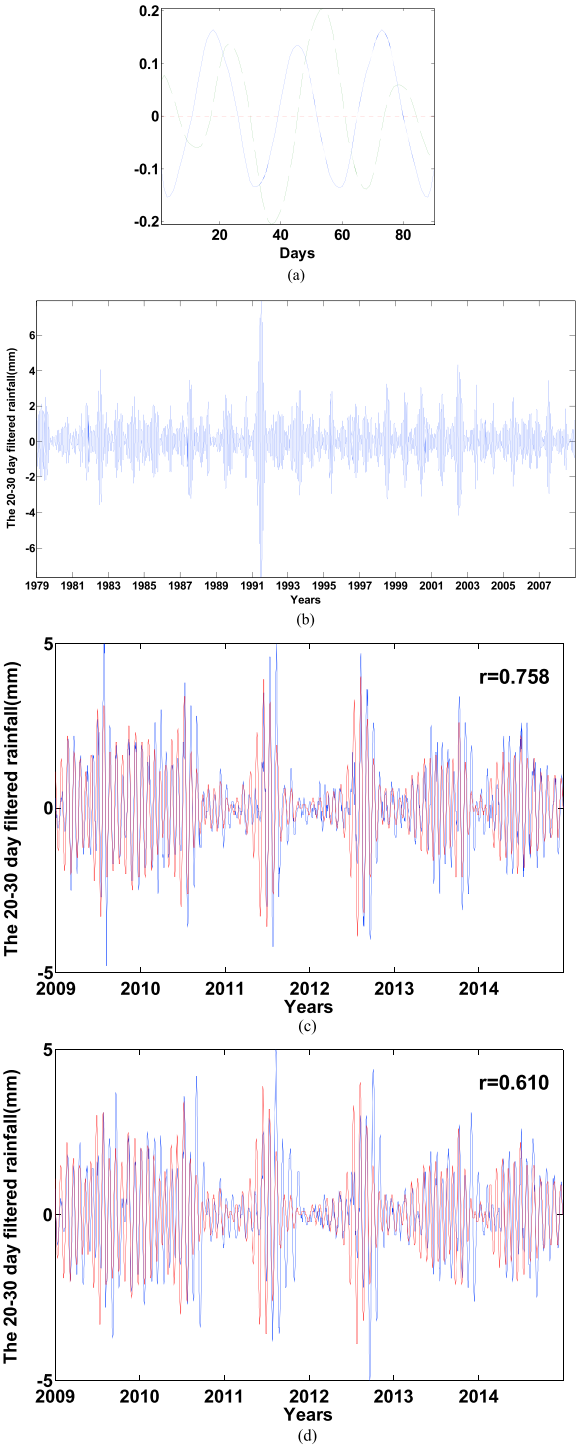


Figure 4. (a) T-EOF 5 (solid line) and T-EOF 6 (dashed line) of the SSA for the rainfall over the LYRV; (b) the 20 to 30 day low-frequency rainfall (the reconstructed component for modes 5 and 6: RC5 + RC6) during the period from 1 January 1979 to 31 December 2008; (c) time series of low-frequency rainfall (RC5 + RC6; projected onto the T-EOF5 and T-EOF6) using the entire data record available during the period from 1 January 2009 to 31 December 2014 (red line); real-time low-frequency rainfall (blue line) with the T-EOF extension, with *r* as their correlation coefficient; (d) time series of the low-frequency rainfall (red line) using the entire data record and the real-time low-frequency rainfall (blue line) without the T-EOF extension.

4. Results

4.1. Forecast Experiment of the Low-frequency Rainfall in the Lower Reaches of the Yangtze River Valley (LYRV)

4.1.1. Forecast Skills Averaged Over All Days

Since atmospheric circulation and rainfall consisted of multiple time scales, it is expected that the high-order ECAR model would perform better, which reflects the diversity of ISOs. We consider a range of orders from 1 to 9, representing temporal memories ranging from 1 day to 9 days prior to the forecast day. A 20 to 30 day lead-time is chosen as the evaluation period for determining the parameters *p*, which represented the model's extended-range forecasting skill. The model with the minimum value of the Akaike information criterion (AIC; Akaike, 1974) is selected as the potential optimal model. This AIC may be minimized over choices of *p* to form a trade-off between the fit of the model, which lowers the sum of squared residuals, and the model's complexity, which is measured by *p*. For the period of 1999–2008, the forecasting experiments are performed by the *p*-order ECAR model (using data within the training period). Figure 5a shows the RMSE, which is measured by the difference between the forecast and observation of the real-time low-frequency rainfall over the LYRV during the period of 1999–2008 (the real-time low-frequency components with the T-EOF extension). For these forecasting experiments, the variations of AIC are found to decrease with the order *p* (Figure 5b) and reached a minimum at the fifth order ($p = 5$), by which ECAR model obtained a 20 to 30 day rainfall ISO predictability of up to approximately 27 days (Figure 5a, green line) with the RMSE less than one standard deviation. It is found that the higher orders beyond the fifth increase the AIC. Thus, it is determined that the optimal parameter *p* is the fifth-order based on the AIC, as shown in Figure 5b. In this study, the fifth-order ECAR is used to obtain the skills of a 6 year real-time forecasting experiment for the independent sample, as shown in Figure 6. It can be seen in the figure that the useful skills could be detected at a lead time of approximately 28 days for the 20 to 30 day low-frequency rainfall component over the LYRV during the period of 2009–2014, in which the correlation coefficient is greater than 0.5, with the significance level of 98% (solid line in Figure 6a), considering the influence of sequence persistence. Meanwhile, the root-mean-square error (RMSE) for forecast and observation is less than 1.0 at a 28 day lead (solid line in Figure 6b), in which the daily low-frequency rainfalls of forecast and observation are normalized by the standard deviation of the data for the period of 2009–2014. It is indicated that this time-varying high-order ECAR could effectively predict the changes of the low-frequency rainfall components that are related to the 20 to 30 day oscillation propagations of circulation in the East Asian subtropical region during the future 30 days. Furthermore, Figure 6c shows the interannual variations of the forecasting skills of this high-order EACR model at the lead times of 11 (green line), 14 (red line), 17(blue line), and 20 (purple line) days for the period from 2009 to 2014, respectively. The skill of the high-order ECAR shows a significantly positive correlation at all of the lead times for the future 20 days for all of the examined years, and the useful predictions are defined by the pattern correlation between the predicted signal and the true signal above 0.5.

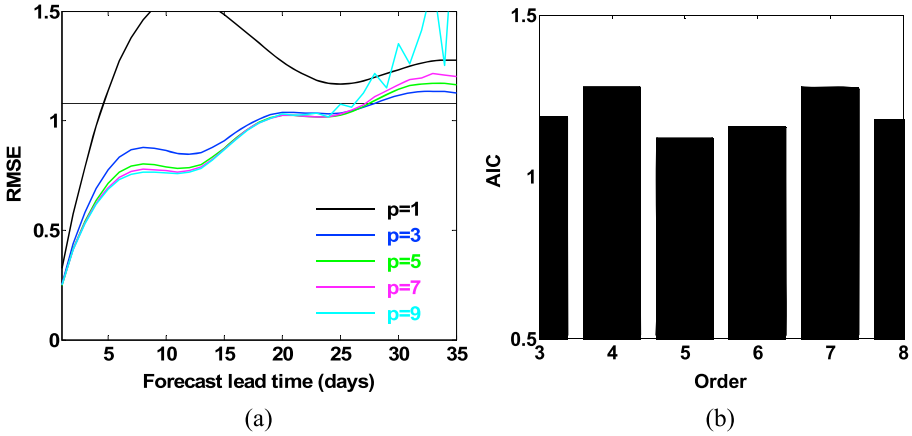


Figure 5. (a) The RMSE of the real-time low-frequency rainfall component r_{lcj} in the LYRV for a lead time 1–35 days during the period from 1999 to 2008 for the *p*-order ECAR, which the black, blue, green, magenta, and cyan lines correspond to p = 1, 3, 5, 7, and 9, respectively. The horizontal solid line in the figure represents one standard deviation. Units are mm; (b) variations of AIC for the *p*-order ECAR model during the period of 1999–2008.

Also, the model statistics and prediction skill are verified by utilizing the aforementioned six consecutive years (2009 to 2014). It is found that despite the discrepancy in the true signal variances due to the strength of the 20 to 30 day rainfall ISO activities in the different years, the autorrelations of up to 90 days (Figure 6d), the distributions of probability density functions (PDFs) (Figure 6e), and the peak of the noninteger power spectrums (Figure 6f) of the forecast at a lead time of 20 days for rainfall ISO all resembled those of the real-time rainfall ISO. Therefore, the information of the 20 to 30 day low-frequency rainfall component over the LYRV is well reflected by this high-order ECAR model for the six consecutive years.

In addition, the assessment for the first-order AR model, which has been directly established by the low-frequency component of the LYRV rainfall, is found to only achieve a useful skill at a lead time of less than 5 days. The forecasting after 5 days shows significant instabilities, with the forecasting skill being less than 0.1 (the dashed line in Figure 6a) and an RMSE of greater than 1.0 (the dashed line in Figure 6b). Thereby, on the time scale of 20–30 days, this AR model has reduced forecast ability for the 20 to 30 day rainfall ISO of the LYRV, when compared with the high-order ECAR model. This study found that one important reason for these results is that the first-order AR model has only reflected the change information of the low-frequency rainfall over the LYRV in a real space (the poor forecasting skill), which is due to the failure of capturing the oscillation structure.

4.1.2. Skill Dependence on 20 to 30 day ISO Intensity

Figure 7 shows the variations in the forecasting skill of the 20 to 30 day ISO with the different intensities using the high-order EACR model. A rainfall ISO is defined as a strong case when its intensity is larger than 1.0, and a weak case when equal to or less than 1.0, during the period ranging from 1979 to 2008, in which the ISO intensity is considered to be a 31 day running (from day −15 to day 15) standard deviation of the daily filtered rainfall. It is evident that the initially strong cases showed a systematically higher forecasting skill than the initially weak cases beginning from the 17th day of the forecast (Figure 7a). Meanwhile, the RMSE increase is observed to be relatively slower for the initially strong cases (Figure 7b), in which the daily low-frequency rainfalls of forecast and observation are normalized by the standard deviation of the data for the period of 2009–2014. The initially strong cases are determined to have a skill of 29 days and the weak cases, 27 days. It is speculated that the relatively lower forecasting skill for the initially weak cases could be partly attributed to the lack of signal in the initial conditions. The above results imply that the high-order ECAR model has a high forecasting skill when starting from an existing rainfall ISO. It should be noted that, for the initially weak cases, the forecasting skills of the high-order ECAR are still determined to be up to 27 days, which indicate the strength of this ECAR model with better forecasting stability at the longer lead times (25 to 30 days). This indicates that this high-order ECAR model with long-range dependence (multiple steps) has better stable based on diversity of irregular variation of ISO in a complex domain, which is not sensitive to the intensity of ISO.

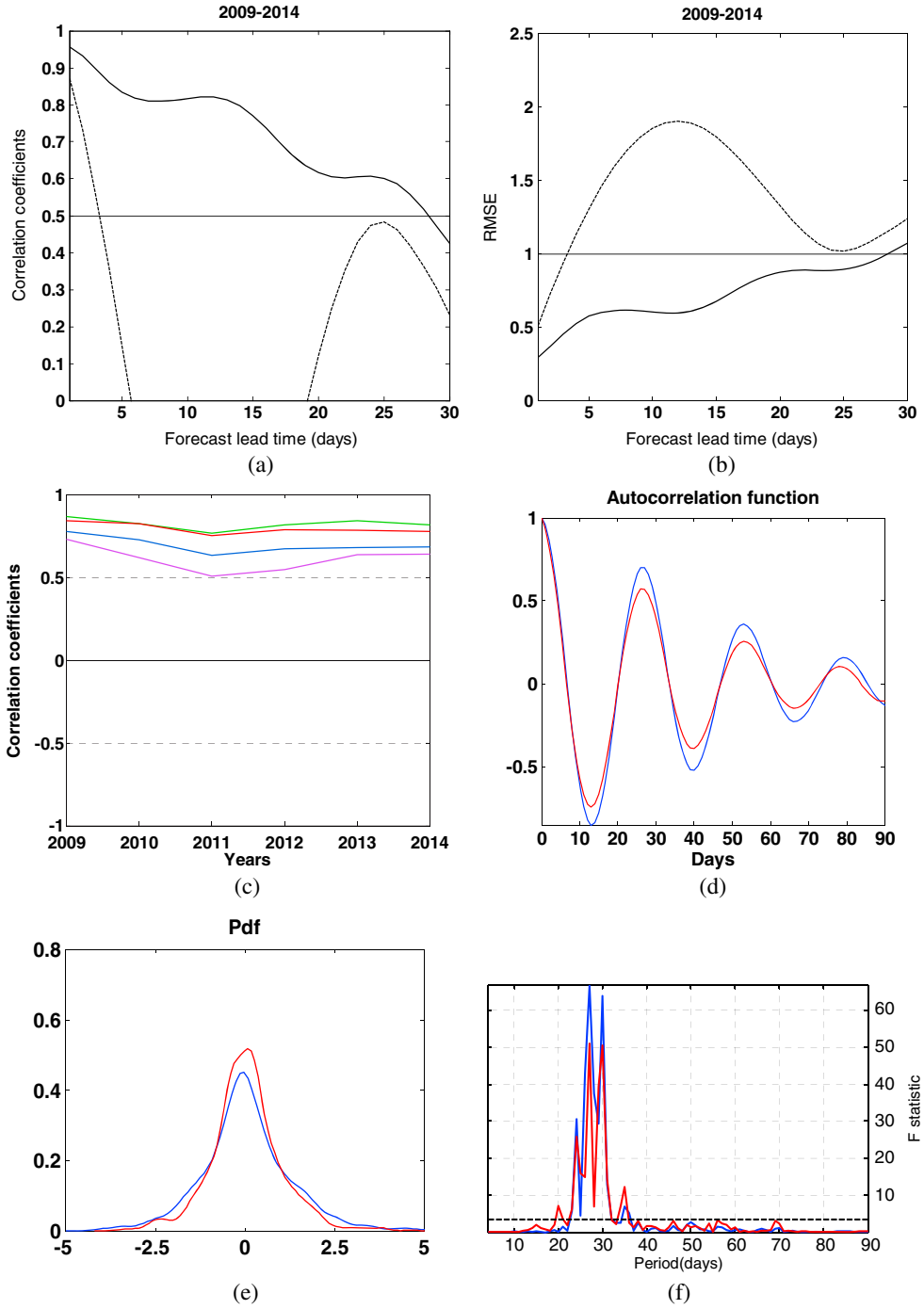


Figure 6. (a) The correlation skill of 1 to 30 day forecasting of the low-frequency rainfall component r_{lcj} in the LYRV during 2009 to 2014 on the time scale of 20 to 30 days. Solid line: the fifth-order ECAR model; dashed line: AR model. The horizontal solid line shows the threshold of the correlation coefficient $r = 0.5$; (b) root-mean-square error (RMSE, unit: standard deviation). The solid line represents the fifth-order ECAR model, and the dashed line represents the AR model, in which the daily low-frequency rainfalls of forecast and observation are normalized by the standard deviation of the data for the period of 2009–2014. The horizontal solid line in the figure represents the threshold of RMSE = 1.00; (c) interannual variations of the correlation skills for the fifth-order ECAR model. The lead times are 11 days (green), 14 days (red), 17 days (blue), and 20 days (purple), respectively. The horizontal dashed line indicates the threshold of the correlation coefficient $r = 0.5$. The useful predictions are defined by the pattern correlation between the predicted signal and the true signal above 0.5; comparison of the statistics of the real-time low-frequency rainfalls and the model signals, for the (d) autocorrelation function, (e) PDFs, and (f) noninteger power spectrum, in which the blue (red) line represents the observations (forecasts at a lead time of 20 days) of the low-frequency rainfall over the LYRV for the 20 to 30 day oscillation.

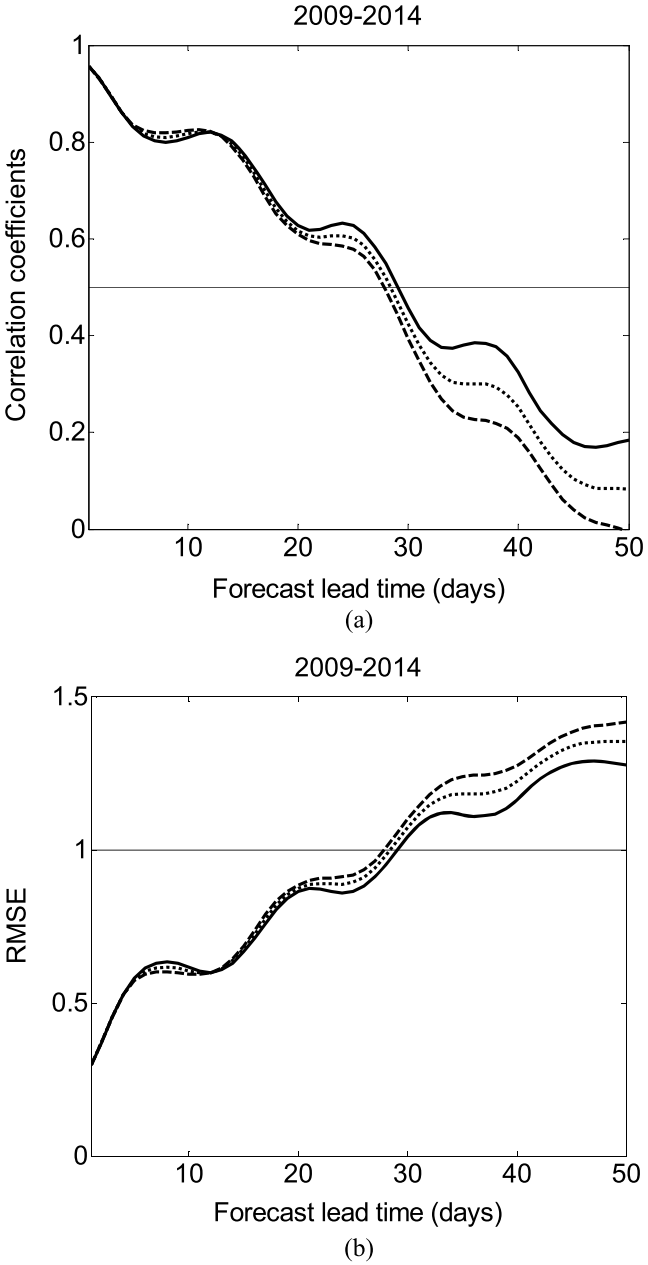


Figure 7. (a) The prediction skills as a function of the ISO strength at the start of the forecast: initially strong cases (solid line), initially weak cases (dashed line), and all of ISO cases (dotted line), utilizing the fifth-order ECAR for the 20 to 30 day oscillation of rainfall over the LYRV during the period of 2009–2014. The horizontal solid line shows the threshold of correlation coefficient $r = 0.5$; (b) the RMSE for the initially strong cases (solid line); initially weak cases (dashed line); all of the ISO cases (dotted line), unit: standard deviation. The horizontal solid line in the figure represents the threshold of RMSE = 1.00, in which the daily low-frequency rainfalls of forecast and observation are normalized by the standard deviation of the data for the period of 2009–2014.

Furthermore, Figures 8a and 8b present the skill scores based on this fifth-order ECAR from the period ranging from 2009 to 2014. The useful predictions of these 6 years all exceeded a 20 day lead time, and in particular, for the years 2009 to 2010, there are skillful predictions out to about 30 days. In these predictions, the ECAR model obtains a 20 to 30 day rainfall ISO predictability of up to approximately 36 days for the year 2009, with the correlation skill being greater than 0.5 (Figure 8a) and RMSE of less than 1.0 (Figure 8b), in which the daily low-frequency rainfalls of forecast and observation are normalized by the standard deviation of the data for each year. Figures 9a–9f show the daily change curves of the real-time forecasting (dashed line) for the daily 20 to 30 day low-frequency rainfall component at a 20 day lead during these 6 years (2009 to 2014). The forecasting skills are determined to be 0.73, 0.62, 0.52, 0.54, 0.63, and 0.64, respectively (365 forecasts for 2009, 2010, 2011, 2013, and 2014, or 366 forecasts for 2012, and the initial date of the forecasting is the previous 12 December, ..., current 11 December). These represent a significance at a 98% level in 2009, 2010, 2013, and 2014 and a 95% level in 2011 and 2012. It can be seen that the observed low-frequency component (solid line) in 2009 (Figure 9a), 2013 (Figure 9e), and 2014 (Figure 9f) shows heavier rainfall, in which the 20 to 30 day oscillation is not actually a regular "oscillation", but rather an episodic mode of intraseasonal variability, and also the time between the active phases varied from event to event. Among these 6 years (2009 to 2014), the year 2012 was recorded as a drought year. A significant error was found in predicting the subdued 20 to 30 day rainfall ISO activity during June and July of 2012 (Figure 9d), which explains its lower overall prediction skill during this particular year when compared to the majority of the other years. On the other hand, despite being a drought year, the 20 to 30 day rainfall ISO activity during 2010 was persistently strong from January to June (Figure 9b), which still indicated a systematically higher forecasting skill (Figure 8a, blue line), and a skill level of 30 days. In addition, the major error in predicting the 20 to 30 day rainfall ISO for the year 2011 was in fact due to the model's failure in capturing the oscillation intensity during the months of June and September (Figure 9c), in which there were two abrupt changes of rainfall ISO, with weak cases in spring, strong cases throughout the boreal summer, and then weak cases again in fall. Furthermore, as seen from the associated PDF in Figure 9, it is also found that the PDFs of the low-frequency rainfall over the LYRV from the data and those from the high-order ECAR mode match well for these 6 years, especially in 2011 and 2012. It is indicated that the fifth-order ECAR model has a higher skill in predicting the extreme events, in which the information of the 20 to 30 day rainfall ISO over the LYRV are well reflected.

The above analysis suggests that the intensities of the rainfall ISO activity play important roles in determining the prediction skills in different years. For example, the year 2009 had overall strong and regular ISO activity during the whole year, which resulted in a long predictability (Figure 8a, black line), while the signal-to-noise ratios in the years 2011(Figure 8a, red line) and 2012 (Figure 8a, green line) were smaller than the other years, and thus, the predictability is greatly affected. In addition, Figure 10 shows the prediction for 1991, which displayed the strongest 20 to 30 day rainfall ISO (Figures 1b and 4b). It can be seen that for the 20 to 30 day rainfall component, with time lead increasing from 10 to 58 days, correlation skill drops from 0.98 to 0.5 (Figure 10a), while the

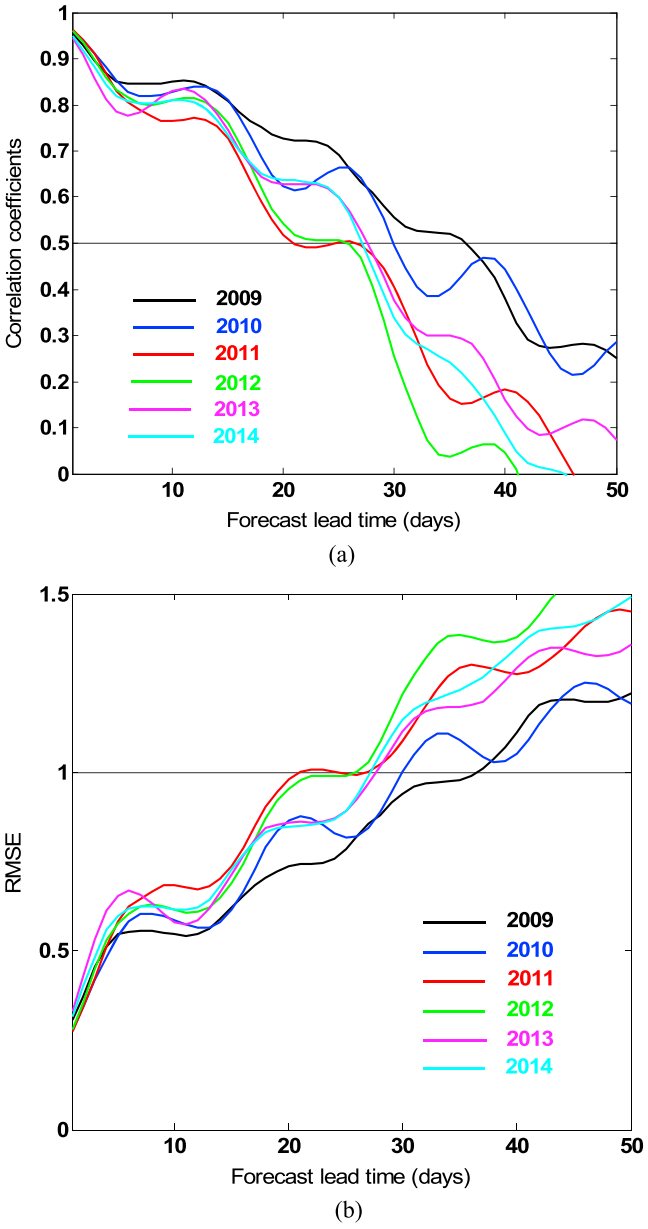


Figure 8. (a) The skill scores with pattern correlations for predictions utilizing the fifth-order ECAR in different years. The horizontal solid line shows the threshold of correlation coefficient $r = 0.5$. Useful predictions are defined by the pattern correlation between the predicted signal and the true signal above 0.5; (b) the RMSE; unit: standard deviation. The horizontal solid line shows the threshold of the RMSE = 1.0. The useful predictions are defined by the normalized root-mean-square error below 1.0, in which the daily low-frequency rainfalls of forecast and observation are normalized by the standard deviation of the data for each year.

RMSE increases from 0.23 to 1.00 (Figure 10b), in which the daily low-frequency rainfalls of forecast and observation are normalized by the standard deviation of the data in 1991. Therefore, the useful predictive skill can be extended to 58 days. Moreover, the forecast skill at a 20 day lead is 0.93 (Figure 10c) (365 forecasts, with the initial date of forecast set from the previous 12 December 1990, …, current 11 December 1991). As can be clearly seen in Figure 10c, the year 1991 had an overall strong and regular 20 to 30 day ISO activity associated with the heavy rainfall over the LYRV during the boreal summer, which resulted in a long predictability. On the other hand, the model statistics are highly consistent with those of the full 20 to 30 day ISO time series, including the autocorrelations of up to 90 days (Figure 10d), the PDFs (Figure 10e), and the power spectrums (Figure 10f). These results suggest that, even though 1991 is within the training period, this forecasting shows that the fifth-order ECAR model could effectively predict extreme events.

In previous studies, it has been confirmed that air-sea couplings play very important roles in ISO generation and maintenance and that external heating forces caused by the sea surface temperature anomaly (SSTA) are closely related to the changes in the atmospheric ISO intensity for a certain time scale (Wang, 2008). Figure 11 describes the spatial distribution of the teleconnection between the 20 to 30 day ISO intensity for the precipitation over the LYRV in summer (Figure 1b, blue line) and the data of the global SSTA in the preceding spring for the period of 1979–2014 (based on the global sea surface temperature data (HadSST3) acquired from the British Hadley Climate Prediction and Research Center; Kennedy et al., 2011). It is clearly indicated that the teleconnections between the intensity of the 20 to 30 day rainfall ISO and SSTA in the middle latitudes of the North Pacific in spring are quite significant. Meanwhile, there are significant correlations over the tropical western Indian Ocean and southern Indian Ocean, as well as the middle latitudes of the South Atlantic. These areas are key regions of the interactions between the SSTA and the 20 to 30 day rain ISO and heavy rainfall in the LYRV. However, there are insignificant correlations over the tropical center and eastern Pacific Ocean, in which the 20 to 30 day rain ISO over the LYRV is not related to the ENSO. Accordingly, the predictability of the 20 to 30 day ISO of the rainfall over the LYRV is high during the years when the springtime SSTA in the northwestern Pacific Ocean is large and positive, which is also favorable for the initially strong 20 to 30 day rainfall ISO to obtain higher prediction skills.

4.1.3. Importance of the Selected Extended Complex Matrix (ECM) to Skills

The abovementioned large amounts of the forecast experiments indicate that, based on the relationship between the 20 to 30 day ISO in East Asia and the global dominant ISO patterns, and aiming at the leading lagged relationship of a certain phase in the complex space, this study's high-order ECAR forecasting model could be established. It is found to effectively predict the changes in the low-frequency rainfall components of the LYRV during the 10 to 30 day extended-range, which is significantly extended to a lead time of about 28 days. It not only effectively reflects the relationship between low-frequency rainfall and the main circulation but also provides an important basis for the forecasting of the heavy rainfall processes of the LYRV for the following 10 to 30 days.

Therefore, based on the extended low-frequency complex sequence, an ECM using a Fourier transformation is established, and an ECAR model is constructed in a complex space (Yang, 2014). This is an entirely

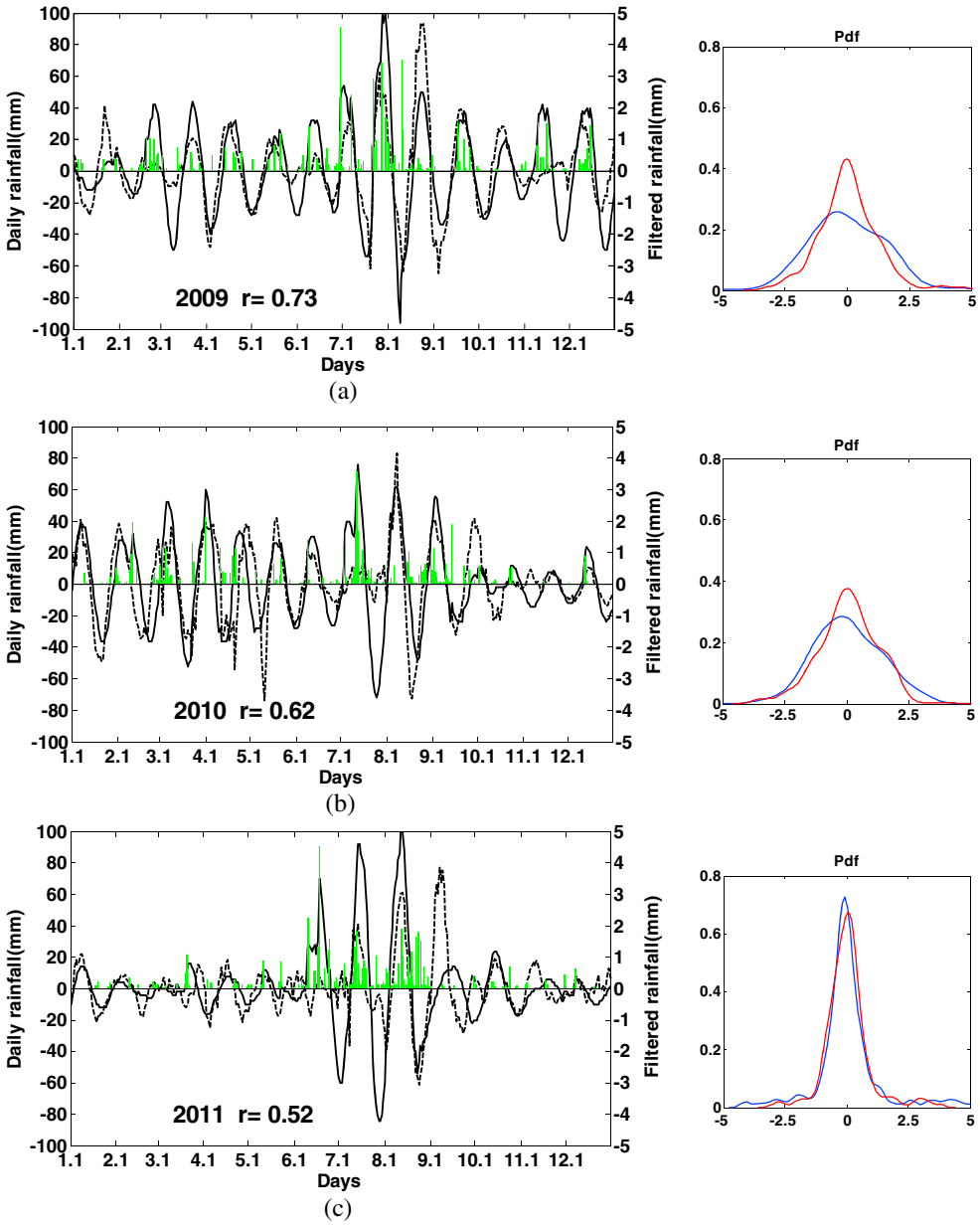


Figure 9. Forecasts (dashed line) and observations (solid line) of the fifth-order ECAR model at a lead time of 20 days from January to December: (a) 2009, (b) 2010, (c) 2011, (d) 2012, (e) 2013, and (f) 2014 for the 20 to 30 day low-frequency rainfall over the LYRV. The histogram represents the daily rainfall, unit: mm; *r* represents the correlation coefficient between the forecasts and the observations; the initial date of forecast is set as the previous December 12, …, current December 11. For the PDFs, the blue (red) line represents the observations (forecasts) of the low-frequency rainfall.

data-driven forecasting model, which requires no preprogrammed or predetermined rules. This method also provides a new description for determining the dynamical processes of the interactions between the internal components in climate systems. It is also conducive to further explaining the structure and properties of the climate system. Consequently, it is determined to provide a longer autocorrelation time in a complex space and a higher forecast skill. This method thereby improves the performances of the forecasting of the various dominant climate change modes and significantly prolongs the predictable lead times.

It is found noteworthy that the ECAR framework is flexible and could be easily adapted for extended-range forecasting over other extratropical regions where the forecasting skills of the current dynamical models

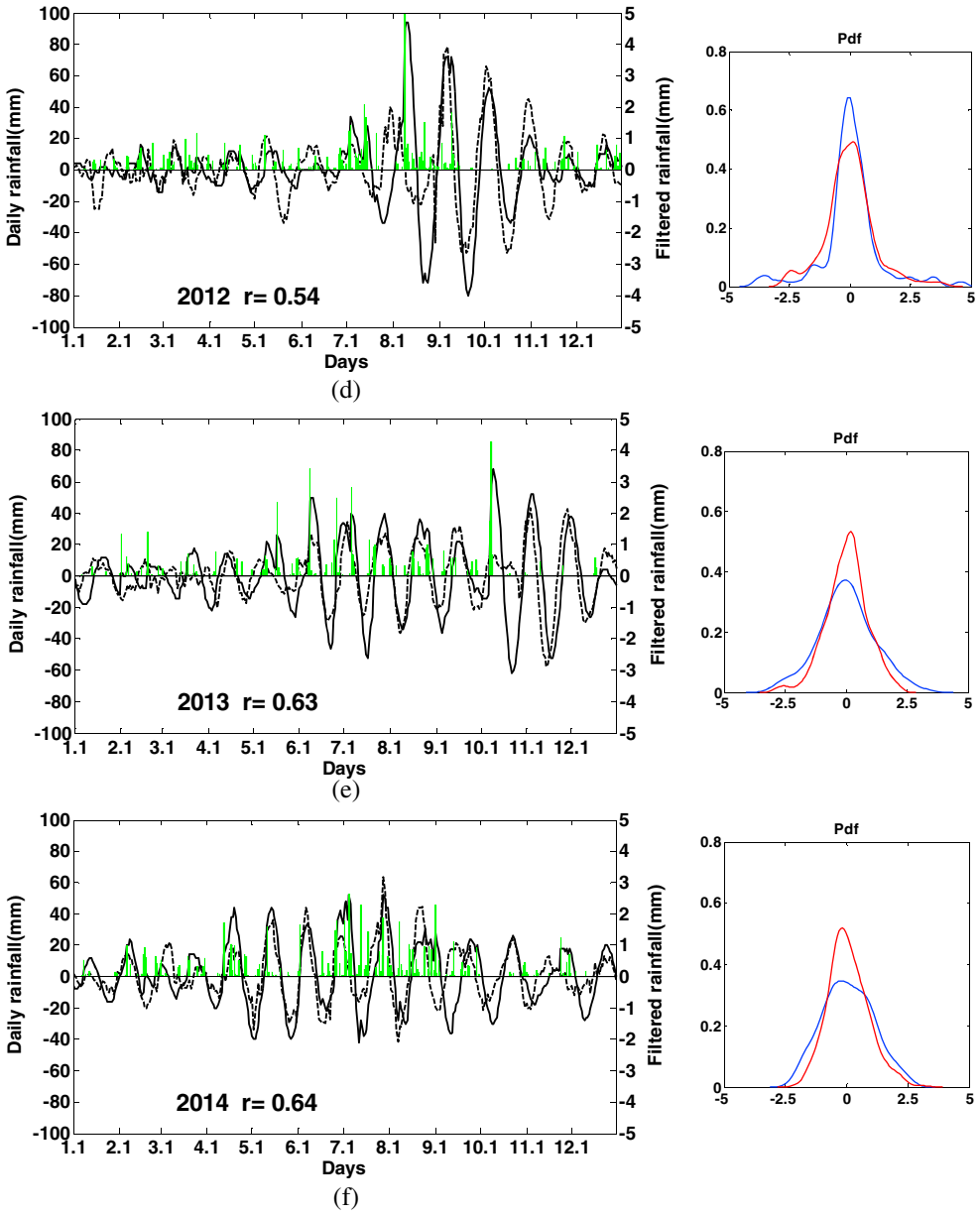


Figure 9. (continued)

have been found to be poor in predicting future precipitation events (for example, regions in Europe and South America and southern Africa).This forecast method helps avoid dynamical model error and is able to operate in real time. Also, the ECAR displays the ability to forecast different predictands (such as cold surges and heat waves), provided that the ECM could be appropriately selected. Furthermore, this method could be readily adopted to predict other timescale variabilities, such as interannual or interdecadal climate variations.

4.2. Importance of Initial Conditions and Boundary Effects of Filtering Method

In this study, based on the higher leading (lagged) correlations between the principal components of the East Asian and global circulations in a complex spaces (the role of the low-frequency component in the Southern Hemisphere is more prominent), improvements are made to the extended-range forecasting of the low-frequency rainfall components over the LYRV. In regard to the real-time ISO forecasting, the 10 to 15 day information of prior and posterior to the sequence is lost during the process of the ISO characteristic

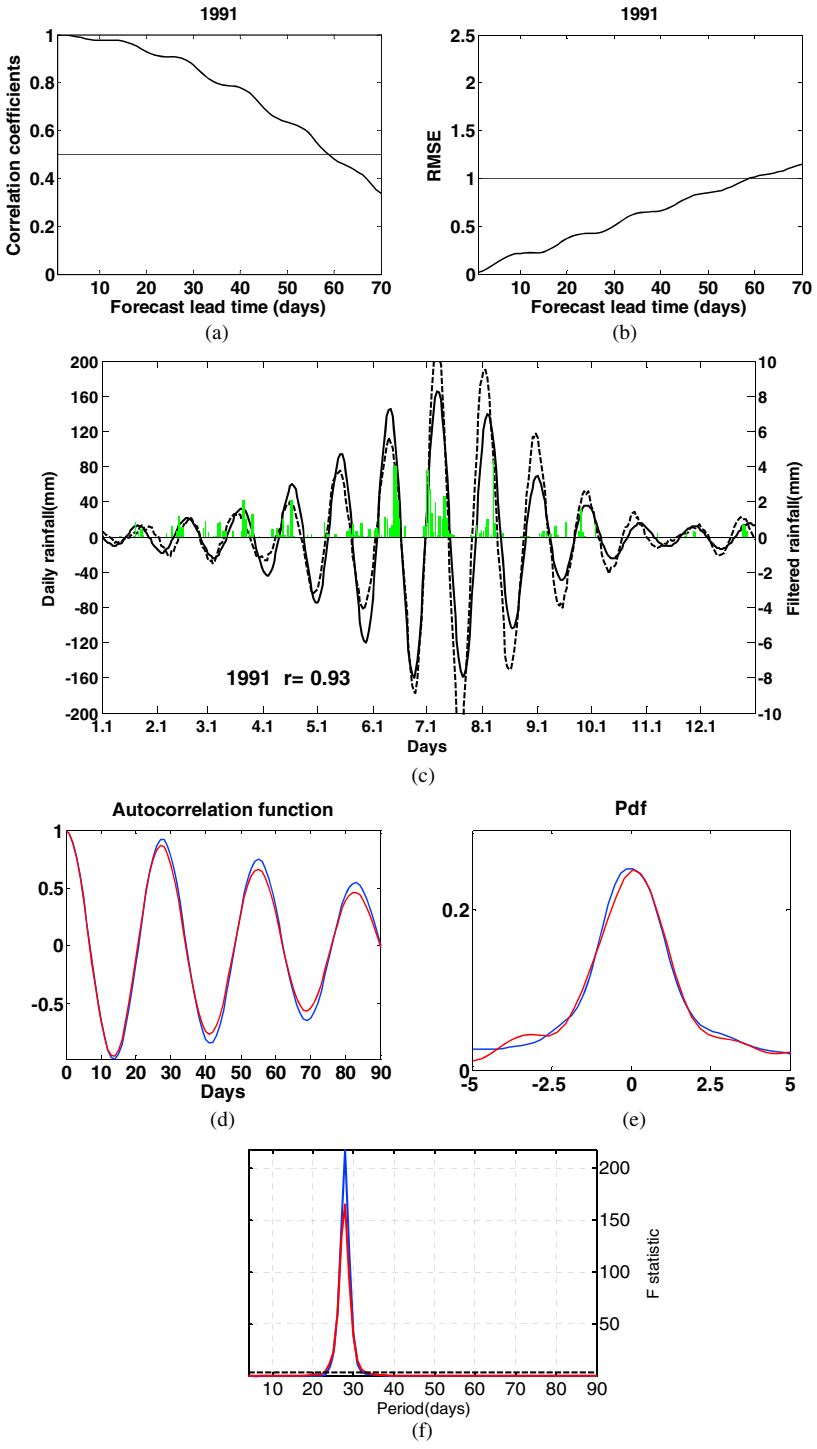


Figure 10. (a) Correlation skill of the 1 to 70 day forecasts of the low-frequency rainfall component r_{Icj} in the LYRV in 1991, on the time scale of 20 to 30 days, utilizing the fifth-order ECAR. The horizontal solid line indicates the threshold of the correlation coefficient $r = 0.5$; (b) root-mean-square error (RMSE, unit: standard deviation). The horizontal solid line in the figure represents the threshold of the RMSE = 1.0, in which the daily low-frequency rainfalls of forecast and observation are normalized by the standard deviation of the data in 1991; (c) forecasting (dashed line) of the fifth-order ECAR model at a lead time of 20 days and observations (solid line) from January to December. The histogram represents the daily rainfall, unit: mm; r is the correlation coefficient between the forecast and the observation. A comparison of the statistics of the real-time low-frequency rainfall and the model signals for (d) autocorrelation function and (e) PDFs and (f) noninteger power spectrum, in which the blue (red) line represents the observations (forecasts) of the low-frequency rainfall over the LYRV for the 20 to 30 day oscillation.

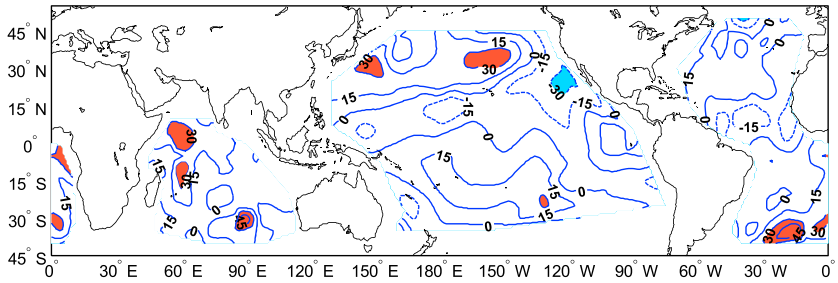


Figure 11. Teleconnections between the intensity of the low-frequency rainfall over the LYRV in summer and global sea surface temperature anomalies (SSTA) in the preceding spring for the period of 1979–2014. The values in the figure have been multiplied by 100, in which the confidence levels greater than 95% are shaded.

extractions using a traditional filtering method, which cause a significant influence on the forecasting precision of the lead time in the time series associated with the right-end effects of the filtered sequence. This uncertainty in initial conditions has become a well-known obstacle in the extended-range forecast.

However, in the forecast scheme proposed in this study, the SSA is used for low pass filtering on the original sequence, not only to filter the high-frequency noises but also to filter the periodic weak signals. The results indicate the accurate position of the main signal change. A T-EOF extension is adopted to minimize the end effects for applications of the real-time predictions for the right end of the filtered sequence, due to the end effects that bias the SSA reconstruction and estimated value of the low-frequency mode phase at the forecast starting time. In this T-EOF extension, for the forecasting starting time t_0 and a given l (l = 75 days), this study uses a T-EOF based on the prediction method (Lee, 2002) in order to forecast the value of $t_0 + 1$, $t_0 + 2$, …, $t_0 + l$, and also to extend the record of the end point of t_0 into the record of $t_0 + 1$, $t_0 + 2$, …, $t_0 + l$, on the daily rainfall over the LYRV. Then, this extended series, with the length of the series $t_0 + l$, is projected onto the T-EOF5 and T-EOF6 (Figure 4a; the T-EOF is calculated by the daily data during the period ranging from 1 January 1979 to 31 December 2008), and the extended RC5 + RC6 is obtained. The record of the t_0 for this extended RC5 + RC6 is a record of the real-time low-frequency rainfall (end point of the filtered sequence at the initial time t_0). Furthermore, for the forecasting start times t_0 during the period ranging from 1 January 2009 to 31 December 2014 (2,191 days), a real-time low-frequency component sequence of the observed rainfall over the LYRV (Figure 4c, blue line) could be obtained on a time scale of 20–30 days. It is found that the correlation coefficient of this real-time low-frequency rainfall with the T-EOF extension to that of the observations when the entire available data record is used for the period from ranging from 1 January 2009 to 31 December 2014 (Figure 4c, red line) is 0.758, which indicates a 99.9% significance level. However, the correlation of the real-time low-frequency rainfall without the T-EOF extension (Figure 4d, blue line) to that of the observations using the entire data record (Figure 4d, red line) is determined to be 0.610, which is significantly lower. Therefore, the information of the full 20 to 30 day rainfall ISO are confirmed to be better reflected by the real-time low-frequency rainfall (Figure 4c, blue line) with a T-EOF extension. It should be noted that no preprocessing, such as band-pass filtering, or seasonal partitioning, is performed. It is also important to note that the advantages of the SSA filter are that these modes are data adaptive, and therefore, they are not restrictive (like the Fourier filter), and unlike moving average filters, there are no losses of the end points. Also, more precise initial values (right end of the filtered sequence) are obtained, which minimizes the boundary effects of the traditional filtering method through the use of a T-EOF extension. Therefore, through above corrections inside SSA, it is found to be very suitable for real-time extended-range forecasting. This is also the main reason for the significant prolonging of the lead time of the ECAR model, which is established in this study.

Furthermore, the SSA also has the ability of identifying oscillation signals with irregular temporal variations, which are not single-frequency sinusoidal oscillations, and therefore had a certain forecasting ability for the nonlinear changes in the low-frequency components of the rainfall. Consequently, this ECAR model could potentially be used to predict the irregular enhancements of a 20 to 30 day low-frequency rainfall component, as well as the symbol changes (phase transformations) in the extended range. Therefore, this model could provide larger amounts of information for the forecasting of heavy rainfall events in summer over the LYRV for a future 30 days.

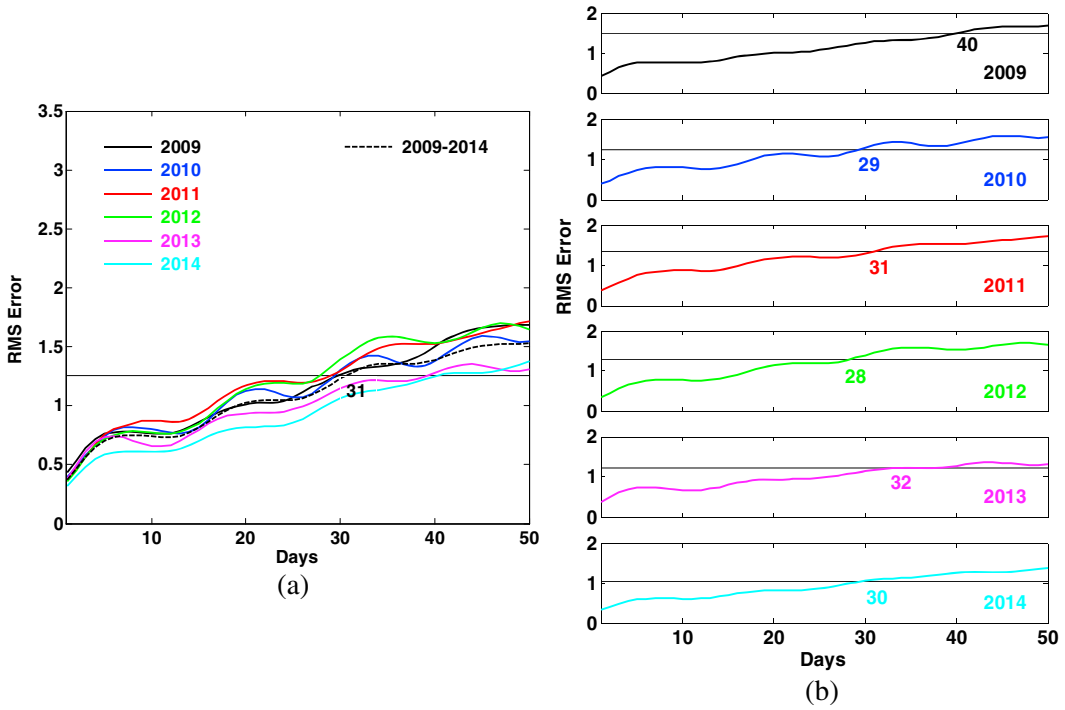


Figure 12. Mean in RMS error is estimated from the firth-order ECAR (a) over the years 2009–2014 and (b) for each year during the period of 2009–2014. The dashed curve in Figure 12a and thick curve in Figure 12b are the mean error growth during 2009–2014 and for each year in the firth-order ECAR model, respectively. The thin horizontal line indicates the standard deviation of the data over the years 2009–2014 in Figure 12a and for each year in Figure 12b, respectively. Units are mm. The predictability estimated by these ECAR models is presented in number.

4.3. Predictability

The forecast error is the difference between the forecasted value and the observed value and therefore includes the error due to the uncertainty in the initial condition as well as the error due to the model imperfection. This forecast error gives the lower bound of predictability of a model (Lorenz, 1982). The daily error growth of the low frequency rainfall over the LYRV is computed as root-mean-square (RMS) of the difference between forecast and observation, and the limit of the forecast skill is estimated. The standard deviation of the data for each year is computed, in which this standard deviation changes from year to year. We use this standard deviation as the threshold for determining the predictability. When the RMS error at a certain lead time is beyond this threshold, the predictability is reached.

The growth of errors is studied by computing the RMS of the difference in the daily anomalies of the low frequency rainfall between forecast and observation. The mean in the RMS error is calculated over the years 2009–2014 (Figure 12a). It can be seen that for the low-frequency rainfall over the LYRV, with time lead increasing from 1 to 31 days, mean RMS error (dashed line) increases from 0.5 to 1.3, in which the skillful prediction is about 31 days (using one standard deviation as the threshold during 2009–2014). Further, the 6 year independent forecast for each year, based on the standard deviation of the data for each year, shows significant skill at the lead time of 40, 29, 31, 28, 32, and 30 days during the period from 2009 to 2014 (Figure 12b), respectively. Note that the errors grow slower in the year 2009 with the stronger ISO and they grow faster in the year 2014 with the weak ISO, but in either case, the useful skills still are beyond 28 days. Hence, the analysis of the 20 to 30 day ISO predictability reveals a predictability limit of about 28–40 days, based on the 6 year (2009–2014) independent forecast.

5. Conclusions and Discussion

Global large-scale dynamical components are commonly used for the monitoring of various physical features, such as the varying components of global circulations on the time scales of 20 to 30 days, 30 to

50 days, 50 to 80 days, and so on. Among these oscillations, the intensity of the 20 to 30 day oscillation has a strongly positive correlation with the heavy rainfall processes over the LYRV in summer. In this study, based on the temporal lag-lead relationship of these components of global circulations to the low-frequency rainfall over the LYRV in a complex space, a statistical forecasting model of the time-varying high-order ECAR models with the 20 to 30 day oscillations is set up for the extended-range prediction of heavy rainfall process over the LYRV for 10–30 days in advance. The predictability of the low-frequency rainfall with a suite of ECAR models is presented for 20 to 30 day oscillations affecting the heavy rainfall processes over the LYRV, in which the global low-frequency wind field data, and the daily 20 to 30 day rainfall ISO over the LYRV, are used to build the forecasting models for the long sequences over a period from 1979 to 2008. A lot of forecast experiments are carried out for the period from 2009 to 2014. Furthermore, the variations of predictability for these 20 to 30 day oscillations are also analyzed with different intensities of ISO in different years.

Using the principal low-frequency modes of the global 850 hPa meridional winds, which correlated to the changes of the 20 to 30 day low-frequency wave train in the extratropics of the Southern Hemisphere, a series of time-varying high-order ECAR forecast models are established in a complex space that could effectively predict the changes in an extended-range for the 20 to 30 day low-frequency rainfall over the LYRV, for the period ranging from 2009 to 2014. The results provide significant forecasting signals in regard to when the extended-range heavy rainfall processes would occur over the LYRV. Based on the time-varying fifth-order ECAR forecast models, a superior prediction of the growth of the enhancement processes for the low-frequency components that correspond to the heavy rainfall process with a 25 to 30 day lead is achieved. The lead time is extended for the 20 to 30 day low-frequency rainfall over the LYRV to about 28 days for this 6 year period (2009 to 2014), and it is also found to be about a 27 day lead time for the forecasts that are initiated from weak ISO conditions. This suggests that the fifth-order ECAR is able to significantly improve the prediction of ISO and also displays better stability. Meanwhile, the autocorrelations, PDFs, and noninteger power spectrums of the forecasts for the low-frequency rainfall with the fifth-order ECAR model are all determined to be quite consistent with the observations. These results imply that nearly identical statistical and dynamical features have been described for the 20 to 30 day rainfall oscillations. Furthermore, the analysis of the 20 to 30 day ISO predictability reveals a predictability limit of about 28–40 days. However, the first-order AR model, which is directly established by the low-frequency component of the rainfall over LYRV, only has a lead time of less than 5 days. Therefore, it could be concluded that greater amounts of low-frequency change information of the generating complex sequence by a Fourier transform could be isolated in a complex space, compared with that of the original sequence in a real space. These constructed time-varying high-order ECAR models are found to be more stable when describing the lagged correlations between the low-frequency components and distinctly improved the real-time forecasting accuracy, along with extending the lead times. Therefore, the forecasting abilities are determined to be higher than those of the traditional first-order AR model. The forecasting framework that is used in this study can potentially assist in a better realization of the potential of the real-time forecasts for the 20 to 30 day oscillations related to the heavy rainfall processes over the LYRV. Also, the increased reliability, as well as improved extended-range rainfall forecasts, suggests that these time-varying high-order ECAR models could be useful operational forecasting tools. It is found that an ECAR modeling method based on an extended complex matrix of the low-frequency sequence specific to the 20 to 30 day ISO also provides a new description for determining the dynamical processes of the internal interaction in climate systems.

One of the main ways to extend the lead time in an extended-range is by using a climate forecasting method based on data-driven modeling. Then, by utilizing the high-dimensional data analyses, noise filtering, and system complexity reduction, the various low-frequency change processes of the global atmosphere, which are known to affect the heavy rainfall over the LYRV with the time scale of 20–30 days, can be reasonably identified and extracted from a large number of observations. The big data sets can be converted into small data sets, which can potentially reveal some of the change laws of a few of the principal low-frequency components within a low dimensional space. By using a Fourier transform, the low-frequency principal component in the real space can be transformed to the complex space. Based on the extended low-frequency complex sequence, an ECM is established. Then, the new laws of the complicated lagged correlation changes between the multiple low-frequency principal components of global circulation and the rainfall over the LYRV can be determined in the larger extended complex space. This can potentially more effectively and comprehensively describe the complex change information of the principal components of a climate

system in a low dimensional space. Therefore, a series of the time-varying high-order ECAR models, with a higher forecasting ability based on an ECM, can effectively prolong the predictable lead times in the extended range, which will reflect the interactions between the scientific big data and small data. Differing from the data-driven modeling strategies for the original sequence in real number domains, such as multilevel regression models (Kondrashov et al., 2013; Kravtsov et al., 2005), physics-constrained models (Chen et al., 2014; Majda & Harlim, 2013), and so on, the high-order ECAR method is developed for generating complex sequence by a data transformation in this study, including a Fourier transformation, an extended complex matrix, and the higher orders for the complex number autoregression. It may help provide the essential time evolution information that can be potentially useful in complex number domains. Furthermore, it is not necessary to input any predetermined rules in advance, and the method is found to be close to the forecasted model with no theoretical support. It is a systematical method and can be applied to any time series, provided that the ECM could be appropriately selected. Therefore, this method is based on the major lagged correlations with multiple different laggings hidden in a large number of the observational data in a complex number space with the time scales of 20–30 days, along with being completely driven by the data. The method is found to be able to combine together the extended-range theoretical forecasting and the practical applications in real time, which is an important way to significantly improve the accuracy of the extreme weather extended-range forecasting.

Acknowledgments
This study is supported by the National Natural Science Foundation of China (grant 41175082). The NCEP/NCAR reanalysis data were obtained from the Climate Diagnostics Center (http://www.cdc.noaa.gov). The author appreciates the valuable comments and helpful suggestions from three anonymous reviewers.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–722.

Alexander, R., Zhao, Z., Szekely, E., & Giannakis, D. (2017). Kernel analog forecasting of tropical intraseasonal oscillations. *Journal of the Atmospheric Sciences*, *74*, 1321–1342.

Baldwin, M. P., Stephenson, D. B., Thompson, D. W. J., Dunkerton, T. J., Charlton, A. J., & O'Neill, A. (2003). Stratospheric memory and skill of extended-range weather forecasts. *Science*, *301*, 636–640.

Barnett, T. P. (1983). Interaction of the monsoon and Pacific trade wind system at interannual time scales part I: The equatorial zone. *Monthly Weather Review*, *111*, 756–773.

Borah, N., Sahai, A. K., Chattopadhyay, R., Joseph, S., Abhilash, S., & Goswami, B. N. (2013). A self-organizing map-based ensemble forecast system for extended range prediction of active/break cycles of Indian summer monsoon. *Journal of Geophysical Research: Atmospheres*, *118*, 9022–9034. https://doi.org/10.1002/jgrd.50688

Brunet, G., Shapiro, M., Hoskins, B., Moncrieff, M., Dole, R., Kiladis, G. N., … Shukla, J. (2010). Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction. *Bulletin of the American Meteorological Society*, *91*, 1397–1406.

Bürger, G. (1993). Complex principal oscillation pattern analysis. *Journal of Climate*, *6*, 1972–1986.

Cassou, C. (2008). Intraseasonal interaction between the Madden-Julian Oscillation and the North Atlantic Oscillation. *Nature*, *455*, 523–527.

Cavanaugh, N. R., Teddy, A., Subramanian, A., Mapes, B., Seo, H., & Miller, A. (2014). The skill of atmospheric linear inverse models in hindcasting the Madden-Julian Oscillation. *Climate Dynamics*. https://doi.org/10.1007/s00382-014-2181-x

Chen, N., Majda, A. J., & Giannakis, D. (2014). Predicting the cloud patterns of the Madden-Julian Oscillation through a low-order nonlinear stochastic model. *Geophysical Research Letters*, *41*, 5612–5619. https://doi.org/10.1002/2014GL060876

Chou, J., Zheng, Z. H., & Sun, S. P. (2010). The think about 10–30 days extended range numerical weather prediction strategy facing the atmosphere chaos. *Sciences Meteorological Sinica (in Chinese)*, *30*, 569–573.

Feng, G. L., Sun, S. P., Zhao, J. H., & Zheng, Z. H. (2013). Analysis of stable components for extended-range (10–30 days) weather forecast: A case study of continuous overcast-rainy process in early 2009 over the mid-lower reaches of the Yangtze River. *Science China Earth Sciences*, *56*(9), 1576–1587. https://doi.org/10.1007/s11430-012-4527-8

Fu, X., Wang, W. Q., Lee, J.-Y., Wang, W., & Vitart, F. (2013). Intraseasonal forecasting of Asian summer monsoon in four operational and research models. *Journal of Climate*, *26*(12), 4186–4203. https://doi.org/10.1175/JCLI-D-12-00252.1

Hasselmann, K. (1988). PIPs and POPs: The reduction of complex dynamical systems using principal interaction and principal oscillation patterns. *Journal of Geophysical Research*, *93*, 11,015–11,021. https://doi.org/10.1029/JD093iD09p11015

Hoskins, B. (2013). The potential for skill across the range of the seamless weather-climate prediction problem: A stimulus for our science. *Quarterly Journal of the Royal Meteorological Society*, *139*, 573–584.

Hsu, P.-C., Li, T., You, L., Gao, J., & Ren, H. L. (2015). A spatial-temporal projection method for 10–30-day forecast of heavy rainfall in Southern China. *Climate Dynamics*, *44*(5-6), 1227–1244. https://doi.org/10.1007/s00382-014-2215-4

Hudson, D., Alves, O., Hendon, H. H., & Marshall, A. G. (2011). Bridging the gap between weather and seasonal forecasting: Intraseasonal forecasting for Australia. *Quarterly Journal of the Royal Meteorological Society*, *137*, 673–689.

Jia, X., Yang, S., Li, X., Liu, Y., Wang, H., Liu, X., & Weaver, S. (2013). Prediction of global patterns of dominant quasi-biweekly oscillation by the NCEP Climate Forecast System version 2. *Climate Dynamics*, *41*(5), 1635–1650.

Jiang, X., Waliser, D. E., Wheeler, M. C., Jones, C., Lee, M. I., & Schubert, S. D. (2008). Assessing the skill of an all-season statistical forecast model for the Madden-Julian Oscillation. *Monthly Weather Review*, *136*(6), 1940–1956. https://doi.org/10.1175/2007MWR2305.1

Jones, C., Carvalho, L. M. V., & Liebmann, B. (2012). Forecast skill of the South American monsoon system. *Journal of Climate*, *25*(6), 1883–1889. https://doi.org/10.1175/JCLI-D-11-00586.1

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., … Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077%3C0437:TNYRP%3E2.0.CO;2

Kang, I.-S., & Kim, H. M. (2010). Assessment of MJO predictability for boreal winter with various statistical and dynamical models. *Journal of Climate*, *23*, 2368–2378.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Saunby, M., & Parker, D. E. (2011). Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 1: Measurement and sampling errors. *Journal of Geophysical Research*, *116*, D14103. https://doi.org/10.1029/2010JD015218

Kikuchi, K., & Wang, B. (2009). Global perspective of the quasi-biweekly oscillation. *Journal of Climate*, *22*, 1340–1359.
Kim, H.-M., & Webster, P. J. (2010). Extended-range seasonal hurricane forecasts for the North Atlantic with a hybrid dynamical-statistical model. *Geophysical Research Letters*, *37*, L21705. https://doi.org/10.1029/2010GL044792
Kondrashov, D., Chekroun, M. D., Robertson, A. W., & Ghil, M. (2013). Low-order stochastic model and "past-noise forecasting" of the Madden-Julian Oscillation. *Geophysical Research Letters*, *40*, 5305–5310. https://doi.org/10.1002/grl.50991
Kravtsov, S., Kondrashov, D., & Ghil, M. (2005). Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability. *Journal of Climate*, *18*, 4404–4424.
Lee, J.-Y., Wang, B., Wheeler, M. C., Fu, X., Waliser, D. E., & Kang, I.-S. (2013). Real-time multivariate indices for the boreal summer intraseasonal oscillation over the Asian summer monsoon region. *Climate Dynamics*, *40*, 493–509.
Lee, Y. A. (2002). A T-EOF based prediction method. *Journal of Climate*, *15*, 226–234.
Lin, H., & Brunet, B. (2011). Impact of the North Atlantic Oscillation on the forecast skill of the Madden-Julian Oscillation. *Geophysical Research Letters*, *38*, L02802. https://doi.org/10.1029/2010GL046131
Lorenz, E. N. (1982). Atmospheric predictability experiments with a large numerical model. *Tellus*, *34*, 505–513.
Love, B. S., & Matthews, A. J. (2009). Real-time localized forecasting of the Madden-Julian Oscillation using neural network models. *Quarterly Journal of the Royal Meteorological Society*, *135*, 1471–1483.
Love, B. S., Matthews, A. J., & Janacek, G. J. (2008). Real-time extraction of the Madden-Julian Oscillation using empirical mode decomposition and statistical forecasting with a VARMA model. *Journal of Climate*, *21*, 5318–5335. https://doi.org/10.1175/2008JCLI1977.1
Madden, R., & Julian, P. (1971). Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *Journal of the Atmospheric Sciences*, *28*, 702–708.
Majda, A., & Harlim, J. (2013). Physics constrained nonlinear regression models for time series. *Nonlinearity*, *26*, 201–217.
Martin, G. M., Milton, S. F., Senior, C. A., Brooks, M. E., Ineson, S., Reichler, T., & Kim, J. (2010). Analysis and reduction of systematic errors through a seamless approach in modeling weather and climate. *Journal of Climate*, *23*, 5933–5957. https://doi.org/10.1175/2010JCLI3541.1
Miura, H. M., Satoh, M., Nasuno, T., Noda, A. T., & Oouch, K. (2007). A Madden-Julian Oscillation event realistically simulated by a global cloud-resolving model. *Science*, *318*, 1763–1765. https://doi.org/10.1126/science.1148443
Miyakoda, K., Gordon, T., Carerly, R., Stern, W., Sirutis, J., & Bourke, W. (1983). Simulation of a blocking event in January 1977. *Monthly Weather Review*, *111*(4), 846–869. https://doi.org/10.1175/1520-0493(1983)111%3C0846:SOABEI%3E2.0.CO;2
Mo, K. C. (2001). Adaptive filtering and prediction of intraseasonal oscillations. *Monthly Weather Review*, *129*, 802–817.
Neena, J. M., Lee, J. Y., Waliser, D. E., Wang, B., & Jiang, X. (2014). Predictability of the Madden-Julian Oscillation in the intraseasonal variability hindcast experiment (ISVHE). *Journal of Climate*, *27*(12), 4531–4543. https://doi.org/10.1175/JCLI-D-13-00624.1
Qian, W. H. (2012). Physical decomposition principle of regional-scale atmospheric transient anomaly. *Chinese Journal of Geophysics (in Chinese)*, *55*, 1439–1448.
Rashid, H. A., Hendon, H. H., Wheeler, M. C., & Alves, O. (2011). Prediction of the Madden-Julian Oscillation with the POAMA dynamical prediction system. *Climate Dynamics*, *36*, 649–661. https://doi.org/10.1007/s00382-010-0754-x
Sabeerali, C. T., Ajayamohan, R. S., Giannakis, D., & Majda, A. J. (2017). Extraction and prediction of indices for monsoon intraseasonal oscillations: An approach based on nonlinear Laplacian spectral analysis. *Climate Dynamics*, *49*, 3031–3050. https://doi.org/10.1007/s00382-016-3491-y
Schickedanz, P. T., & Bowen, E. G. (1977). The computation of climatological power spectrum. *Journal of Applied Meteorology*, *16*, 359–367.
Seo, K.-H., Wang, W., Gottschalck, J., Zhang, Q., Schemm, J.-K. E., Higgins, W. R., & Kumar, A. (2009). Evaluation of MJO forecast skill from several statistical and dynamical forecast models. *Journal of Climate*, *22*, 2372–2388.
Thompson, D. W. J., & Barnes, E. A. (2014). Periodic variability in the large-scale Southern Hemisphere atmospheric circulation. *Science*, *343*, 641–645.
Thompson, D. W. J., & Woodworth, J. D. (2014). Barotropic and baroclinic annular variability in the Southern Hemisphere. *Journal of the Atmospheric Sciences*, *71*, 1480–1493.
Vautard, R., & Ghil, M. (1989). Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*, *35*, 395–424.
Vitart, F., & Molteni, F. (2010). Simulation of the Madden-Julian Oscillation and its teleconnections in the ECMWF forecast system. *Quarterly Journal of the Royal Meteorological Society*, *136*, 842–855.
von Storch, H., & Xu, J. S. (1990). Principal oscillation pattern analysis of the 30- to 60-day oscillation in the tropical troposphere. Part I: Definition of an index and its prediction. *Climate Dynamics*, *4*, 175–190.
Waliser, D. E. (2012). Predictability and forecasting. In K. M. Lau, & D. E. Waliser (Eds.), *Intraseasonal Variability of the Atmosphere-Ocean Climate System* (2nd ed., pp. 433–476). Berlin: Springer.
Waliser, D. E., Jones, C., Schemm, J. K., & Graham, N. E. (1999). A statistical extended range tropical forecast model based on the slow evolution of the Madden-Julian Oscillation. *Journal of Climate*, *12*(7), 1918–1939. https://doi.org/10.1175/1520-0442(1999)012%3C1918:ASERTF%3E2.0.CO;2
Waliser, D. E., Lau, K. M., Stern, W., & Jones, C. (2003). Potential predictability of the Madden-Julian Oscillation. *Bulletin of the American Meteorological Society*, *84*(1), 33–50. https://doi.org/10.1175/BAMS-84-1-33
Wang, B. (2008). Thrusts and prospects on understanding and predicting Asian monsoon climate. *Acta Meteorologica Sinica*, *22*, 383–403.
Wang, B., & Xie, X. (1997). A model for the boreal summer intraseasonal oscillation. *Journal of the Atmospheric Sciences*, *54*, 72–86.
Wheeler, M. C., & Hendon, H. H. (2004). An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly Weather Review*, *132*, 1917–1932.
Xavier, P. K., & Goswami, B. N. (2007). An analog method for real-time forecasting of summer monsoon subseasonal variability. *Monthly Weather Review*, *135*, 4149–4160.
Yang, Q. M. (1998). Forecast experiments for low frequency oscillation of tropical circulation in west Pacific using principal oscillation pattern analysis. *Journal of Applied Meteorology and Science (in Chinese)*, *9*(3), 345–351.
Yang, Q. M. (2009). The 20–30-day oscillation of the global circulation and heavy precipitation over the lower reaches of the Yangtze River valley. *Science in China Series D: Earth Sciences*, *52*, 1485–1501.
Yang, Q. M. (2011). Extended range forecast experiments of the low frequency circulation for the period of the heavy precipitation over the lower reaches of the Yangtze River Valley in mid-July 2011. *Science and Technology Review (in Chinese)*, *29*, 61–66.
Yang, Q. M. (2014). Extended complex autoregressive model of low frequency rainfalls over the lower reaches of Yangtze river valley for extended range forecast in 2013. *Acta Physica Sinica (in Chinese)*, *63*, 199202. https://doi.org/10.7498/aps.63.199202

Yang, Q. M. (2015). Prospects and progresses in the research of the methods for 10–30 days extended-range weather forecast. *Advanced Earth Science (in Chinese)*, *30*(9), 970–984.

Yang, Q. M., Li, Y., Song, J., & Huang, S. (2012). Extended-range forecasts of the principal 20–30-day oscillation of the circulation over East Asia during the summer of 2002. *Acta Meteorologica Sinica*, *26*(5), 554–565. https://doi.org/10.1007/s13351-012-0502-8

Yang, Q. M., Song, J., Li, Y., Zhiqing, X., Shicheng, H., & Wei, Q. (2012). Review of impacts of the global atmospheric intraseasonal oscillation on the continuous heavy rainfall over the Yangtze River valley. *Advances in Earth Science (in Chinese)*, *27*, 876–884.

Zhang, C., Gottschalck, J., Maloney, E. D., Moncrieff, M. W., Vitart, F., Waliser, D. E., … Wheeler, M. C. (2013). Cracking the MJO nut. *Geophysical Research Letters*, *40*, 1223–1230. https://doi.org/10.1002/grl.50244

Zheng, Z. H., Huang, J. P., Feng, G. L., & Chou, J. (2013). Forecast scheme and strategy for extended-range predictable components. *Science China Earth Sciences*, *56*, 878–889.

Zhu, Z., & Li, T. (2017). The statistical extended-range (10–30-day) forecast of summer rainfall anomalies over the entire China. *Climate Dynamics*, *48*, 209–224.

Zhu, Z. W., Li, T., Hsu, P.-C., & He, J. H. (2015). A spatial-temporal projection model for extended-range forecast in the tropics. *Climate Dynamics*, *45*, 1085–1098.