Forecasting experiments of a dynamical-statistical model of the sea surface temperature anomaly field based on the improved self-memorization principle

Mei Hong[1,2], Ren Zhang[1,2], ~~Xi Chen[1], Ren Zhang[1,2],~~ Dong Wang[3], Shuanghe Shen[2], Xi Chen[1], and Vijay P. Singh[4]

[1]*Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disaster, Nanjing University of Information Science &Technology, Nanjing 210044, China*

[1]~~Institute~~ [2]Institute of Meteorology and Oceanography, National University of Defense Technology, Nanjing 211101, China

[2]~~*Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disaster, Nanjing University of Information Science &Technology, Nanjing 210044, China*~~

[3]*Key Laboratory of Surficial Geochemistry, Ministry of Education; Department of Hydrosciences, School of Earth Sciences and Engineering, Collaborative Innovation Center of South China Sea Studies, State Key Laboratory of Pollution Control and Resource Reuse, Nanjing University, Nanjing 210093, China*

[4]*Department of Biological and Agricultural Engineering, Zachry Department of Civil Engineering, Texas A & M University, College Station, TX 77843, USA*

Corresponding authors address:

1. Ren Zhang, Research Centre of Ocean Environment Numerical Simulation, Institute of Meteorology and Oceanography, National University of Defense Technology, Nanjing 211101, China

E-mail: 254247175@qq.com

~~1.~~2. Xi Chen, Research Centre of Ocean Environment Numerical Simulation, Institute of Meteorology and Oceanography, National University of Defense Technology, Nanjing 211101, China

E-mail: chenxigfkd@163.com

2. Ren Zhang, Research Centre of Ocean Environment Numerical Simulation, Institute of Meteorology and Oceanography, National University of Defense Technology, Nanjing 211101, China

E-mail: 254247175@qq.com

Abstract: With the objective of tackling the problem of inaccurate long-term El Niño Southern Oscillation (ENSO) forecasts, this paper develops a new dynamical-statistical forecast model of sea surface temperature anomaly (SSTA) field. To avoid single initial prediction values, a self-memorization principle is introduced to improve the dynamic reconstruction model, thus making the model more appropriate for describing such chaotic systems as ENSO events. The improved dynamical-statistical model of the SSTA field is used to predict SSTA in the equatorial eastern Pacific and during El Niño and La Niña events. The long-term step-by-step forecast results and cross-validated retroactive hindcast results of time series T_1 and T_2 are found to be satisfactory, with a pearson correlation coefficient of approximately 0.80 and a mean absolute percentage error (MAPE) of less than 15%. The corresponding forecast SSTA field is accurate in that not only is the forecast shape similar to the actual field, but the contour lines are essentially the same. This model can also be used to forecast the ENSO index. The temporal correlation coefficient is 0.8062, and the MAPE value of 19.55% is small. The difference between forecast results in summer spring and those in winter autumn is not high,

indicating that the improved model can overcome the spring predictability barrier to some extent. Compared with six mature models published previously, the present model has an advantage in prediction precision and length, and is a novel exploration of the ENSO forecast method.

Keywords: Dynamical-statistical forecast model; self-memorization principle; sea surface temperature field; long-term forecast of ENSO

1. Introduction

The El Niño Southern Oscillation (ENSO), the well-known coupled atmosphere−ocean phenomenon, was firstly proposed by Bjerknes (1969). The ENSO phenomenon can influences regional and global climates, so the prediction of ENSO has received considerable public interest (Rasmusson and Carpenter, 1982; Glantz et al., 1991).

Over the past two to three decades, one might reasonably expect the ability to predict warm and cold episodes of ENSO at short and intermediate lead times to have gradually improved (Barnston et al., 2012). Many countries have been focusing on ENSO forecasts since the 1990s, and the ENSO forecast has become one of the important research topics in the International Climate Change and Predictability Research plan. The U.S. International Research Institute for Climate and Society, the U.S. Climate Prediction Centre, Japan Meteorological Agency, and European Centre for Medium-Range Weather Forecasting have developed different coupled atmosphere–ocean models to forecast ENSO (Saha et al., 2006; Molteni et al., 2007) .

71   The forecast models can generally be divided into two types (Palmer et al., 2004).

72   The first type is typified by a dynamic model, which mathematically expresses

73   physical laws that govern how the ocean and the atmosphere interact. The second type

74   is typified by a statistical model, which requires large a amount of historical data and

75   analyses the data to do forecasting (Chen et al., 1995; Moore et al., 2006).

76   Over the past three decades, ENSO predictions have made remarkable progress,

77   reaching a stage where reasonable statistical and numerical forecasts (Jin et al.,

78   2008)can be made 6–12 months in advance (Wang et al., 2009a). . However, there are

79   three problems remaining to be resolved (Zhang et al., 2003aa): (1) The current

80   ENSO predictions are mainly limited to the short term, such as annual and seasonal

81   predictions; (2) Although the representation of ENSO in coupled models has

82   advanced considerably during the last decade, several aspects of the simulated

83   climatology and ENSO are not well reproduced by the current generation of coupled

84   models. The systematic errors in SST are often very large in the equatorial Pacific,

85   and model representations of ENSO variability are often weak and/or incorrectly

86   located (Neelinet al. 1992; Mechoso et al. 1995; Delecluse et al. 1998; Davey et al.

87   2002). (3) Coupled models of ENSO predictions initialized from observed initial

88   states tend to adjust towards their own climatological mean and variability, leading to

89   forecast errors. The errors associated with such adjustments tend to be more

90   pronounced during boreal spring, which is often called the ''spring predictability

91   barrier'' (Webster et al., 1999). More efficient models are therefore desired (Belkin

92   and Niyogi, 2003; Weinberger and Saul, 2006). Therefore, the idea of combining

93   dynamical and statistical methods to improve weather and climate prediction has been

94   developed in many studies (~~Chou, 1974;~~ Huang et al., 1993;Yu et al., 2014a; Yu et

95   al.,2014b). By introducing genetic algorithms (GAs), Zhang et al. (2006) inverted and

96   reconstructed a new dynamical-statistical forecast model of the tropical Pacific sea

97   surface temperature (SST) field using historic statistical data (Zhang et al., 2008).

98   However, there is one flaw in the forecast model: the time-delayed SST field. This is

99   because ENSO is a complicated system with many influencing factors. To overcome

100 information insufficiency in the forecast model, Hong et al. (2014) selected the

101 tropical Pacific SST, SSW and SLP fields as three modelling factors and utilized the

102 GA to optimize model parameters.

103   However, the above dynamical prediction equations which were ,proposed by

104 Hong et al.(2014), greatly depend on a single initial value, creating long-term

105 forecasts over 8 months that diverged significantly. These unsatisfactory results

106 indicate that this model needs to be improved. Cao (1993) first proposed the

107 self-memorization principle, which transforms the dynamical equations with the

108 self-memorization equations, wherein the observation data can determine the memory

109 coefficients. This method has been widely used in forecast problems in environmental,

110 hydrological and meteorological fields (Feng et al., 2001; Gu, 1998; Chen et al.,

111 2009). The method can avoid the question of initial conditions for the differential

112 equations, so it can be introduced here to improve the proposed dynamical forecast

113 model.

114   Therefore, an improved dynamical-statistical forecast model of the SST field

115 and its impact factors with a self-memorization function was developed. The

116 improved model can absorb the information from past observations.

117 This paper is organized as follows: Research data and forecast factors are

118 introduced in section 2. In Section 3 the reconstruction of the dynamical model of

119 SSTA field is described. To improve the reconstruction model, the self-memorization

120 principle is introduced in Section 4. Model forecast experiments are described in

121 Section 5, and conclusions are given in Section 6.

2. Research data and forecast factors

2.1 Data

124 The monthly average SST data ~~from January 1951 to January 2010, 720~~

125 ~~months in total,~~ were obtained from the UK Met Office Hadley Centre for the region

126 (30°S-30°N；120°E -90°W). The gridded $1°\times1°$ Met Office Hadley Sea Ice and

127 SST dataset (HadISST1; Rayner et al. 2003) includes both in situ and available

128 satellite data. The sea areas provide important information on ocean-atmosphere

129 coupling in the East and West Pacific Ocean and the El Niño /La Niña events. The

130 reanalysis data, zonal winds and sea level pressures were obtained from the National

131 Center for Environmental Forecast of America and the National Center for

132 Atmospheric Research (Kalnay et al., 1996). The sea surface height (SSH) field was

133 obtained from Simple Ocean Data Assimilation (SODA) data (James and Benjamin,

134 2008). Outgoing longwave radiation (OLR) was obtained from the National Oceanic

135 and Atmospheric Administration (NOAA) satellites, at a resolution of $0.5°\times$

136 0.5°(Liebmann and Smith, 1996).~~The sea areas provide important information on~~

6

137 ~~ocean-atmosphere coupling in the East and West Pacific Ocean and the El Niño and~~

138 ~~La Niña events. The reanalysis data and zonal winds were obtained from the National~~

139 ~~Center for Environmental Forecast (NECP) of America and the National Center for~~

140 ~~Atmospheric Research (NCAR) (Kalnay et al., 1996).~~ The Southern Oscillation Index

141 (SOI) data were obtained from the Climate Prediction Center (CPC). The time series

142 of all data were from Jan. 1951 to ~~Jan~~Dec. 2010, 720 months in total.

2.2 EOF deconstruction

144 The sea surface temperature anomaly (SSTA) field can be calculated from the

145 SST field and can be deconstructed into time (coefficients)-space (structure) using the

146 empirical orthogonal function (EOF) method. Detailed information on the EOF

147 method can be seen in the related references (Dommenget & Latif, 2002). We have

148 used covariance matrix，because the covariance matrix was selected to diagnose the

149 primary patterns of co-variability in the basin-wide SSTs, rather than the patterns of

150 normalized covariance (or correlation matrix).

151

152 We used the smooths function with MATLAB to smooth the SSTA field before

153 the EOF deconstruction, which is five points two times moving, mainly filtering out

154 some noise points and outliers. Then a~~An~~ ~~empirical orthogonal function (EOF)~~

155 analysis of smoothed anomalies was performed, and the first two SSTA EOFs are

156 shown in Figs. 1a and 1c. The principal component (PC) time series corresponding to

157 the first and second EOFs are shown in Figs. 1b and 1d. The first EOF pattern, which

158 accounted for 61.33% of the total SSTA variance, represented the mature ENSO phase

(El Niño or La Niña), and the corresponding PC time series was highly correlated (with a correlation coefficient of 0.85) with the cold tongue index (SST anomaly averaged over 4°S–4°N, 180°–90°W) over the whole period. The second EOF, accounting for 14.52% of the total SSTA variance, indicated the ENSO signal beginning to enhance ~~the ENSO signal beginning to decay~~. Compared with the first mode, these were slightly attenuated in terms of the scope and intensity. The above analysis is similar to the EOF analysis of the SSTA field in the previous studies (Johnson et al., 2000;Timmermann et al., 2001). This indicates that the front two variance contribution modes can describe the main characteristics of the SSTA field and El Niño/La Niña. Therefore, we can choose the T_1, T_2 time series EOF decomposition modes as the modelling objects.

2.3 Selection of other prediction model factors

Considering the complexity of computation, the amount of variables in the equations of our model can't be too large, usually 3 or 4 for the best. This has been explained in our previous studies (Zhang et al., 2006; Zhang et al., 2008). If there are more than 4 variables in the modeling equation, it will cause the amount of parameters such as $a_1, a_2, ...a_n, b_1, b_2, ...b_n, ...$ too large. The huge computation makes it difficult to be precisely modeled. Thus, the total number of parameters in the model of five variables was 102, which may cause an overfitting problem. Hence, when we selected the model of five or six variables which entailed large amounts of computation that made precision difficult, and too many parameters might cause an overfitting phenomenon. If we choose only two or even fewer variables, the forecast

performance is poor too. Too few variables cause too small reconstructed parameters, resulting in amounts of important information missing out in the model. Thus, four variables are best for dynamically and accurately modeling. Because we have chosen two time series in section2.2 as the modeling objects, now we should select the other two ENSO intensity impact factors.

The ENSO intensity impact factor is an important issue in ENSO prediction. Previous studies have been completed in this area, which found that teleconnection patterns, temperature, precipitation, wind and SSH may affect ENSO strength. For example, Trenberth et al. (1998) noted that PNA, SOI and OLR in the Pacific Intertropical Convergence Zone (ITCZ) are all closely related to ENSO. Webster(1999) pointed out after the 1970, Indian Ocean dipole (IOD) is not only affected by ENSO, but also affected the strength of ENSO (Ashok et al., 2001). Yoon and Yeh (2010) reported that the Pacific Decadal Oscillation (PDO) disrupts the linkage between El Ni~no and the following Northeast Asian summer monsoon (NEASM) through inducing the Eurasian pattern in the mid-high latitudes. The vast majority of studies (Tomita and Yasunari, 1996; Zhou and Wu, 2010; Kim et al., 2017) have concentrated on the impacts of ENSO on the East Asian winter monsoon(EAWM). During the EAWM season, ENSO generally reaches its mature phase and has the most prominent impact on the climate. Wang et al. (1999a) and Wang et al. (1999b) suggested that the zonal wind factors in the eastern and western equatorial Pacific play a critical role in the phase of transition of the ENSO cycle, which could excite eastward propagating Kelvin waves and affect the SSTA in the

equatorial Pacific. Zhao et al. (2012) analyzed the characteristics of the tropical Pacific SSH field and its impact on ENSO events.

Based on the above analysis, we have selected nine factors, which may be closely related with the ENSO index (Niño3.4).

(1)The zonal wind in the eastern equatorial Pacific factor (u1) was calculated as the grid-point average of zonal wind in the area [5 °S ~ 5 °N, 150 °W ~ 90 °W].

(2) The zonal wind in the western equatorial Pacific factor (u2) was calculated as the grid-point average of zonal wind in the area [0 °~ 10 °N; 135 °E ~ 180 °E].

(3) The PNA teleconnection factor was obtained from the CPC.

(4) the dipole mode index factor (DMI) was obtained from SSTA for June-July-August (JJA) based on Saji(1999) method.

(5) The SOI factor was obtained from the CPC.

(6) The PDOI factor was obtained from department of Atmospheric Sciences in the university of Washington. The web is http://tao.atmos.washinton.edu/pdo/RDO.latest.

(7) The EAWM index (EAWMI) factor was proposed by Yang et al. (2002), which is defined by the meridional 850-hPa winds averaged over the region (20 ° ~40 °N, 100 °~140 °E).

(8) The OLR in the ITCZ factor was calculated as the grid-point average of OLR in the area [10 °N～20 °N，120 °E～150 °E].

(9) The SSH factor was calculated as the grid-point average of the SSH data in the area [10 °S ~ 10 °N; 120 °E ~ 60 °W].

A correlation analysis of the above factors was carried out and the results are shown in Table 1.

Table 1 shows that SOI and EAWMI have the stronger correlation with the front two time series T_1, T_2 than the other 7 factors. The results are also consistent with previous research (Clarke and Van Gorder, 2003; Drosdowsky, 2006; Zhang et al., 1996; Wang et al., 2008; Yang and Lu, 2014). Therefore, the first time series T_1, the second time series T_2, SOI and EAWMI will be selected as prediction model factors.

The ENSO intensity impact factor is an important issue in the ENSO prediction. Previous studies have found that teleconnection patterns, temperature, precipitation, wind and SSH may affect the ENSO strength (Trenberth et al.,1998; Webster,1999; Ashok et al., 2001; Yoon and Yeh, 2010; Tomita and Yasunari, 1996). For example, Trenberth et al. (1998) noted that the Pacific North American Oscillation Index (PNA) and SOI in the Pacific Intertropical Convergence Zone (ITCZ) were all closely related to ENSO. Liao et al. (2007) also noted that the decadal variation during ENSO events had a close relationship with the SOI index.The vast majority of studies (Tomita and Yasunari, 1996; Zhou and Wu, 2010) have concentrated on the impacts of ENSO on the East Asian winter monsoon (EAWM). During the EAWM season, ENSO generally reaches its mature phase and has the most prominent impact on the climate. Wang et al. (1999a) and Wang et al. (1999b) suggested that the zonal wind factors in the eastern and western equatorial Pacific played a critical role in the transition phase of the ENSO cycle, which could excite eastward propagating Kelvin waves and affect the SSTA in the equatorial Pacific.

247   Based on the above analysis, we selected four factors, which may be closely

248 related with the ENSO index (Niño 3.4) and were obtained as follows:

249   (1) The zonal wind in the eastern equatorial Pacific factor (u1) was calculated

250 as the grid-point average of zonal wind in the area [5 °S ~ 5 °N, 150 °W ~ 90 °W].

251   (2) The PNA teleconnection factor was obtained from the CPC.

252   (3) The SOI factor was obtained from the CPC.

253   (4) The EAWM index (EAWMI) factor was proposed by Yang et al. (2002),

254 which is defined by the meridional 850-hPa winds averaged over the region (20 °

255 ~40 N, 100 °~140 E).

256   All the four data selected ranged from January 1951 to January 2010.

257   Actually, how many variables and which variables are used in our model

258 become a key issue to be resolved. We can introduce a stepwise regression principle

259 to choose more reasonable predictors (Yim et al., 2015), because the stepwise

260 procedure can help selecting statistically important predictors at each step. The

261 significance of each predictor selected was based on its significance in increasing the

262 regressed variance by the standard F test (Panofsky and Brier, 1968). A 95 %

263 statistical significance level was used as a criterion to select a new predictor at each

264 step. Once selected into the model, a predictor can only be removed if its significance

265 level falls below 95 % by the addition/removal of another variable. For example, for

266 the model of only one variable, because we forecast the ENSO index, we should

267 choose T_1 or T_2 as the variable. Considering that T_1 accounts for 61.33% of the total

268 SSTA variance, so we chose T_1 as the variable. For the model of two variables, there

are five factors (T_2, u_1, PNA, SOI and EAWMI) which can be chosen for the second variable. Taking advantage of the stepwise regression ideas and selecting statistically important predictors by a standard F test, we can find the largest F test value among the five factors. That is T_2. Continuing this step, we can also select the reasonable factors for the model of three variables. Based on this thought, when the number of variables is determined, we can choose the most statistically important variables to reconstruct the prediction model. The forecast results of these models can be seen in table 1.

From table 1, the forecast results of all six models are satisfactory, where the temporal correlations of the models are all greater than 0.60 and the root mean square errors are all less than 0.81. Among all six models, the forecast results of four variables are the best for the following reasons:

(1) In general, the amount of parameters is less than 10% of the sample size, which can avoid over-fitting (Tetko et al., 1995) .The number of parameters $a_1, a_2, ... a_{14}, b_1, b_2, ... b_{14}, c_1, c_2, ... c_{14}, d_1, d_2, ... d_{14}$ of the model of four variables $T_1, T_2, SOI, EAWMI$ is 56, but we deleted the parameters which contributed little to the prediction. That means that there are 56 parameters in equation (1) in section 3, but there are only 34 parameters in equation (3) in section 3which is our final prediction equation. In section 5.1, because p is identified as 6, the number of parameters of the self-memorization function β_i is 28. Therefore, the total number of parameters in the model of four variables is 62, which is less than 10% of the sample size (720 months). The number of parameters $a_1, a_2, ... a_{20}, b_1, b_2, ... b_{20}, c_1, c_2, ... c_{20}, d_1, d_2, ... d_{20}, e_1, e_2, ... e_{20}$ of the model

13

of five variables $T_1, T_2, SOI, EAWMI, u_1$ is 100. Although the parameters which contributed a little were deleted, the number was still 72, and the number of self-memorization parameters was 30 (p determined as 5). Thus, the total number of parameters in the model of five variables was 102, which was more than 10% of the sample size (720 months).This will cause an overfitting problem. Hence, when we selected the model of five or six variables which entailed large amounts of computation that made precision difficult, and too many parameters caused an overfitting phenomenon. That is why the forecast results of five or six variables were worse than those of four variables.

(2) The models of one, two and three variables can avoid the overfitting problem, but too few variables will result in too few reconstruction parameters, causing important information missing from the model. Especially, when the model of one or two variables was considered, we only studied the self-memorization of the ENSO system but did not consider the mutual-memorization between factors. Thus, equations of our model only contained a self-memory term, not an exogenous effect term. That is why the forecast results of one, two and three variables were worse than those of four variables.

Based on the above analysis, we finally chose T_1, T_2, SOI and EAWMI as predictors for the model.

3. Reconstruction of dynamical model based on GA

Takens' delay embedding theorem (Takens, 1981) provides the conditions under which a smooth attractor can be constructed from observations made with a generic

313 function. Later results replaced the smooth attractor with a set of arbitrary

314 box-counting dimensions and the class of generic functions with other classes of

315 functions. Takens had shown that if we measured any single variable with sufficient

316 accuracy for a long period of time, it would be possible to construct the underlying

317 dynamical structure of the entire system from the behavior of that single variable

318 using delay coordinates and the embedding procedure. It was therefore possible to

319 construct a dynamical model of system evolution from the observed time series.

320 Introducing this idea here, four time series of the T_1, T_2, SOI and EAWMI factors

321 were chosen to construct the dynamical model.

322   The basic idea of statistical-dynamical model construction is discussed in

323 Appendix A and was introduced in our previous work (Zhang et al., 2006; Hong et al.,

324 2014).

325   A simplified second-order nonlinear dynamical model can be used to depict the

326 basic characteristics of atmosphere and ocean interactions (Fraedrich, 1987). Suppose

327 that the following nonlinear second-order ordinary differential equations are taken as

328 the dynamical model of reconstruction. In the equations, x_1, x_2, x_3, x_4 were used to

329 represent the time coefficient series of T_1, T_2, SOI and EAWMI.

$$\frac{dx_1}{dt} = a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + a_5 x_1^2 + a_6 x_2^2 + a_7 x_3^2 + a_8 x_4^2 + a_9 x_1 x_2 + a_{10} x_1 x_3 + a_{11} x_1 x_4 + a_{12} x_2 x_3 + a_{13} x_2 x_4 + a_{14} x_3 x_4$$

330
$$\frac{dx_2}{dt} = b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_1^2 + b_6 x_2^2 + b_7 x_3^2 + b_8 x_4^2 + b_9 x_1 x_2 + b_{10} x_1 x_3 + b_{11} x_1 x_4 + b_{12} x_2 x_3 + b_{13} x_2 x_4 + b_{14} x_3 x_4$$

$$\frac{dx_3}{dt} = c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4 + c_5 x_1^2 + c_6 x_2^2 + c_7 x_3^2 + c_8 x_4^2 + c_9 x_1 x_2 + c_{10} x_1 x_3 + c_{11} x_1 x_4 + c_{12} x_2 x_3 + c_{13} x_2 x_4 + c_{14} x_3 x_4$$

$$\frac{dx_4}{dt} = d_1 x_1 + d_2 x_2 + d_3 x_3 + d_4 x_4 + d_5 x_1^2 + d_6 x_2^2 + d_7 x_3^2 + d_8 x_4^2 + d_9 x_1 x_2 + d_{10} x_1 x_3 + d_{11} x_1 x_4 + d_{12} x_2 x_3 + d_{13} x_2 x_4 + d_{14} x_3 x_4$$

331

Based on the parameter optimization search method of GA in Appendix A, the time coefficient series of T_1, T_2, SOI and EAWMI from January 1951 to April 2008 are chosen as the expected data to optimize and retrieve model parameters. In order to eliminate the dimensionless relationship between variables, data standardization is to transform data from different orders of magnitude to the same order of magnitude, thus making the data comparable. So we used $x_{nor} = \dfrac{x - x_{\min}}{x_{\max} - x_{\min}}$ to normalize the raw value of each of the four predictors, then we used the normalized value to model and forecast. ~~To avoid the overfitting problem, we used $x_{nor} = \dfrac{x - x_{\min}}{x_{\max} - x_{\min}}$ to normalize the raw value of each of the four predictors, then we used the normalized value to model and forecast.~~ Finally, we made forecast results revert back to the raw data magnitude by $x = x_{nor}(x_{\max} - x_{\min}) + x_{\min}$.

In order to quantitatively compare the relative contribution of each item of our model to the evolution of the system, we calculated the relative variance contribution. The formula is as follows: $R_i = \dfrac{1}{n}\sum_{j=1}^{n}[\dfrac{T_i^2}{\sum_{i=1}^{14} T_i^2}], i = 1, 2, ..., 14$, Where n is the length of the data, $T_i = a_1 x_1, a_2 x_2, ..., a_{14} x_3 x_4$ is the item in the equation. According to our previous research (Hong et al., 2007), the variance contribution of the real item reflecting the performance of the model has a large proportion, while the variance contribution of the false term is almost zero, so we delete the weak items of $R_i < 0.01$.

352   After <u>deleting the weak items</u><s>eliminating weak items with small dimension</s>

353 <s>coefficients</s>, the nonlinear dynamical model of the first time series T_1, the second time

354 series T_2 ,SOI and EAWMI can be reconstructed as follows:

$$\frac{dx_1}{dt} = F_1 = -0.3328x_1 + 1.2574x_2 - 0.3511x_3 - 0.0289x_1^2 + 3.1280x_3^2 + 0.0125x_1x_2 + 2.7805x_1x_3 - 1.5408x_2x_4$$

355
$$\frac{dx_2}{dt} = F_2 = 1.0307x_1 - 3.1428x_2 + 0.3095x_4 + 4.2301x_1^2 - 1.2066x_2^2 + 2.5024x_4^2 - 0.2891x_1x_3 + 0.7815x_1x_4 - 0.4266x_3x_4$$

$$\frac{dx_3}{dt} = F_3 = -2.3155x_1 + 3.2166x_3 + 1.5284x_4 - 1.4527x_2^2 - 0.0034x_3^2 - 4.1206x_4^2 - 0.0025x_1x_4 + 0.0277x_2x_3 + 1.2860x_2x_4$$

$$\frac{dx_4}{dt} = F_4 = 0.4478x_2 - 0.0268x_4 + 0.8995x_1^2 - 2.3890x_3^2 + 0.2037x_4^2 + 1.3035x_1x_2 + 2.0458x_1x_4 - 2.0015x_2x_4$$

356   (2)

357   <s>The appropriate model coefficient estimates determine the robustness of the</s>

358 <s>model and the accuracy of forecast results. We should now judge whether the model</s>

359 <s>coefficients are appropriate or not.</s>

360   <s>Frist, the largest Lyapunov exponent (LLE) is one of the indexes that can</s>

361 <s>represent the characteristics of chaotic systems. The final Lyapunov exponents of Eq.</s>

362 <s>(2) were [0.0433, 0.0012, 0.1285], containing both a negative Lyapunov exponent</s>

363 <s>and two positive Lyapunov exponents, which demonstrate that our dynamic system is</s>

364 <s>indeed a chaotic system.</s>

365   <s>Second, we calculated the equilibrium roots of Eq. (2). Only the third</s>

366 <s>equilibrium was adjudged to be stable, based upon higher-order terms within the</s>

367 <s>Taylor series, the indices of which were mostly in accordance with the actual weather</s>

368 <s>system. The indices in the unstable equilibria could not accurately describe the actual</s>

369 <s>weather. Based on these two aspects, we can see that the model coefficient estimates</s>

370 <s>were reasonable and reflected the dynamical characteristics of the model.</s>

371   The model required testing. Because the training period was from January 1951

17

372   to April 2008, we chose T_1, T_2,SOI and EAWMI of May 2008, which were not used

373   as initial forecast data in the modeling. Next, the Runge–Kutta method was used to do

374   the numerical integration of the above equations, and every step of the integration was

375   regarded as 1 month's worth of forecasting results. As a result, forecast results of four

376   time series over a period of 20 months were obtained. Here, the focus was on the

377   forecast results of T_1 and T_2, as shown in Fig.2.

378   <u>The pearson correlation coefficient (CC) (Wang et al. 2009b) and the mean</u>

379   <u>absolute percentage error (MAPE)(Hu et al. 2001) are employed as objective</u>

380   <u>functions to calibrate the model. The CC evaluates the linear relationship between the</u>

381   <u>observed and predicting values and MAPE measures the difference between the</u>

382   <u>observed and predicting values.</u>

383   From Fig. 2, forecast performance of T_1 and T_2 within 5 months was better.

384   Using T_1 as an example, <u>the</u> ~~at this time, CC~~<u>CC</u>~~the temporal correlation~~ between model

385   predictions and corresponding observations <u>over the first five months forecasts</u> was

386   0.8966 and ~~the mean absolute percentage error (~~MAPE~~) (Hu et al.,~~

387   ~~2001),~~ $MAPE = \dfrac{1}{n}\sum\limits_{i=1}^{n}\left|\dfrac{D_e(i)-D_0(i)}{D_0(i)}\right| \times 100, \cancel{(n=5),}$ was 8.32%. However, after 5

388   months, MAPE increased rapidly, and was 31.29% at 10 months. The model forecast

389   then significantly diverged from observations, and the forecast became inaccurate.

390   After 10 months, the forecast results became increasingly worse, which indicated that

391   the forecast of the model after 5 months was unacceptable. The forecast results of

392   T_2 were similar to those of T_1.

393   The model's skill should be further assessed by cross-validated retroactive

394   hindcasts of the time series. As in the above example, omitting a portion of the time

395   series (12 months, ~~January~~ Jan. 1951 to ~~January~~ Dec. 1951~~2~~) from observations, we

396   trained the model based on the data from ~~February~~ Jan. ~~1952~~ 1951 to ~~December~~ Dec.

397   2010, and then predicted the omitted segments (12 months, Jan. 1951 to Dec.

398   1951~~January 1951 to January 1952~~). Then in the next prediction experiment, the

399   omitted segment is Jan.1952 to Dec. 1952 and the training samples are Jan. 1951 to

400   Dec.1951 and Jan.1953 to Dec.2010. So the forecast time series is Jan.1952 to Dec.

401   1952.We then repeated this procedure by moving the omitted segment along the

402   entirety of the available time series. Each experiment ~~have~~ has used the different

403   training sample and have established the different model equation (but the method is

404   the same). The similar process of the cross-validated retroactive hindcasts has also

405   been used in the previous literatures (Hu et al., 2017).

406   Finally, we obtained cross-validated retroactive hindcast results of T_1 and T_2, as

407   shown in Fig. 3. So the forecast results of 60 cross experiment (each experiment is the

408   prediction of the 12 month as Fig.2) according to the time sequence can merger into a

409   new time series (from Jan.1951-Dec.2010), and then the pearson correlation

410   coefficient (CC) and the mean absolute percentage error (MAPE) can be calculated by

411   the new prediction time series and the time series of the actual value. Figure 3 is

412   combined results of the 60 forecast experiments.

413   As Fig. 2, the forecast performance of T_1 and T_2 in Fig. 3 was not satisfactory.

414   The model forecast significantly diverged from observations, and the forecast became

415   inaccurate. The ~~temporal correlations~~CC of T_1 and T_2 between model predictions and

416   corresponding observations were 0.3411 and 0.4176, respectively. Additionally, the

mean absolute percentage errors (MAPE) of T_1 and T_2 were 65.42% and 57.56%,

respectively. This indicates that the forecast of the model in the long -term was

inaccurate and unacceptable.

The forecast result may be inaccurate when the integral forecasting time is long.

There will be a significant divergence which will cause an ineffective forecast. To

improve the forecast accuracy, the forecast not only depends on the integral equation

but also on a single initial value. Choosing the different initial value will cause

different forecast accuracy. For example, in a total of 60 cross-validated retroactive

hindcasts examples, the minimum MAPE was 37.65%, while the maximum MAPE

was 89.88%. A forecast, depending on a single initial value, will cause instability of

the forecast results. These two problems are addressed by introducing the

self-memorization principle in the next section.

4. Introduction of self-memorization dynamics to improve the reconstructed model

In the above discussion, it was shown that the accuracy of the forecast results of

equation (2) were unsatisfactory. To improve long-term forecasting results, the

principle of self-memorization can be introduced into the mature model (Gu, 1998;

Chen et al., 2009). The principle of self-memorization dynamics (Cao, 1993; Feng et

al., 2001) can be seen in Appendix B.

Based on Eq. (B10) in Appendix B, the improved model can be expressed as

438 follows:

$$\begin{cases} x_{1t} = \sum\limits_{i=-p-1}^{-1} \alpha_{1i} y_{1i} + \sum\limits_{i=-p}^{0} \theta_{1i} F_1(x_{1i}, x_{2i}, x_{3i}, x_{4i}) \\ x_{2t} = \sum\limits_{i=-p-1}^{-1} \alpha_{2i} y_{2i} + \sum\limits_{i=-p}^{0} \theta_{2i} F_2(x_{1i}, x_{2i}, x_{3i}, x_{4i}) \\ x_{3t} = \sum\limits_{i=-p-1}^{-1} \alpha_{3i} y_{3i} + \sum\limits_{i=-p}^{0} \theta_{3i} F_3(x_{1i}, x_{2i}, x_{3i}, x_{4i}) \\ x_{4t} = \sum\limits_{i=-p-1}^{-1} \alpha_{4i} y_{4i} + \sum\limits_{i=-p}^{0} \theta_{4i} F_4(x_{1i}, x_{2i}, x_{3i}, x_{4i}) \end{cases}$$ (3)

439 where y_i is replaced by the mean of two values at adjoining times; i.e.,

440 $y_i \equiv \frac{1}{2}(x_{i+1} + x_i)$; F is the dynamic core of the self-memorization equation, which

441 can be obtained from Eq. (2); and α and θ are the memory coefficients, the formula

442 for which can be found in Appendix B.

443 If the values of α and θ can be obtained, Eq. (3) can be used to obtain the

444 results of final prediction. The memory coefficients α and θ in Eq. (3) were

445 calibrated using the least-squares method with the same data (January 1951 to April

446 2008) as those used in Section 3. Eq. (3) can be deconstructed as follows (M is the

447 length of the time series):

448
$$X = \begin{bmatrix} x_{11} \\ x_{12} \\ \cdot \\ \cdot \\ \cdot \\ x_{1M} \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_{-p-1} \\ \alpha_{-p} \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{-1} \end{bmatrix}, Y = \begin{bmatrix} y_{-p-1,1} & y_{-p,1} & \cdots & y_{-1,1} \\ y_{-p-1,2} & y_{-p,2} & \cdots & y_{-1,2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ y_{-p-1,M} & y_{-p,M} & \cdots & y_{-1,M} \end{bmatrix}, \Theta = \begin{bmatrix} \theta_{-p} \\ \theta_{-p+1} \\ \cdot \\ \cdot \\ \cdot \\ \theta_{0} \end{bmatrix},$$

449
$$F = \begin{bmatrix} F_{-p,1} & F_{-p+1,1} & \cdots & F_{0,1} \\ F_{-p,2} & F_{-p+1,2} & \cdots & F_{0,2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ F_{-p,M} & F_{-p+1,M} & \cdots & F_{0,M} \end{bmatrix}$$

21

450   The matrix equation is:

$$X = Y\alpha + F\theta \qquad (4)$$

452   where $Z = [Y \vdots F]$, $W = \begin{bmatrix} \alpha \\ \vdots \\ \Theta \end{bmatrix}$.

453   Eq. (4) can be written as:

$$X = ZW \qquad (5)$$

455   The memory coefficients vector W can be calibrated using the least squares

456   method:

$$W = (Z^T Z)^{-1} Z^T X \qquad (6)$$

458   The memory coefficients a, θ can be obtained from Eq. (6). We then made a

459   prediction using the self- memorization equation (3), which used the p values before

460   t_0 .

461   The coefficients in F and W were used with the same training data from January

462   1951to Apr.il 2008. In the forecast examples, we trained both the coefficients in F and

463   W at the same time, but in the paper we describe them separately to facilitate the

464   reader for better understanding.

465   **5. Model prediction experiments**

466   **5.1 Forecast of time series T_1 and T_2**

467   The training sample for the model was from January1951 to April 2008. Here, from

468   Eq. (3), the forecast results using T_1, T_2, SOI and EAWMI factors can be calculated, called

469   as step-by-step forecast.

470   When the retrospective order p is confirmed, step-by-step forecasts can be

22

471 carried out. For example, when the T_1, T_2, SOI and EAWMI values of May 2008 were

472 forecast, y_i was obtained from the previous p +1 time of T_1, T_2, the SOI and the

473 EAWMI data, and $F_i(x_{1i}, x_{2i}, x_{3i}, x_{4i})$ was obtained from the previous p times of

474 T_1, T_2, the SOI and the EAWMI data. All four equations were integrated simultaneously.

475 Taking these in Eq. (3), we can get the T_1, T_2, SOI and EAWMI values of May 2008,

476 which these can be taken as the initial values for the next prediction step. Then, the

477 T_1, T_2, SOI and EAWMI values from June 2008 and so on, can be generated.

478 5.1.1 Determination of *p*

479 Based on the self-memorization principle, the self-memorization of the system

480 determines the retrospective order *p* (Cao, 1993). If the system forgets slowly,

481 parameters a and θ will be small and the *p* value should be high. The SSTA field

482 forecasts were on a monthly scale, the change of which was slow in contrast to

483 large-scale atmospheric motion. So parameters a and θ were small, and generally,

484 the *p* value was in the range 5 to 15.

485 The retrospective order *p* was obtained by a trial calculation method. We selected

486 the *p* values in the range 4 to 16 to construct the model. The ~~correlation~~

487 ~~coefficients~~CC ~~(CC)~~ and MAPE of long-term fitting test (from February 1951 to

488 December 2010) are shown in Table 2, which can be used as the standard to determine

489 the retrospective order *p*.

490 Table 2 indicates that when $p = 6$, the MAPE values of long-term fitting test

491 were the smallest and the ~~correlation coefficients~~CCs were the largest. Also, when *p*

492 from 5 to 9, ~~CCs~~ The CCs were all more than 0.58 and the forecast results were all

493   good, which is consistent with our interpretation of the physical mechanisms in

494   section 6.2 below. SOI and EMWMI were 5-12 months lead relationships with SST

495   (Xu et al., 1993; Chen et al, 2010; Wang et al., 2003). Using a cumulative period of

496   SOI-, EMWMI 5-8 months ahead as initial values can help improve the final forecast

497   results. Our results in table 2 are consistent with the actual physical ENSO process.

498   Therefore, we selected the retrospective order as p=6.

499   Then, the prediction experiments can be carried out, based on improved

500   self-memorization Eq. (3).

501   The improved self-memorization equation of T_1, T_2, SOI and EAWMI can then be

502   established. After the differential equation was discretely dealt with, the memory

503   coefficients were solved by the least-squares method given in section 4 (Training

504   period is January 1951 to April 2008). Finally, the improved prediction equation of

505   T_1, T_2, SOI and EAWMI, based on the self-memorization principle, can be expressed

506   as:

507
$$\begin{cases} x_{1t} = \sum_{i=-7}^{-1} \alpha_{1i} y_{1i} + \sum_{i=-6}^{0} \theta_{1i} F_1(x_{1i}, x_{2i}, x_{3i}, x_{4i}) \\ x_{2t} = \sum_{i=-7}^{-1} \alpha_{2i} y_{2i} + \sum_{i=-6}^{0} \theta_{2i} F_2(x_{1i}, x_{2i}, x_{3i}, x_{4i}) \\ x_{3t} = \sum_{i=-7}^{-1} \alpha_{3i} y_{3i} + \sum_{i=-6}^{0} \theta_{3i} F_3(x_{1i}, x_{2i}, x_{3i}, x_{4i}) \\ x_{4t} = \sum_{i=-7}^{-1} \alpha_{4i} y_{4i} + \sum_{i=-6}^{0} \theta_{4i} F_4(x_{1i}, x_{2i}, x_{3i}, x_{4i}) \end{cases} \quad (7)$$

508   where

$$\alpha=[\alpha_{ij}]=\begin{bmatrix} 0.0315 & -2.113 & 0.0284 & 2.1468 & 0.0688 & -0.7014 & 1.3248 \\ 0.4088 & -1.887 & -1.0233 & 1.5485 & 0.9028 & 1.0255 & -0.6443 \\ -0.9088 & -0.2557 & 0.9671 & -0.0054 & 1.0568 & 2.9764 & -0.5234 \\ 0.2088 & -1.0567 & 0.4891 & -0.5066 & -0.4890 & 1.4555 & 1.0966 \end{bmatrix}$$

$(i=0,1,...,4; j=-7,-6,...,-1)$

$$\theta=[\theta_{ij}]=\begin{bmatrix} 0.0485 & 0.0425 & -1.7688 & 0.8543 & 2.8901 & -0.1788 & -0.9066 \\ 0.07642 & 0.0941 & -1.2466 & -0.2288 & 0.1097 & 2.3221 & -1.4228 \\ -0.5288 & 1.2368 & -0.5568 & -0.0155 & 0.2886 & -0.1560 & 1.2775 \\ 1.5335 & -0.2887 & -0.5336 & -0.6072 & -0.5611 & 1.0225 & -1.0625 \end{bmatrix}$$

$(i=0,1,...,4; j=-6,-5,...,0)$

The step-by-step forecast was performed. The retrospective order $p=6$ means that earlier seven observation data ($p+1=7$) should be used during the forecasting process. The forecast results per month were saved for the next period predictions.

5.1.2 Long-term step-by-step forecasts of T_1 and T_2

To test the actual forecast performance of the above improved model, long-term step-by-step forecasts of T_1 and T_2 from May 2008 to December 2010 for 20 months were carried out, as shown in Fig. 4. The forecast results of T_1 and T_2 were good. Within 8 months, the ~~correlation coefficients~~CCs of T_1 and T_2 were 0.9163 and 0.9187. MAPEs of T_1 and T_2 were small, only 5.86% and 6.78%. The forecast time series from 8 months to 14 months gradually diverged, but the trend was acceptable. The CC~~correlation coefficient~~s of T_1 and T_2 reached 0.8375 and 0.8251, and MAPEs of T_1 and T_2 were 8.32% and 9.11%. After 14 months, forecast began to diverge and the error started to increase, but the ~~correlation~~ CC~~coefficient~~s of T_1 and T_2 remained about 0.6899 and 0.6782, and MAPEs reached 18.31% and 19.44%, which can be acceptable.

5.2 Cross-validated retroactive hindcasts of time series T_1 and T_2

As in section 3, the model's skill should be further assessed by cross-validated retroactive hindcasts of the time series. Because our step-by-step forecasts need the earlier seven observation data ($p + 1 = 7$), we can obtain cross-validated retroactive hindcast results of T_1 and T_2 from August 1951 to December 2010, as shown in Fig. 5.

From Fig. 5, the forecast performance of T_1 and T_2 was good. The CC~~correlation coefficient~~s of T_1 and T_2 were 0.7124 and 0.7036, respectively. The MAPEs of T_1 and T_2 were small, only 19.57% and 19.79%, respectively. The peaks and valleys of T_1 and T_2 were also forecasted accurately. The forecast results indicated that the cross-validated retroactive hindcast results of T_1 and T_2 were close to the observed values. Compared to Fig. 3, the improved model had better forecast abilities than the original model.

Many researchers (Zhang et al., 2003b; Smith, 2004) have used Oceanic Niño Index (ONI) which is used by the U.S. NOAA Climate Prediction Center to determine the El Niño and La Niña years. It defined that the ONIs of five consecutive months in winter were all more than 0.5 (less than -0.5) is the ElNiño (La Niña) year. Based on the above criterion, we can divide the total 60 years (1951-2010) into three categories. It includes the 18 examples of ElNiño year (such as 1958, 1964, 1966, etc.), 22 examples of LaNiña year (such as 1951, 1955, 1956, etc.) and the remaining 20 experiments of the neutral year. Since the details in Fig.5 is not clear, we list the forecast results of 60 experiments (including 18 El Niño examples, 22 La Niña examples and 20 Neutral examples) in table 3.

549   From table 3, the average of ~~CC~~ CC of both T_1 and T_2 of 60 experiments

550   within 6 months was more than 0.84 and MAPE was less than 8%. The average of ~~CC~~

551   CC within 12 months was more than 0.74 and MAPE was less than 12%. According

552   to the literature (Barranel et al., 1999), when MAPE was less than 15%, which means

553   the error was not great and the forecast results were good. Obviously, the forecast

554   results of ElNiño / LaNiña experiments were a little worse than those of neutral

555   examples, which means the forecast ability of our model for the abnormal situation

556   was a little worse than those for the normal situation. But even for ElNiño / LaNiña

557   experiments, the average of ~~CC~~ CC was still more than 0.7 and MAPE was less than

558   15%, which means the error was not too large and was still within an acceptable

559   range.

5.3 Forecast of the SSTA field

561   When we obtained the forecast results of the time coefficient series T_1 and T_2,

562   we submitted them into the following equation to reconstruct the forecast SSTA field:

563
$$\hat{x}_t = \sum_{n=1}^{2} E_n \bullet T_{nt}, t = 1, 2, ..., 12 \qquad (8)$$

564   where E_n, T_{nt} are the EOF space fields and forecast time coefficients,

565   respectively, and $\hat{x}_{tj}$ is the forecast SSTA field reconstructed by EOF.

566   After reconstruction of the space mode (treated as constant) and time coefficient

567   series (model prediction), the forecast of the SSTA fields was obtained, based on the

568   forecast results of T_1 and T_2 in Section 5.2. For economy of space, we cannot draw

569   all of the forecasted SSTA fields, so we selected a strong El Niño event (December

570   1997), a strong La Niña event (December 1999) and a neutral event (November 2002)

571   as examples.

572   Fig. 6 shows the forecast SSTA field during a strong El Niño event. From the

573   actual SSTA field in December 1997 (Fig. 6a), an obvious warm tongue structure

574   occurred in the area of [10°S～5°N, 90°W～150°W] in the Eastern Equatorial Pacific,

575   and a warm anomalous distribution arose in the west Pacific, which indicated a weak

576   El Niño event. The forecasted SSTA field of December 1997 is shown in Fig. 6b.

577   Although the range of warm tongue was a litter bigger than the actual situation, the

578   forecast shape was similar to the actual field and also the contour lines were similar.

579   The average MAPE between the forecast field and the actual field is 8.56%, which

580   was controlled within 10%. The forecast results of the improved model event were

581   quite good for the El Niño event.

582   Fig.7 shows the forecasted SSTA field of a strong La Niña event. From the actual

583   SSTA field in December 1999 (Fig. 7a), an obvious cold pool occurred in the area of

584   [10°S～10°N, 120°W～180°W] in the Equatorial Pacific, which covered the Niño3.4

585   area. This SSTA field presented a strong strength La Niña event. The forecast SSTA

586   field from December 1999 is shown as Fig. 7b. Although the strength of the cold pool

587   was weaker than the actual situation, the forecast shape was similar to that of the

588   actual field. The average MAPE between the forecast field and the actual field was

589   9.69%. The errors were larger than that of the El Niño event, but they can be

590   controlled within 10%, which is acceptable.

591   Fig. 8 shows the forecasted SSTA field of a neutral event. From the actual SSTA

592   field in November 2002 (Fig. 8a), a warm pool occurred in the area of [10°S～10°N,

593   120°W～180°W] in the Equatorial Pacific, which covered the Niño3.4 area. However,

594 the warm pool was small and weak, which represented a neutral event. The forecasted

595 SSTA field from November 2002 is shown in Fig. 8b. Comparing Figures 6, 7 and 8,

596 we can see that the forecasted SSTA field of a neutral event was a little worse than

597 that of the El Niño and La Niña events. The forecasted shape of the SSTA field

598 basically described the actual situation, but the warm pool in the Niño3.4 area was

599 stronger and bigger than that of the actual situation, which indicated a borderline El

600 Niño event. The average MAPE between the forecasted field and the actual field was

601 14.50%, which was big but can be accepted.

602 We obtained the average values of MAPE of 18 El Niño events, 22 La Niña

603 events and 20 neutral events, which were 9.52%, 9.88% and 14.67%, respectively,

604 representing a good SSTA field forecasting ability of our model.

5.4 Forecast of ENSO index

606 The ENSO index can be represented as the sea surface temperature anomaly

607 (SSTA) in the Niño-3.4 region (5 °N-5 °S, 120 °-170 °W) and the ENSO index

608 forecast was the 3-month forecast (Barnston et al. 2012). So we also can pick up the

609 ENSO index from the above forecasted SSTA field. The forecast results of the ENSO

610 index within 20 months can also be obtained. The definition of lead time can be seen

611 in the reference (Barnston et al. 2012). Therefore, similar to the forecast experiment in

612 section 5.1, a succession of running 3-month mean SST anomalies with respect to the

613 climatological means for the respective prediction periods, averaged over the Niño 3.4

614 region, can be obtained, as demonstrated in Fig. 9.

615 The evaluation criteria of the ENSO index is the temporal correlation (TC), its

definition and specific calculation steps can be seen in these literatures (Kathrin et al.,2016; Nicosia et al. 2013); The TC is often used to measure the prediction effect of the ENSO index. For example, Barnston et al.in 2012 also used the TC to compare the forecast skill of 21 real-time seasonal ENSO models.

The forecast results within lead times of 18 months are shown in Fig. 9, which demonstrate that the forecast results of the ENSO index are good. Within lead time of 12 months, the ~~correlation coefficient~~TC was 0.8985 and the MAPE value was small, only 8.91%. In addition, the borderline La Niña event in 2008–2009 was predicted well. After lead times of 12 months, forecasts began to diverge and the errors started to increase. Although the ~~correlation coefficient~~TC remained approximately 0.61, MAPE reached 18.58%. Therefore, a moderate strength El Niño event that occurred in 2009/10 was not predicted.

We should give more examples to test the ENSO prediction ability of our model. As in section 5.3, we can divide 60 examples as three types, which are examples of ElNiño year, LaNiña year and neutral year. Finally, we can obtain the forecast results of different types of examples in different lead times, as shown in table 4.

From table 4, the average ~~CC~~ TC of 60 experiments was 0.712 and the average MAPE was 7.62% within 12 months for all seasons of lead time, which indicates that the overall ENSO forecast ability of our model was good. The forecast results of the El Niño examples were significantly worse than those of La Niña examples, while the forecast results of La Niña examples were significantly worse than those of neutral examples, which show the model forecast ability of the abnormal state was worse than

30

638   the normal state of the ENSO index. Even for the forecast results of El Niño examples,

639   the average ~~CC~~ TC was still above 0.6 and the average MAPE can be controlled

640   below 10%, which means the forecast results were still in the acceptable range. Our

641   model not only accurately predicted the stronger El Niño and La Niña phases but also

642   the neutral states. ~~But the forecast results in summer were a little worse than those in~~

643   ~~winter, as shown in Fig.10.~~

644   The ENSO forecast often had a spring predictability barrier (Webster, 1999),

645   which was most prominent during decades of relatively poor predictability

646   (Balmaseda et al., 1995).To test our model, the skill should be computed over the

647   entire time series and separately for seasonal subsets of the time series. From the

648   table4, we can see that ~~The average cumulative correlation coefficient and MAPE of~~

649   ~~winter were compared with those of summer, as shown in Fig.10. The average~~

650   ~~cumulative correlation and average cumulative MAPE values between the forecast~~

651   ~~values and the actual values changed with time, from which good trends of forecast~~

652   ~~results can be seen. As long as the forecast time increased, the cumulative MAPE~~

653   ~~increased and the correlation decayed gradually. The forecast results appeared to~~

654   ~~diverge. A~~although the forecast results of the present model in the ~~summer~~ spring

655   were worse than in the ~~winter~~autumn, the margin was not high, which means the

656   model can overcome the "spring predictability barrier," to some extent.

657 **5.5 Compared with six mature models**

658   Barnston et al. (2012) compared many ENSO forecast models. Based on his

659   research, we selected four high quality dynamical models, including ECMWF, JMA,

660 the National Aeronautics and Space Administration Global Modelling and

661 Assimilation Office (NASAGMAO) and the National Centre for Environmental

662 Prediction Climate Forecast System (NCEP CFS; Version1). Two high quality

663 statistical models also be selected, including the University of California, Los Angeles

664 Theoretical Climate Dynamics (UCLA-TCD) multilevel regression model and the

665 NOAA/NCEP/CPC constructed Analogue (CA) model. The detail of the above

666 models can be seen in these references (Reynolds~~Reynoldset~~ al., 2002; Luo et al.,

667 2005; Barnston et al., 2012).

668 We then compared the forecast ability of the above six models with that of our

669 model. All of the experiments of our model and six other models were conducted

670 under the same conditions using the same historical data for modelling and the same

671 initial values to forecast. In the CPC website, there are detailed explanations of six

672 models' training samples and the initial values. So we do not need to install all these

673 models on their own machines and run them for forecasting. We just made training

674 samples and initial values of our model were the same with those of selected six

675 models. At an 8-month lead time, the ~~correlation ability~~TC of our model for all

676 seasons combined was 0.613 (Fig. ~~11~~10). In brief, the forecast ability of the ECMWF

677 model was slightly better than that of our model but the ability of the other 5 models

678 was worse than that of our model. While, in regard to the forecast length, the ~~temporal~~

679 ~~correlation~~TC within 12 months of our model is greater than 0.6, which was superior

680 to the ECMWF model. In addition, the forecast results of the UCLA-TCD model and

681 the CPC CA model reduced quickly after 5-month lead times, so the forecast ability of

our model was more stable than them.

The root mean square error (RMSE) was also examined to assess the performance of discrimination and calibration. Barnston et al. (2012) believed that all seasonal RMSE values contributed equally to a seasonally combined RMSE. So we drew figure 12 11 to show seasonally combined RMSE.

From Fig. 11 11 0 and Fig. 12 11, we can see the highest correlation tend to have lower RMSE. So the RMSE of our model was slightly higher than that of ECMWF model, but it was much lower than those of the other 5 models. Figure 11 and Figure 12 is the average CC TC and RMSE of the 240 experiments of compared with six mature models, covers a variety of different types of ENSO and different lead time. So those samples should be really representative .

6. Conclusions and discussion

6.1 Conclusions

A new forecasting model of the SSTA field was proposed based on a dynamic system reconstruction idea and the principle of self-memorization. The approach of the present paper consisted of the following steps:

(1) The SST field can be time (coefficients)-space (structure) deconstructed using the EOF method. Take T_1, T_2, SOI and EAWMI and consider them as trajectories of a set of four coupled quadratic differential equations based on the dynamic system reconstruction idea. The parameters of this dynamic model were estimated using a GA.

(2) The forecast results of the dynamic model can be improved by the

self-memorization principle. The memory coefficients in the improved self-memorization model were obtained using the GA method.

(3) The long-term step-by-step forecast results and cross-validated retroactive hindcast results of time series T_1 and T_2 are all found to be good, with the ~~a~~ correlation ~~coefficient~~CC ~~–~~of approximately 0.80 and ~~a mean absolute percentage error~~ the MAPE of less than 15%.

(4) The improved model was used to forecast the SSTA field. The forecasted SSTA fields of three types of events are accurate. Not only is the forecast shape similar to the actual field but also the contour lines are similar.

(5) The improved model was also used to forecast the ENSO index. The average ~~correlation coefficient~~TC of 60 examples within 12 months is 0.712, and the MAPE value is small, only 7.62%, which proves that the improved model has better forecasting results of the ENSO index. Although the forecast results of the model in the summer were worse than in the winter, the margin was not high, which means that the model can overcome the spring predictability barrier to some extent. Finally, compared with the six mature models, the new dynamical-statistical forecasting model has a scientific significance and practical value for the SST in the eastern equatorial Pacific and El Niño/La Niña event predictions.

6.2 Discussion

L'Heureux et al.(2013) reported that using different data sets and time periods, the 2nd EOF is not stable, being entirely due to the strong trend. So we need to do more experiments to prove that we choose the second mode of EOF to be appropriate,

and whether different time periods will make us forecast unstable or not. Our original data is the monthly average SST data from January 1951 to Dec, 2010, which are 60 years. We will increase the length of the data for 20 years (Jan.1931 –Dec,2010), for 10 years (Jan.1941- Dec.2010) and decrease the length of the data for 10 years (Jan.1961- Dec.2010), for 20 years (Jan.1971- Dec.2010). And then we use the same method to reconstruct a model and forecast the ENSO index as section5.4. The prediction results are shown in the table5.

From the table, we can see that in the 60 experiments, the prediction results of the data period increased by 20 years are the best, and the prediction results of the data period decreased by 20 years is the worst. This is because the more data we use, the more information it contain. But from the table we can also see the difference among forecast results of both TC and MAPE of five different sample data are less, and no abnormal change suddenly worse or better appear. All these indicate that using different data sets and time periods, even though may have a certain impact on the pattern of the 2nd EOF, but the impact on our forecast is not great and it will not make our forecast unstable.

Actually, how many variables and which variables are used in our model become a key issue to be resolved. We are a complex four factor differential equations coupling model. We are a complex coupled model of four factor differential equations, so we are more concerned with the correlation between each other. The correlation must be considered as an important criterion to select the factors, but in order to further verify the correctness of the selection criterion, we have carried out

the prediction experiments (the 60 cross-validated retroactive hindcasts experiments of the ENSO index for all seasons combined at lead times of 8 months) of different variables.

We can see that for all the forecast results of the models of different variables, the prediction results of T_1, T_2, SOI is the best among those of the three factors and the prediction result of $T_1, T_2, \text{SOI}, \text{EAWMI}$ is the best among those of the four factors. But the prediction result of $T_1, T_2, \text{SOI}, \text{EAWMI}$ is best among all, which proves that our selection factors are correct. In our previous study (Hong et al., 2015), the model of the Western Pacific subtropical high was established by using the correlations as a criterion to select factors and their forecast results are also good. Now we use the correlations as a criterion to select factors is also in line with our previous research.

Because the formula of our model includes a linear combination of 4 variables ($T_1, T_2,$ SOI, EAWM), statistical forecasting requires independence between predictors. We can calculate the correlation coefficients between variables, as shown in table 5. In fact, as Table 5 shows, the correlation coefficients between the factors were all less than 0.45, indicating the independence between factors. So this does not generate too much redundancy and can avoid an overfitting problem, which can destroy the stability of the model.

The definition of overfitting: The learned hypothesis may fit the training set very well, but fail to predict to new examples (fail to fit additional data or predict future observations reliably).

The potential for overfitting depends not only on the number of parameters and

data but also the conformability of the model structure with the data shape, and the magnitude of model error compared to the expected level of noise or error in the data(Burnham and Anderson, 2002). So there are many reasons causing the overfitting phenomenon. But this does not mean having many parameters relative to the number of observations inevitably causes the overfitting problem (Golbraikh et al., 2003). There is no evidence that more parameters will be certain to result in overfitting. Based on the definition of overfitting and the previous studies(Golbraikh et al., 2003; Everitt and Skrondal,2010), we can judge whether a model is overfitting or not by the accuracy of prediction results of independent samples (Golbraikh and Tropsha, 2002; Qin and Li, 2006).

In the sample training, our model does not purposely pursue the high degree of the training samples fitting and improve the effectiveness of the independent generalization. In fact in our paper the forecast results of the Cross-validated retroactive hindcasts (section 5.2) and the independent samples validation (table3 and table4) are both good. Especially, the independent samples validation of the ENSO index as the table4, we have carried out the 240 independent sample validation prediction of four seasons of different ENSO events and the coverage of independent samples test is very wide. Moreover, compared with 6 mature prediction models, the forecast results of our model are also good, which prove the overfitting problem does not exist in our model. According to the previous literature (Islam and Sivakumar, 2002; Sivakumar et al.,2001), we can see that prediction principle and structure of the phase space reconstruction (PSR) of dynamical system is not the same with the

traditional neural network, and in the small sample situation the forecasting results of PSR model are better than those of the traditional neural network (Sivakumar et al., 2002), which can be verified in the independent sample test (table3 and table4). So according to the definition of overfitting, we can say the over fitting phenomenon does not exist in our model.

The introduction of self-memorization essentially introduces a lot of new coefficients, which may cause an overfitting problem. Because we have selected a model of four variables, there is a total of 62 parameters. In order to avoid the overfitting problem, the sample sizes are more than 10% of the amount of parameters. So our sample size is greater than 620 data to avoid the overfitting problem. If we choose the model of three variables, the parameters in which will be less, the sample size in this situation can be less. But the forecast results may be a little worse, based on the analysis in section 2.3. So the length of training samples is related to the number of parameters of our model.

Also, we have tried to detrend our data before the model constructed. But we found the results didn't change too much. That is mean our model is not very sensitive to climate change, so the detrended data has little effect for our model to improve the forecast effect.

Compared with the original model, why the improved model has good forecast results and can overcome the spring predictability barrier to some extent are as follow: Recently, many studies have pointed out that spring is the most unstable season of the air - sea interaction and the error is likely to develop or grow in the spring, resulting in

814   the spring predictability barrier (Zhang et al, 2012; Philander et al., 1992). When the

815   original model uses the indexes in summer as the initial values to predict, the SOI

816   factor representing the air-sea interaction is most unstable in the spring and the

817   EMWMI factor does not have much influence on ENSO in summer, so the forecast

818   results using the indexes in summer as the initial values are certainly much worse than

819   those using the indexes in the winter as the initial values. That is why our original

820   model does not overcome the spring predictability barrier.

821   However, the introduction of the self-memorization dynamics principle can help

822   our model overcome the spring predictability barrier to some extent. Although the

823   lead time is still summer (such as JJA), the information of the initial value actually

824   contains the previous p + 1 month (in this case p = 6, which contains the information

825   of the previous seven months, including the information of T_1, T_2 , SOI, EMWMI

826   factor in winter (January, February), spring (March, April, May) and summer (June

827   and July)). From the dynamical analysis, in this situation, the information and

828   interaction relationship of four factors have been a long period (from winter to

829   summer) accumulated, containing much air-sea interaction processes and winter

830   monsoon continued abnormal information, so the forecast results of our improved

831   model will be much better than the original model which simply uses only one initial

832   value. That is why the improved model overcomes the spring predictability barrier to

833   some extent.

834   The forecast results of our model are good, but it still has some problems:

835   (1) The inclusion of these terms and the physical processes do these terms in

equation (2) represent are important, especially for the discussion of dynamical characteristics of the dynamical model. But now we are difficult to give a clear meaning. Now the main work of our paper is the prediction experiments of the model. For the reason of time and length, this paper mainly discusses the prediction results of the model. The physical processes do these terms represent and the discussion of the dynamical characteristics of the model will be the focus of our next work. Before this, we have also used the Takens' delay embedding theorem to reconstruct the dynamical model of the Western Pacific subtropical high(WPSH). And Based on the reconstructed dynamical model, dynamical characteristics of WPSH are analyzed and an aberrance mechanism is developed, in which the external forcings resulting in the WPSH anomalies are explored, which have been published (Hong et al., 2016). We also study the bifurcation and catastrophe of the West Pacific subtropical high ridge index of a nonlinear model (Hong et al., 2017). Based on our previous method and work, our next work is to analyse the physical processes and the dynamical characteristics of the SST field.

Although the reason why the improved model has good forecast results has disucussed in the section6.2, the deep physical mechanisms that the proposed model has dealt with is not very clear, so its dynamical characteristics should be further analysed.

—— (2)The experiments in the present study have proven that the forecasting results of the improved model are good for large-scale systems, such as ENSO events, and the forecasting period has been extended. However, for small-scale systems, such as

858 Hurricanes, whether the forecast results could be improved using the present

859 improved model needs to be further verified.

860   (3) Our paper focuses primarily on these defined indices with T_1, T_2 to

861 reconstruct a prediction model. Maybe, we can select variables (predictor) based on

862 EOF analysis and our model may be a more physically oriented model. Maybe we can

863 learn from Yim et al. (2013; 2015) to draw correlation maps between these fields and

864 the SSTA field and select the predictors from physical considerations. All these above

865 questions require that a lot of experiments to be carried out.

866   These items will be our future work.

867

875

876 **APPENDIX A: THE PRINCIPLE OF DYNAMICAL MODEL**

877 **RECONSTRUCTION**

878   Suppose that the physical law of a nonlinear system going by over time can be

879 expressed as the following difference form:

$$\frac{q_i^{(j+1)\Delta t} - q_i^{(j-1)\Delta t}}{2\Delta t} = f_i(q_1^{j\Delta t}, q_2^{j\Delta t}, ..., q_i^{j\Delta t}, ..., q_N^{j\Delta t}) \quad j = 2,3,......M-1 \tag{A1}$$

where f_i is the generalized nonlinear function of $q_1, q_2, ...q_i, ...q_N$, N is the number of variables, and M is the length of observed data. $f_i(q_1^{j\Delta t}, q_2^{j\Delta t}, ..., q_i^{j\Delta t}, ..., q_N^{j\Delta t})$ can be assumed to contain two parts: G_{jk} representing the expanding items which contain variable q_i, P_{ik} just representing the corresponding parameters which are real numbers ($i = 1, 2, ...N$, $j = 1, 2, ...M$, $k = 1, 2, ..., K$).

It can be supposed as follows:

$$f_i(q_1, q_2, ...q_n) = \sum_{k=1}^{K} G_{jk} P_{ik} \tag{A2}$$

$D = GP$ is the matrix form of Eq.(A2), in which

$$D = \begin{Bmatrix} d_1 \\ d_2 \\ ... \\ d_M \end{Bmatrix} = \begin{Bmatrix} \dfrac{q_i^{3\Delta t} - q_i^{\Delta t}}{2\Delta t} \\ \dfrac{q_i^{4\Delta t} - q_i^{2\Delta t}}{2\Delta t} \\ ... \\ \dfrac{q_i^{M\Delta t} - q_i^{(M-2)\Delta t}}{2\Delta t} \end{Bmatrix}, \quad G = \begin{Bmatrix} G_{11}, G_{12}, ...G_{1K} \\ G_{21}, G_{22}, ...G_{2,K} \\ ... \\ G_{M1}, G_{M2}, ...G_{M,K} \end{Bmatrix}, \quad P = \begin{Bmatrix} P_{i1} \\ P_{i2} \\ ... \\ P_{iK} \end{Bmatrix} \tag{A3}$$

Parameters of the above equation can be determined through inverting the observed data. Vector P which satisfies the above equation can be solved, based on a given vector D. Assuming q is unknown, it is a nonlinear system. However, assuming P is unknown, it is a linear system.

With the restriction $S = (D-GP)^T(D-GP)$ as a minimum, GA is introduced as an optimization solution search in the model parameters space.

Assuming that the parameters matrix P is the population (solutions), the $S = (D-GP)^T(D-GP)$ is an objective function, $l_i = \dfrac{1}{S}$ is the value of individual fitness, and $L = \sum_{i=1}^{n} l_i$ is the value of total fitness. The operating steps of GA include: creation and coding of initial population (solutions), fitness calculation, the choice of

900 male parents, crossover and variation, etc. A detailed theoretical explanation can be

901 got from Wang (2001). The step length is 1 month during the calculation. After

902 optimization searches and genetic operations, the target value can be rapidly

903 converged on and each optimal parameter of the dynamical equations can be obtained.

904 Through the above approach, we can obtain parameters of a nonlinear

905 dynamical system, and reconstruct the nonlinear dynamical equations from observed

906 data.

907

908 **APPENDIX   B: THE   MATHEMATICAL   PRINCIPLE   OF**

909 **SELF-MEMORIZATION DYNAMICS OF SYSTEMS**

910 The dynamical equations of a system can be expressed as:

911
$$\frac{\partial x_i}{\partial t} = F_i(x, \lambda, t) \ i = 1, 2, ..., J \tag{B1}$$

912 where J is an integer，x_i is the ith variable of the system state, and λ is

913 the parameter. Equation (B1) represents the relationship between a source function

914 F and a local change of x. Obviously, x is a scalar function with time t and

915 space r_0. A set of time $T = [t_{-p} .. t_0 .. t_q]$ can be considered, where t_0 is an initial

916 time. A set of space $R = [r_a ... r_i ... r_\beta]$ can be considered, where r_i is a spatial point.

917 An inner product in space $L^2 : T \times R$ is defined by:

918
$$(f, g) = \int_a^b f(\xi) g(\xi) d\xi, f, g \in L^2 \tag{B2}$$

919 Accordingly, a norm can be defined as:

920
$$\|f\| = [\int_a^b (f(\xi)^2 d\xi]^{1/2}$$

43

921   For a completion L^2, it can become a Hilbert space H. A generalized one

922   in H can be regarded as a solution of the multi-time model. By introducing a

923   memorization function $\beta(r,t)$, we can obtain:

924    $$\int_{t_0}^{t} \beta(\tau)\frac{\partial x}{\partial \tau}d\tau = \int_{t_0}^{t} \beta(\tau)F(x,\tau)d\tau \qquad (B3)$$

925   where r in $\beta(r,t)$ can be dropped through fixing on the spatial point r_0. Suppose

926   that function $\beta(r,t)$ and variable x etc. are all continuous, differentiable and

927   integrable, an integration by the left parts of Eq. (B3) can be made as:

928    $$\int_{t_0}^{t} \beta(\tau)\frac{\partial x}{\partial \tau}d\tau = \beta(t)x(t) - \beta(t_0)x(t_0) - \int_{t_0}^{t} x(\tau)\beta'(\tau)d\tau \qquad (B4)$$

929   where $\beta'(t) = \partial \beta(t)/\partial t$. The mean value theorem can be introduced into the third

930   term in Eq. (B4), the following equation can be obtained:

931    $$-\int_{t_0}^{t} x(\tau)\beta'(\tau)d\tau = -x^m(t_0)[\beta(t) - \beta(t_0)] \qquad (B5)$$

932   where $x^m(t_0) \equiv x(t_m), t_0 < t_m < t$. Substituting Eq. (B4) and Eq. (B5) in Eq. (B3) and

933   carrying out an algebraic operation, the following equation can be obtained:

934    $$x(t) = \frac{\beta(t_0)}{\beta(t)}x(t_0) + \frac{\beta(t) - \beta(t_0)}{\beta(t)}x^m(t_0) + \frac{1}{\beta(t)}\int_{t_0}^{t} \beta(\tau)F(x,\tau)d\tau \qquad (B6)$$

935   Because the x value which is at initial time t_0 and middle time t_m, only on

936   the fixed point r_0 itself , relates to the first term and the second term in Eq. (B6) ,

937   they are be called as a self-memory term. Also, we can call the third term as an

938   exogenous effect, i.e., which is contributed by other spatial points.

939   Similarly as Eq. (B4), for multi-time t_i, $i = -p, -p+1...,t_0, t$, it gives

940 $$\int_{t_{-p}}^{t_{-p+1}} \beta(\tau)\frac{\partial x}{\partial \tau}d\tau + \int_{t_{-p+1}}^{t_{-p+2}} \beta(\tau)\frac{\partial x}{\partial \tau}d\tau + ... + \int_{t_0}^{t} \beta(\tau)\frac{\partial x}{\partial \tau}d\tau = \int_{t_{-p}}^{t} \beta(\tau)F(x,\tau)d\tau .$$

941   After the same term $\beta(t_i)x(t_i), i = -p+1, -p+2, ..., 0$ was eliminated, we

942 have

943 $$\beta(t)x(t) - \beta(t_{-p})x(t_{-p}) - \sum_{i=-p}^{0}[\beta(t_{i+1}) - \beta(t_i)]x^m(t_i) - \int_{t_{-p}}^{t}\beta(\tau)F(x,\tau)d\tau = 0 \quad \text{(B7)}$$

944   As a matter of convenience, we set $\beta_t \equiv \beta(t), \beta_0 \equiv \beta(t_0), x_t \equiv x(t), x_0 \equiv x(t_0)$; the

945 following text uses similar notations. Then, Eq. (B7) can be expressed as:

946 $$\beta_t x_t - \beta_{-p}x_{-p} - \sum_{i=-p}^{0}x_i^m(\beta_{i+1} - \beta_i) - \int_{t_{-p}}^{t}\beta(\tau)F(x,\tau)d\tau = 0 \quad \text{(B8)}$$

947   Setting $x_{-p} \equiv x_{-p-1}^m, \beta_{-p-1} = 0$, the Eq. (B8) can be written as:

948 $$x_t = \frac{1}{\beta_t}\sum_{i=-p-1}^{0}x_i^m(\beta_{i+1} - \beta_i) + \frac{1}{\beta_t}\int_{t_{-p}}^{t}\beta(\tau)F(x,\tau)d\tau = S_1 + S_2 \quad \text{(B9)}$$

949   S_1 is called as a self-memory term and S_2 is called as an exogenous effect term.

950   For the convenience of calculations, the above self-memorization equation can

951 be discretized. The differential by difference and the summation can replace the

952 integration in Eq. (B9), and the mean of two values which are at adjoining times; i.e.,

953 $x_i^m \approx \frac{1}{2}(x_{i+1} + x_i) \equiv y_i$ can simply replace x_i^m .

954   Taking an equal time interval $\Delta t_i = t_{i+1} - t_i = 1$ and incorporating β_i and β_t,

955 we can obtain a discretized self-memorization equation as follows:

956 $$x_t = \sum_{i=-p-1}^{-1}\alpha_i y_i + \sum_{i=-p}^{0}\theta_i F(x,i) \quad \text{(B10)}$$

957 where F is the dynamic kernel of the self-memorization equation, $\alpha_i = \frac{(\beta_{i+1} - \beta_i)}{\beta_t}$;

958 $\theta_i = \frac{\beta_i}{\beta_t}$.

959   Based on Eq. (B10), the above technique performed computations and the

960   forecast can be called as a self-memorization principle.

961

962

963   **REFERENCES**

964   Ashok K, Guan Z, Yam agata T：Impact of the Indian Ocean Dipole on the decadal relationship

965   between the Indian mon soon rainfall and ENSO, Geophys Res Let,28(23), 4499-4502, 2001.

966   Balmaseda M.A., Davey M.K. and Anderson D.L.T.: Decadal and seasonal dependence of ENSO

967   prediction skill,J Clim.,8, 2705–2715, 1995.

968   Barnston A. G., et al.: Skill of real-time seasonal ENSO model predictions during 2002-2011,Bull.

969   Amer. Meteor. Soc.,93, 631-651, 2012.

970   Belkin M. and P. Niyogi: Laplacian eigenmaps for dimensionality reduction and data

971   representation,Netural Comput.,15,1373-1391, 2003.

972   Bjerknes J.: Atmsopheric telconnections from the equtorail Pacific,Mon. Wea. Rev.,97,163-172, 1969.

973   Burnham, K. P.; Anderson, D. R: Model Selection and Multimodel Inference (2nd ed.),

974   Springer-Verlag, 2002.

975   Cao H. X.: Self-memorization Equation in Atmospheric Motion,Science in China (Series B),36(7),

976   845-855, 1993.

977   Chen D., S. E. Zebiak, A. J. Busalacchi and M. A. Cane: An Improved Procedure for El Nirio

978   Forecasting: Implications for Predictability,Science, 269, 1699-1702, 1995.

979   Chen G., Shao B. M. Han Y., et al.: Modality of semiannual to multidecadal oscillations in global sea

980   surface temperature variability. Journal of Geophysical Research, 115, 1-14, 2010.

981   Chen X. D., Xia J., Xu Q.: Differential Hydrological Grey Model(DHGM) with self-memory function

982   and its application to flood forecasting,Sci China Tech Sci.,52,1039–1049, 2009.

983   Clarke A. J. and S. Van Gorder: Improving El Niño prediction using a space-time integration of

984   Indo-Pacific winds and equatorial Pacific upper ocean heat content, Geophys. Res. Lett.,30,1399.

985   doi:10.1029/2002GL016673, 2003.

986   Delecluse P., Davey M., Kitamura Y., Philander S., Suarez M., Bengtsson L.: TOGA review paper:

987   coupled general circulation modeling of the tropical Pacific,J Geophys Res,103,14357–14373, 1998.

46

Davey M., Huddleston M., Sperber K.R., et al.: A study of coupled model climatology and variability in tropical ocean regions,Clim. Dyn.,18,403–420, 2002.

Dommenget and Latif: A Cautionary Note on the Interpretation of EOFs, Journal of Climate,15(2),216–225, 2002.

Drosdowsky W.: Statistical prediction of ENSO (Niño 3) using sub-surface temperature data,Geophys. Res.Lett., 33 , L03710. doi:10.1029/2005GL024866, 2006.

Everitt B.S., Skrondal A.: Cambridge Dictionary of Statistics, Cambridge University Press, 2010.

Feng G. L., Cao H. X., Gao X. Q., et al.: Prediction of precipitation during summer monsoon with self-memorial model,Adv Atmos Sci.,18 ,701–709, 2001.

Fraedrich K.: Estimating weather and climate predictability on attractors,J .A tmos.Sci.,44,7 22-728, 1987.

Glantz MH, Katz RW, Nicholls N (eds): Teleconnections linking worldwide climate anomalies, 74pp,Cambridge University Press, Cambrige, UK, 1991.

Golbraikh A. and Tropsha A.: Beware of q 2 ! Journal of Molecular Graphics and Modelling, 20 , 269–276, 2002.

Golbraikh A.,Shen M., Xiao Z. Y., Xiao Y. D., Lee Kuo-Hsiung, Tropsha A.: Rational selection of training and test sets for the development of validated QSAR models. Journal of Computer-Aided Molecular Design, 17(2), 241-253, 2003.

Gu X. Q.: A spectral model based on atmospheric self memorization principle,Chinese Science Bulletin,43(20),1692-1702 , 1998.

Hong M., Zhang R., Wu G. X., et al.,: A Nonlinear Dynamic System Reconstruction of the Subtropical High Characteristic Index based on Genetic Algorithm. Chinese Journal of Atmospheric Sciences,31(2):346-352, 2007.

Hong M., Zhang R.andMa C. C.et al.: A Non-Linear Dynamical–Statistical Model for Reconstruction of the Air–Sea Element Fields in the Tropical Pacific Ocean,Atmosphere-Ocean, doi: 10.1080/07055900.2014.908765, 2014.

Hong M., Zhang R., et al.: Reconstruction and forecast experiments of a statistical-dynamical model of the Western Pacific subtropical high and Eastern Asian summer monsoon factors，Weather and Forecasting, 30:206-216 , 2015.

Hong M., Zhang R., et al.: Catastrophe and Mechanism Analyses of Multiple Equilibria in the Western

1018 Pacific Subtropical High System Based on Objective Fitting of Spatial Basis Functions. Monthly

1019 Weather Review, 144, 997-1015, 2016.

1020 Hong M., Zhang R., et al.: Bifurcations and catastrophes in a nonlinear dynamical model of the western

1021 Pacific subtropical high ridge line index and its evolution mechanism, Theor. Appl. Climatol., 129,

1022 363-384, 2017.

1023 Hu, T.S., K.C. Lam, and S.T. Ng: River flow time series prediction with a range-dependent neural

1024 network,Hydrol. Sci. J., 46, 729–745, 2001.

1025 Hu Y. J., Zhong Z., Zhu Y. M. et al.: A statistical forecast model using the time-scale decomposition

1026 technique to predict rainfall during flood period over the middle and lower reaches of the Yangtze

1027 River Valley. Theoretical and Applied Climatology, doi: 10.1007/s00704-017-2094-9,2017.

1028 Huang. J., Y. Yi, S. Wang, et al.: An analogue-dynamical long-range numerical weather prediction

1029 system incorporating historical evolution,Quart J Roy Meteor Soc, 119(511),547-565, 1993.

1030 Islam M.N. Sivakumar B.: Characterization and prediction of runoff dynamics:a nonlinear dynamical

1031 view. Advances in Water Resources, 25, 179-190, 2002.

1032 James A. Carton and Benjamin S. Giese: A Reanalysis of Ocean Climate Using Simple Ocean Data

1033 Assimilation (SODA),Monthly Weather Review,136(8),2999-3011 , 2008.

1034 Jin E. K.，James L. K.，Wang B.，et a1.: Current status of ENSO prediction skill in coupled

1035 ocean-atmosphere models,Climate Dyn,31, 647-664, 2008.

1036 Johnson S.D., Battisit D.S. andSarachik E. S.: Empirically Derived Markov Models and Prediction of

1037 Tropical Pacific Sea SurfaceTemperature Anomalies, Journal of Climate,13,3-17, 2000.

1038 Kalnay E., Kanamitsu M. and Kistler R.:   The NCEP/NCAR 40-year reanalysisproject, Bull. Amer.

1039 Meteor. Soc.,77,437-470 , 1996.

1040 Kathrin Büttner, Jennifer Salau, and Joachim Krieter: Temporal correlation coefficient for directed

1041 networks. Springerplus, 5(1): 1198-1203, 2016.

1042 Kim Ji-Won ,Soon-Il An,Sang-Yoon Jun,Hey-Jin Park,Sang-Wook Yeh.: ENSO and East Asian winter

1043 monsoon relationship modulation associated with the anomalous northwest Pacific anticyclone,

1044 Climate Dynamics, 49(4), 1157–1179, 2017.

1045 L'Heureux Michelle L., Collins Dan C., Hu Zeng-Zhen. Linear trends in sea surface temperature of the
1046 tropical Pacific Ocean and implications for the El Nin˜o-Southern Oscillation, Climate Dynamics, 40,
1047 1223–1236, 2013.

1048 Liebmann B. and C.A. Smith: Description of a Complete (Interpolated) Outgoing Longwave Radiation

Dataset,Bulletin of the American Meteorological Society,77,1275-1277, 1996.

Luo, J.-J., S. Masson, S. Behera, S. Shingu, andT. Yamagata: Seasonal climate predictabilityin a coupled OAGCM using a different approach forensemble forecasts,J. Climate, 18,4474–4497, 2005.

Mechoso C.R., Robertson A.W., Barth N., et al.: The seasonal cycle over the tropical Pacific in coupled atmosphere–ocean generalcirculation models,Mon Weather Rev,123,2825–2838, 1995.

Molteni F., et al.: ECMWF seasonal forecast system3,CLIVAR Exch,43,7-9, 2007.

Moore A. M., Zavala-Garay J. and Tang Y., et al.: Optimal forcing patterns for coupled models of ENSO,J Climate,19,4683-4699 , 2006.

Neelin J.D., Latif M. and Allaart M.A.F.: Tropical air-sea interaction in general circulation models,Clim Dyn.,7,73–104, 1992.

Nicosia V, Tang J, Mascolo C, Musolesi M, Russo G, Latora V: Graph metrics for temporal networks. In: Holme P, Saramäki J, editors. Temporal networks. Berlin: Springer, pp. 15–40, 2013,.

Palmer T. N., Alessandri A. and Andersen U., et al.: Development of a European multi-model ensemble system forseasonal to interannual prediction (DEMETER),Bull Amer Met Soc.,85,853-872 , 2004.

Philander S G., Pacanowski R.C., N-C Lau et al.: Simulation of ENSO with a global atmospheric GCM coupled to a high resolution, tropical Pacific Ocean GCM. *J.Climate, 5,*308-329,1992.

Qin G. H. and Li Z. H.: Over-fitting of BP NN research and its application, Engineering Journal of Wuhan University,39(6),1671-1679 , 2006.

Rasmusson E.M. and Carpenter T.H.: Variations in tropical seasurface temperature and surface wind fields associated with the Southern Oscillation/E1 Niño,Mon Weather Rev.,10, 354-384, 1982.

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res 108(D14):4407. doi:10.1029/2002JD002670, 2003.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C.Stokes, and W. Wang: An improved in situand satellite SST analysis for climate,J. Climate,15,1609–1625, 2002.

Saha S., Nadiga C. and Thiaw J., et al.: The NCEP climate forecast system,Journal of Climate,19 ,3483-3517 , 2006.

Saji N. H., Goswami B. N., V. inayachandran P. N., et al.: A dipole mode in the tropical Indian

Ocean,Nature, 401(6751),360-363, 1999.

Smith T.M.: Improved extended reconstruction of SST(1854-1997). J. Climate, 17, 2466-2477, 2004.

Takens, F.: Detecting strange attractors in fluid turbulence,Lecture Notes in Mathematics,898(2),361-381 , 1981.

Sivakumar B, Berndtsson R, Persson M.: Monthly Runoff Prediction Using Phase -space Reconstruction. Hydrological Sciences Journal, 46(3), 377 -388, 2001.

Sivakumar B., Jayawardena A.W., Fernando T.M.K.G.,: River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. Journal of Hydrology, 265, 225-245, 2002.

Timmermann A., Voss H. U. and Pasmanter R.: Empirical Dynamical System Modeling of ENSO Using Nonlinear Inverse Techniques, Journal of Physical Oceanography, 31,1579-1598 , 2001.

Tomita, T., and T. Yasunari: Role of the northeast winter monsoon on the biennial oscillation of the ENSO/monsoon system,J. Meteor. Soc. Japan, 74,399–413 , 1996.

Trenberth, E. K., et al.: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures,J. Geophys. Res., 107, C7, 14291-14324,1998.

Wang B., Wu R., Lukas R.: Roles of western North Pacific wind variation in thermocline adjustment and ENSO phase transition,J Meteor Soc Japan,77,1-16,1999a.

Wang B., Wu R., Li T.: Atmoshere-warm ocean interaction and its impacts on Asian-Australian monsoon variation. J. Climate, 16, 1195-1211, 2003.

Wang B.，Lee J. Y.，Shukla J.，et al.: Advance and prospectus of seasonal prediction：assessment of the APCC / CIiPAS 14-Model Ensemble Retrospective Seasonal Prediction(1980—2004),Climate Dyn.,33(1),93-117 , 2009a.

Wang C., Weisberg R. H. and Virmani J. I.: Western Pacific interannual variability associated with the El Niño-Southern Oscillation,J Geophy Res.,104,5131-5149, 1999b.

Wang, L., W. Chen, and R. H. Huang: Interdecadal modulation of PDO on the impact of ENSO on the east Asian winter monsoon, Geophys. Res. Lett., 35, L20702, doi:10.1029/2008GL035287, 2008.

Wang, W. C., K. W. Chau, C. T. Cheng, and L. Qiu: A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J. Hydrol., 374, 294–306, doi:10.1016/j.jhydrol.2009.06.019,2009b.

Wang L.: Intelligent Optimization Algorithms and Its Application, pp. 23-24, Tsinghua University Press, Chendu, 2001.

Webster P. J., Moore A. M., Loschnigg J. P., et al.: Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997- 98, Nature, 401(6751),356-360, 1999.

Weinberger K. Q. and L. Saul: Unsupervised learning of image manifolds by semidefinite programming,Int. J. Comput. Vision.,70, 77-90, 2006.

Xu B.C., Wang Z.S., Wu J.P. and Zhou E.M.: Interaction between sea surface temperature (SST) of information regions and southern oscillation index (SOI) in Tropical Pacific Ocean. *Marine Science Bulletin, 12(5)*,211-25,1993.

Yang, S., K. M. Lau, and K. M. Kim: Variations of the East Asian jet stream and Asian-Pacific-American winter climate anomalies, J. Climate, 15,306–325 , 2002.

Yang Se-Hwan and Lu Riyu: Predictability of the East Asian winter monsoon indices by the coupled models of ENSEMBLES, Advances in Atmospheric Sciences, 31(6), 1279–1292, 2014

Yim SY, Wang B, Kwon M: Interdecadal change of the controlling mechanisms for East Asian early summer rainfall variation around the mid-1990s.ClimDyn., 42,1325–1333, 2013.

Yim, S.-Y., B. Wang, W. Xing, M.-M.Lu: Prediction of Meiyu rainfall in Taiwan by multi-lead physicalempiricalmodels.Clim. Dyn., 44 (11-12), 3033-3042, doi:10.1007/s00382-014-2340-0, 2015.

Yoon, J., and S. W. Yeh: Influence of the Pacific Decadal Oscillation on the relationship between El Ni˜no and the northeast Asian summer monsoon, J. Climate,23, 4525–4537, 2010.

Yu H., J. Huang, and J. Chou: Improvement of Medium-Range Forecasts Using the Analogue-Dynamical Method,Mon. Wea. Rev., 142, 1570–1587, doi: http://dx.doi.org/10.1175/MWR-D-13-00250.1, 2014a.

Yu H., J. Huang, W. Li, and G. Feng: Development of the analogue-dynamical method for error correction of numerical forecasts,J. Meteor. Res., 28(5), 934–947, doi: 10.1007/s13351-014-4077-4 , 2014b.

Zhang R. and Hong M., et al.: Non-linear Dynamic Model Retrieval of Subtropical High Based on Empirical Orthogonal Function and Genetic Algorithm,Applied Mathematics and Mechanics,27(12),1645-1654, 2006.

Zhang R. and Hong M.,et al.: Retrieval of the non-linear dynamic forecast model of El Nino/La Nina index based on the genetic algorithm optimization. Chinese Journal of Geophysics,51(5),1354-1362, 2008.

Zhang R. H., Zhou G. Q. and Chao J. P.: ENSO Dynamics and Its Prediction,Chinese Journal of

Atmospheric Sciences,27(4) ,674-688, 2003a.

Zhang, R.H., S. E. Zebiak, R. Kleeman, and N.Keenlyside: A new intermediate coupled model for El Niño simulation and prediction. Geophys. Res.Lett., 30, doi:10.1029/2003GL018010, 2003b.

Zhang, R. H., A. Sumi, and M. Kimoto: Impact of El Ni˜noon the East Asian monsoon: A diagnostic study of the '86/87 and '91/92 events,J. Meteor. Soc. Japan, 74, 49–62, 1996.

Zhang Y. L., Yu Y. Q., Duan W. S.: The spring prediction barrier of ENSO in retrospective prediction experiments as shown by the four coupled ocean-atmosphere models. Acta Meteorologica Sinica, 70(3), 506-519, 2012.

Zhao J. H., Liu X. Y. and Jiang H. Y., et al.: Characteristics of Sea Surface Height in Tropical Pacific and its relationship with ENSO events,Meteorological and Environmental Sciences, 35(2),33-39, 2012.

Zhou, L.-T., and R. G. Wu: Respective impacts of the East Asian winter monsoon and ENSO on winter rainfall in China,J. Geophys. Res.,115, doi: 10.1029/2009JD012502, 2010.

REFERENCES

Ashok K, Guan Z, Yamagata T : Impact of the Indian Ocean Dipole on the decadal relationship between the Indian monsoon rainfall and ENSO, Geophys Res Let,28(23), 4499-4502, 2001.

Balmaseda M.A., Davey M.K. and Anderson D.L.T.: Decadal and seasonal dependence of ENSO prediction skill,J Clim.,8, 2705–2715, 1995.

Barber R.T. and Chavez F.P.: Biological consequences of E1Niño,Science,222,1203-1210, 1983.

Barnston A. G., et al.: Skill of real-time seasonal ENSO model predictions during 2002-2011,Bull. Amer. Meteor. Soc.,93, 631-651, 2012.

Belkin M. and P. Niyogi: Laplacian eigenmaps for dimensionality reduction and data representation,Netural Comput.,15,1373-1391, 2003.

Bjerknes J.: Atmsopheric teleconnections from the equtorail Pacific,Mon. Wea. Rev.,97,163-172, 1969.

Cao H. X.: Self-memorization Equation in Atmospheric Motion,Science in China (Series B),36(7), 845-855, 1993.

Chen D., S. E. Zebiak, A. J. Busalacchi and M. A. Cane: An Improved Procedure for El Nirio Forecasting: Implications for Predictability,Science, 269, 1699-1702, 1995.

Chen X. D., Xia J., Xu Q.: Differential Hydrological Grey Model(DHGM) with self-memory function and its application to flood forecasting,Sci China Tech Sci.,52,1039- 1049, 2009.

Chou J.: The problem of utilizing past data in numerical weather forecasting,Sci. China, 6, 635-644, 1974 (in Chinese).

Clarke A. J. and S. Van Gorder: Improving El Niño prediction using a space-time integration of Indo-Pacific winds and equatorial Pacific upper ocean heat content,Geophys. Res. Lett.,30,1399. doi:10.1029/2002GL016673, 2003.

Delecluse P., Davey M., Kitamura Y., Philander S., Suarez M., Bengtsson L.: TOGA review paper: coupled general circulation modeling of the tropical Pacific,J Geophys Res,103,14357-14373, 1998.

Davey M., Huddleston M., Sperber K.R., et al.: A study of coupled model climatology and variability in tropical ocean regions,Clim. Dyn.,18,403-420, 2002.

Dommenget and Latif: A Cautionary Note on the Interpretation of EOFs, Journal of Climate,15(2),216-225, 2002.

Drosdowsky W.: Statistical prediction of ENSO (Niño 3) using sub-surface temperature data,Geophys. Res.Lett., 33 , L03710. doi:10.1029/2005GL024866, 2006.

Feng G. L., Cao H. X., Gao X. Q., et al.: Prediction of precipitation during summer monsoon with self-memorial model,Adv Atmos Sci.,18 ,701-709, 2001.

Fraedrich K.: Estimating weather and climate predictability on attractors,J .A tmos.Sci.,44,7 22-728, 1987.

Glantz MH, Katz RW, Nicholls N (eds): Teleconnections linking worldwide climate anomalies, 74pp,Cambridge University Press, Cambrige, UK, 1991.

Gu X. Q.: A spectral model based on atmospheric self_memorization principle,Chinese Science Bulletin,43(20),1692-1702 , 1998.

Hong M., Zhang R.andMa C. C.et al.: A Non-Linear Dynamical-Statistical Model for Reconstruction of the Air-Sea Element Fields in the Tropical Pacific Ocean,Atmosphere-Ocean, doi: 10.1080/07055900.2014.908765, 2014.

Hu, T.S., K.C. Lam, and S.T. Ng: River flow time series prediction with a range-dependent neural network,Hydrol. Sci. J., 46, 729-745, 2001.

Huang. J., Y. Yi, S. Wang, et al.: An analogue dynamical long-range numerical weather prediction system incorporating historical evolution,Quart J Roy Meteor Soc, 119(511),547-565, 1993.

Huang J. P. and Yi Y. H.: A Non-linear Dynamic System Reconstructing of Actual data,Science in China,3 (3),331-336 , 1991.

James A. Carton and Benjamin S. Giese: A Reanalysis of Ocean Climate Using Simple Ocean Data Assimilation (SODA),Monthly Weather Review,136(8),2999-3011 , 2008.

Jin E. K., James L. K., Wang B., et al.: Current status of ENSO prediction skill in coupled ocean-atmosphere models,Climate Dyn,31, 647-664, 2008.

Johnson S.D., Battisit D.S. andSarachik E. S.: Empirically Derived Markov Models and Prediction of Tropical Pacific Sea SurfaceTemperature Anomalies, Journal of Climate,13,3-17, 2000.

Kalnay E., Kanamitsu M. and Kistler R.: The NCEP/NCAR 40 year reanalysisproject, Bull. Amer. Meteor. Soc.,77,437-470 , 1996.

Liao D., Zhou Y.H. and Liao X.H.: Modulation of the SSTA Decadal Variation on ENSO Events amd Relationships of SSTA with LOD, SOI, etc.,Acta Astronomica Sinica,48(1),36-47 , 2007.

Liebmann B. and C.A. Smith: Description of a Complete (Interpolated) Outgoing Longwave Radiation Dataset,Bulletin of the American Meteorological Society,77,1275-1277, 1996.

Li L. P., Wang P. X., He J. H. and Wang D. X.: Analysis of interdecadal and interannual Characteristics of the Tropical western Pacific Warm Pool heat status,Journal of Tropical Meteorology,20(5),472-482 , 2004.

Luo, J. J., S. Masson, S. Behera, S. Shingu, andT. Yamagata: Seasonal climate predictabilityin a coupled OAGCM using a different approach forensemble forecasts,J. Climate, 18,4474-4497, 2005.

Mechoso C.R., Robertson A.W., Barth N., et al.: The seasonal cycle over the tropical Pacific in coupled atmosphere-ocean generalcirculation models,Mon Weather Rev,123,2825-2838, 1995.

Molteni F., et al.: ECMWF seasonal forecast system3,CLIVAR Exch,43,7-9, 2007.

Moore A. M., Zavala Garay J. and Tang Y., et al.: Optimal forcing patterns for coupled models of ENSO,J Climate,19,4683-4699 , 2006.

Neelin J.D., Latif M. and Allaart M.A.F.: Tropical air-sea interaction in general circulation models,Clim Dyn.,7,73-104, 1992.

Palmer T. N., Alessandri A. and Andersen U., et al.: Development of a European multi-model ensemble system forseasonal to interannual prediction (DEMETER),Bull Amer Met Soc.,85,853-872 , 2004.

Panofsky H.A., Brier G.W.: Some applications of statistics tometeorology, Pennsylvania State University Press, Pennsylvania , 1968.

Rasmusson E.M. and Carpenter T.H.: Variations in tropical seasurface temperature and surface wind

fields associated with the Southern Oscillation/El Niño,Mon Weather Rev.,10, 354-384, 1982.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C.Stokes, and W. Wang: An improved in situand satellite SST analysis for climate,J. Climate,15,1609-1625, 2002.

Saha S., Nadiga C. and Thiaw J., et al.: The NCEP climate forecast system,Journal of Climate,19 ,3483-3517 , 2006.

Sa ji N. H., Goswami B. N., V. inayachandran P. N., et al.: A dipole mode in the tropical Indian Ocean,Nature, 401(6751),360-363, 1999.

Takens, F.: Detecting strange attractors in fluid turbulence,Lecture Notes in Mathematics,898(2),361-381 , 1981.

Tetko, I. V., Livingstone, D. J., Luik, A. I.: Neural network studies. 1. Comparison of Overfitting and Overtraining, J. Chem. Inf. Comput. Sci.,35 (5), 826-833 , 1995.

Timmermann A., Voss H. U. and Pasmanter R.: Empirical Dynamical System Modeling of ENSO Using Nonlinear Inverse Techniques, Journal of Physical Oceanography, 31,1579-1598 , 2001.

Tomita, T., and T. Yasunari: Role of the northeast winter monsoon on the biennial oscillation of the ENSO/monsoon system,J. Meteor. Soc. Japan, 74,399-413 , 1996.

Trenberth, E. K., et al.: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures,J. Geophys. Res., 107, C7, 14291-14324,1998.

Wang B., Wu R., Lukas R.: Roles of western North Pacific wind variation in thermocline adjustment and ENSO phase transition,J Meteor Soc Japan,77,1-16,1999a.

Wang B., Lee J. Y., Shukla J., et al.: Advance and prospectus of seasonal prediction: assessment of the APCC / CliPAS 14-Model Ensemble Retrospective Seasonal Prediction(1980 - 2004),Climate Dyn.,33(1),93-117 , 2009.

Wang C., Weisberg R. H. and Virmani J. I.: Western Pacific interannual variability associated with the El Niño-Southern Oscillation,J Geophy Res.,104,5131-5149, 1999b.

Wang, L., W. Chen, and R. H. Huang: Interdecadal modulation of PDO on the impact of ENSO on the eastAsian winter monsoon,Geophys. Res. Lett., 35, L20702, doi:10.1029/2008GL035287, 2008.

Wang L.: Intelligent Optimization Algorithms and Its Application, pp. 23-24, Tsinghua University Press, Chendu, 2001.

Wang W., Su J. Y., Hou B. W. et al.: Dynamic prediction of building subsidence deformation with data-based mechanistic self-memory model,Chinese Science Bulletin,57(26),3430-3435, 2012.

Webster P. J., Moore A. M., Loschnigg J. P., et al.: Coupled ocean atmosphere dynamics in the Indian Ocean during 1997-98, Nature, 401(6751),356-360, 1999.

Weinberger K. Q. and L. Saul: Unsupervised learning of image manifolds by semidefinite programming,Int. J. Comput. Vision.,70, 77-90, 2006.

Xu J. J. and Wang D. X.: Diagnosis of interannual and interdecadal variation in SST over Indian-Pacific Ocean and numerical simuation of their effect on Asian summer monsoon,Acta Oceanologica Sinica,22(3),34-43 , 2000.

Yang, S., K. M. Lau, and K. M. Kim: Variations of the East Asian jet stream and Asian Pacific American winter climate anomalies, J. Climate, 15,306-325 , 2002.

Yim SY, Wang B, Kwon M: Interdecadal change of the controlling mechanisms for East Asian early summer rainfall variation around the mid-1990s.ClimDyn., 42,1325-1333, 2013.

Yim, S. Y., B. Wang, W. Xing, M.-M.Lu: Prediction of Meiyu rainfall in Taiwan by multi-lead physicalempiricalmodels.Clim. Dyn., 44 (11-12), 3033-3042, doi:10.1007/s00382-014-2340-0, 2015.

Yoon, J., and S. W. Yeh: Influence of the Pacific Decadal Oscillation on the relationship between El Ni~no and the northeast Asian summer monsoon, J. Climate,23, 4525-4537, 2010.

Yu H., J. Huang, and J. Chou: Improvement of Medium Range Forecasts Using the Analogue Dynamical Method,Mon. Wea. Rev., 142, 1570-1587, doi: http://dx.doi.org/10.1175/MWR-D-13-00250.1, 2014a.

Yu H., J. Huang, W. Li, and G. Feng: Development of the analogue dynamical method for error correction of numerical forecasts,J. Meteor. Res., 28(5), 934-947, doi: 10.1007/s13351-014-4077-4 , 2014b.

Zhang R. and Hong M., et al.: Non-linear Dynamic Model Retrieval of Subtropical High Based on Empirical Orthogonal Function and Genetic Algorithm,Applied Mathematics and Mechanics,27(12),1645-1654, 2006.

Zhang R. and Hong M.,et al.: Retrieval of the non-linear dynamic forecast model of El Nino/La Nina index based on the genetic algorithm optimization. Chinese Journal of Geophysics,51(5),1354-1362, 2008.

Zhang R. H., Zhou G. Q. and Chao J. P.: ENSO Dynamics and Its Prediction,Chinese Journal of Atmospheric Sciences,27(4) ,674-688, 2003.

Zhang, R. H., A. Sumi, and M. Kimoto: Impact of El Ni~noon the East Asian monsoon: A diagnostic

study of the '86/87 and '91/92 events,J. Meteor. Soc. Japan, 74, 49—62, 1996.

Zhao J. H., Liu X. Y. and Jiang H. Y., et al.: Characteristics of Sea Surface Height in Tropical Pacific

and its relationship with ENSO events,Meteorological and Environmental Sciences, 35(2),33-39, 2012.

Zhou, L. T., and R. G. Wu: Respective impacts of the East Asian winter monsoon and ENSO on winter

rainfall in China,J. Geophys. Res.,115, doi: 10.1029/2009JD012502, 2010.

List of Figures:

Fig.1(a, c) First and second modes of the EOF deconstruction of the SSTA field, and (b, d) the corresponding PC time series.

The time series (a) and the spatial mode (b) of the first mode; the time series(c) and the spatial mode (d) of the second mode of the SSTA filed

Fig. 2 Forecast results of the first time coefficient series (a) and the second time coefficient series (b) of the SSTA field by the original model

Fig. 3. The cross-validated retroactive hindcast results of the first time coefficient series (a) and the second time coefficient series (b) of the SSTA field by the original model

Fig. 4. Long-term step-by-step forecast results of the first time coefficient series (a) and the second time coefficient series (b) of the SSTA field by the improved model

Fig. 5. The cross-validated retroactive hindcast results of the first time coefficient series (a) and the second time coefficient series (b) of the SSTA field by the improved model

Fig. 6. The forecast SSTA field (a) and the actual SSTA field (b) of an El Niño event (Dec.1997)

Fig. 7. The forecast SSTA field (a) and the actual SSTA field (b) of a La Niña event (Dec.1999)

Fig. 8. The forecast SSTA field (a) and the actual SSTA field (b) of neutral event (Nov.2002)

Fig. 9. The improved dynamical-statistical model prediction of the ENSO index

Fig.10. The cumulative correlation coefficients (a) and cumulative mean absolute percentage error (b) changing with time of different lead times

Fig. 1110. Temporal correlation between model forecasts and observations for all seasons combined, as

1326 a function of lead time. Each line highlights one model.

1327 **Fig.~~12~~11.** RMSE in standardized units, as a function of lead time for all seasons combined. Each line

1328 highlights one model.

1329

1330

1331

1332

1333

1334

1335

1336

1337 **Table captions:**

1338 **Table 1.** The correlation analysis between the front two time series T_1, T_2 and nine impact factors

1339 ~~Table 1. Forecast results of models of different variables~~

1340 **Table2.** The CC and MAPE of long-term fitting test when the retrospective order p is different

1341 ~~Table 2. The correlation coefficient (CC) and Mean absolute percentage error (MAPE) of long-term~~
1342 ~~fitting test when the retrospective order p is different~~

1343 **Table3.** The forecast results of T_1 and T_2 in different examples within 6 and 12 months

1344 **Table. 4.** The TC and the MAPE between model forecasts and observations within 12 months for
1345 Nov.–Jan., Dec.–Feb., and Jan.–Mar. as lead time of winter, for Feb.–Apr. , Mar.–May and Apr.–June as
1346 lead time of spring, for May-July, June-August and July-Sep. as lead time of summer and for
1347 August-Oct., Sep.-Nov. and Oct.-Dec. as lead time of autumn.
1348 ~~Table4. Temporal correlation (CC) and the mean absolute percentage error (MAPE) between model~~
1349 ~~forecasts and observations within 12 months for November–January , December–February, and~~
1350 ~~January–March as lead time of winter and for May-July, June-August and July-Sep. as lead time of~~
1351 ~~summer.~~
1352 **Table5.** The forecast results of the different data periods~~Table 5. The correlation coefficients among~~
1353 ~~four factors~~

1354

1355 **Figure:**

58
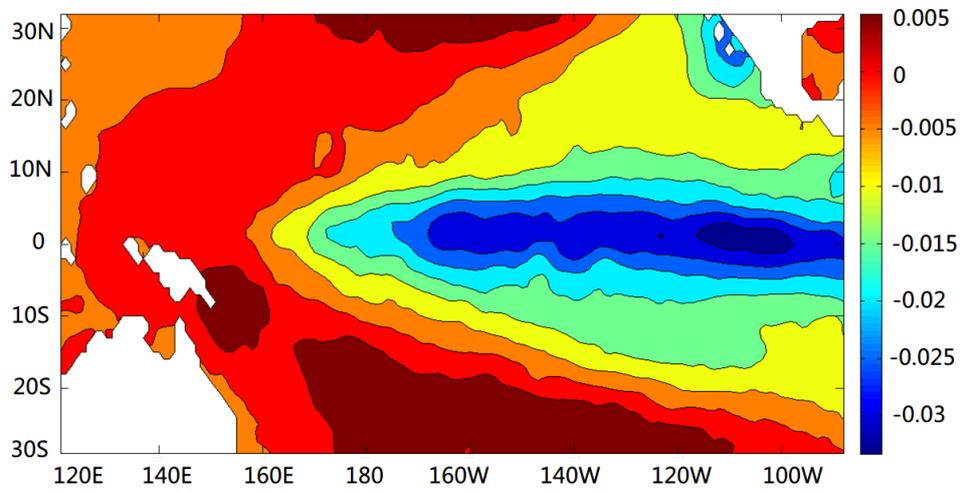
1356

1357

(a)
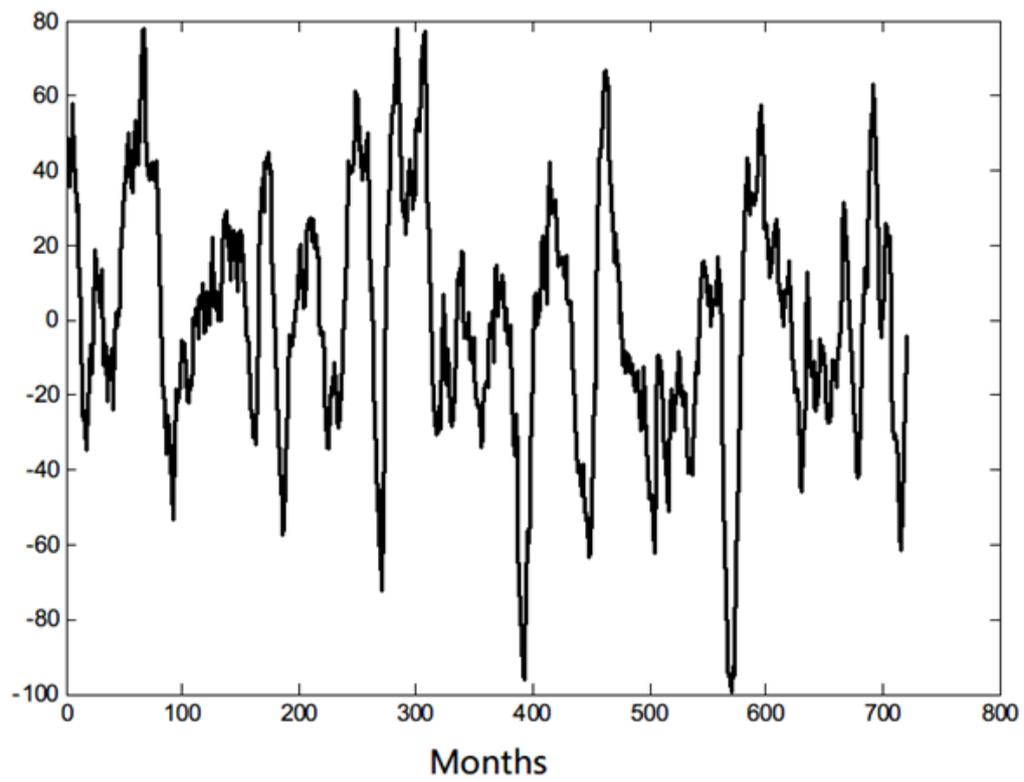
1358

1359

(b)
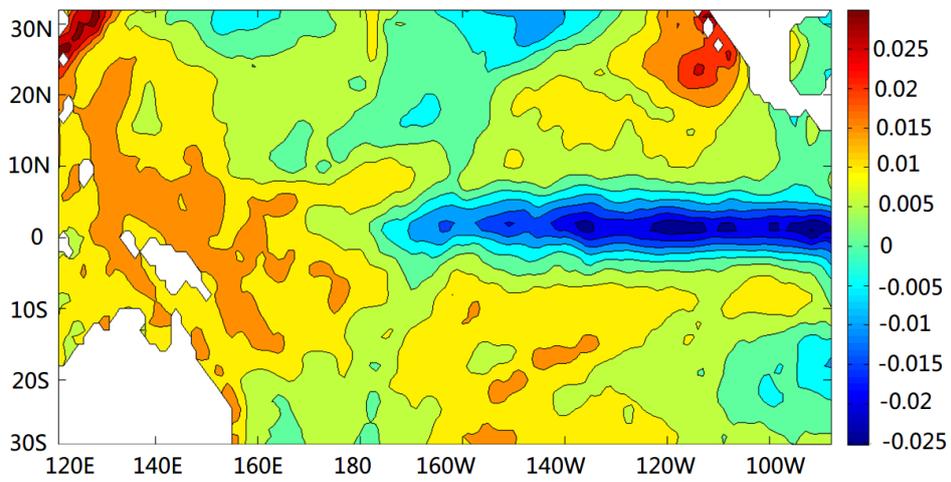
1360

1361   (c)
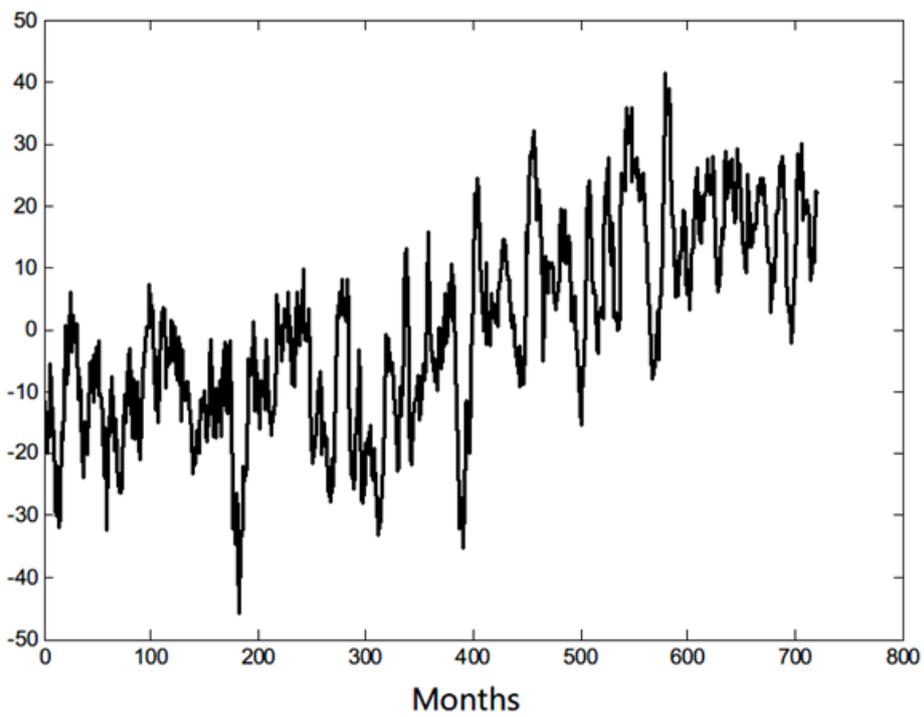

1362

1363   (d)

1364   **Fig. 1** (a, c) First and second modes of the EOF deconstruction of the SSTA field, and (b, d) the

1365   corresponding PC time series.
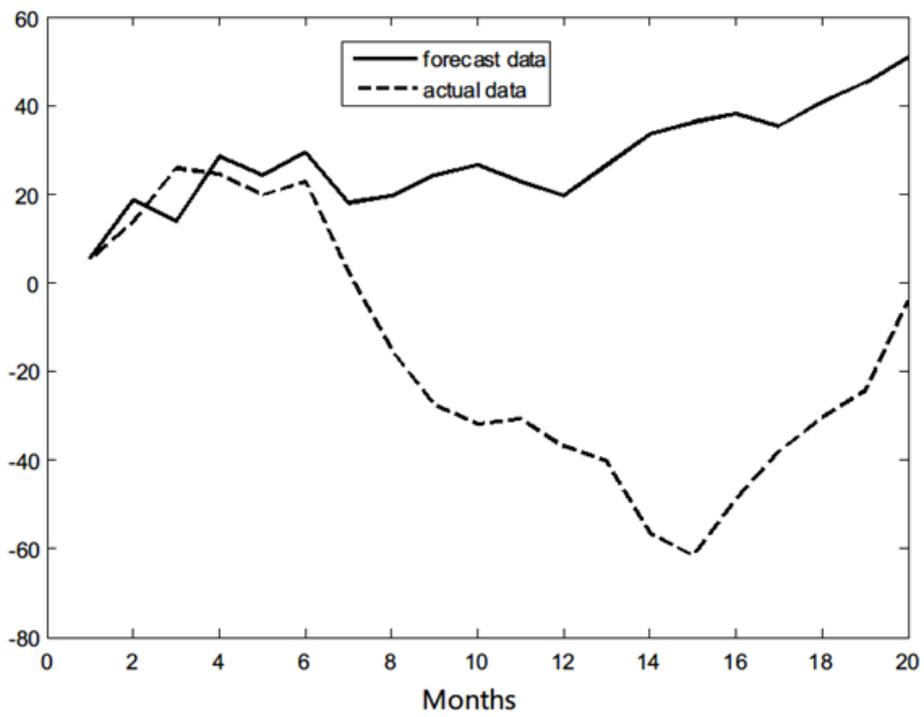
1366

1367   (a)
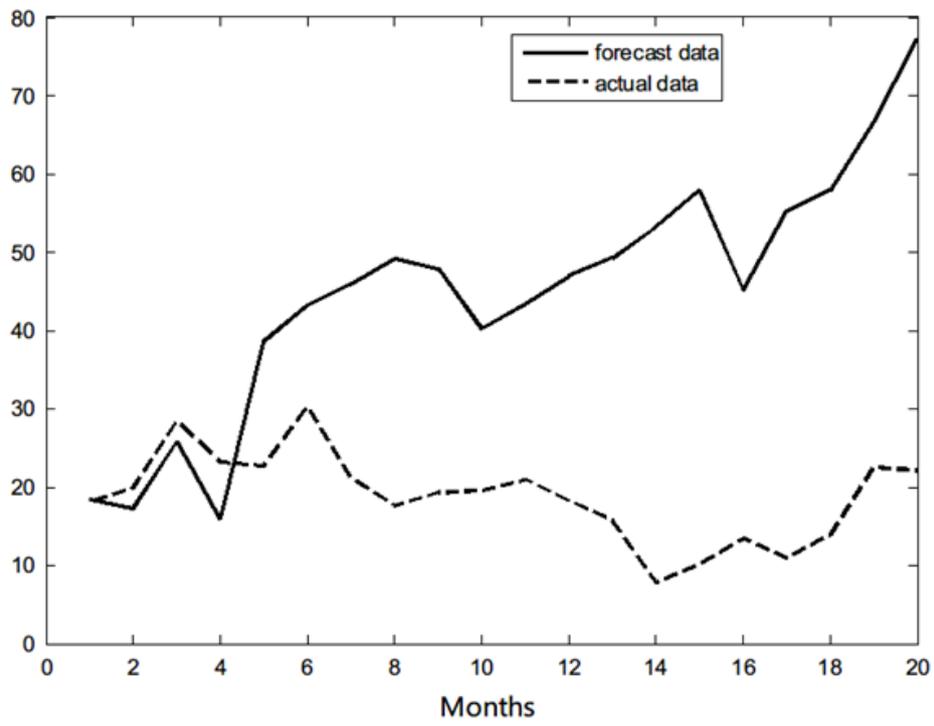
1368

1369   (b)

1370   Fig.2 Forecast results of the first time coefficient series T_1 错误!未找到引用源。 (a) and the second

1371   time coefficient series T_2 ~~T_2~~ (b)of the SSTA field by the original model

1372

1373

1374

1375

1376

1377

1378

1379

1380

1381

1382
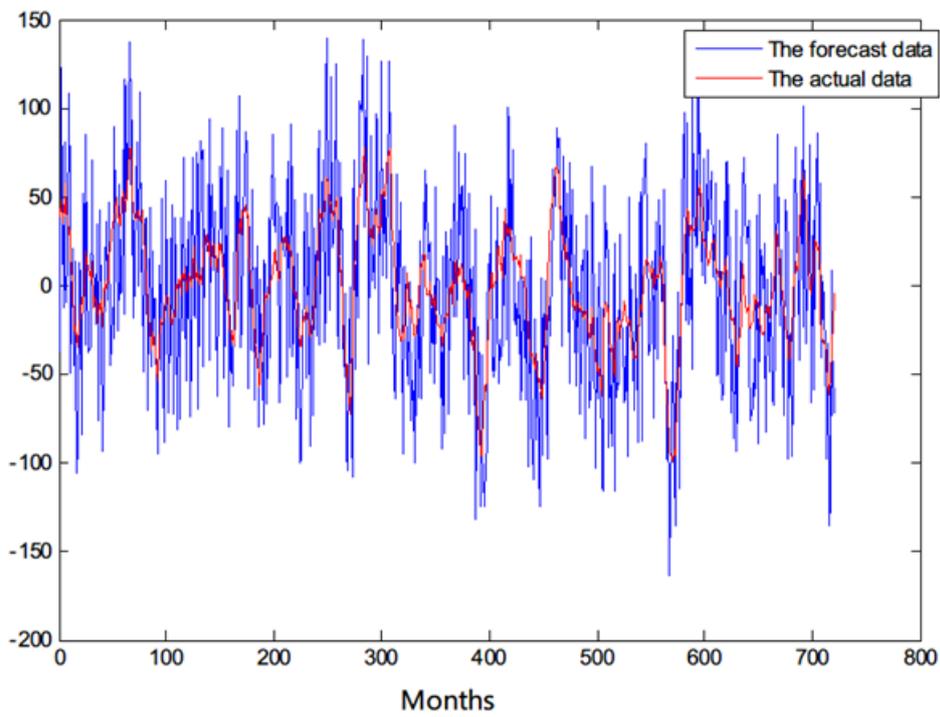
1383

1384   (a)
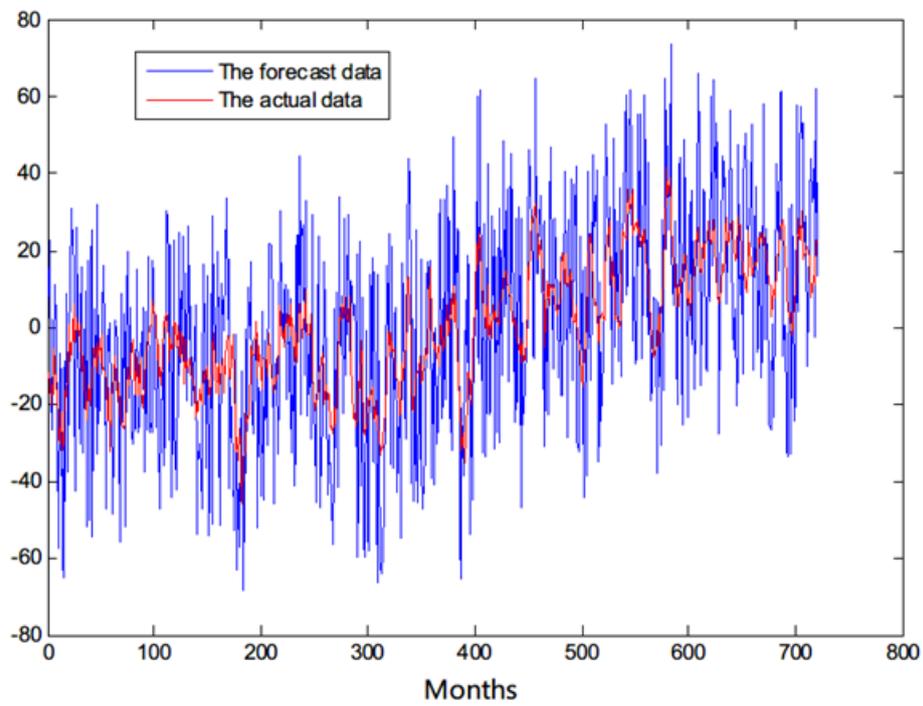
1385

1386   (b)

1387   Fig.3The cross-validated retroactive hindcast results of the first time coefficient series T_1 $\overline{T_1}$ (a)and

1388   the second time coefficient series T_2 $\overline{T_2}$ (b)of the SSTA field by the original model

1389

1390

1391

1392

1393

1394

1395

1396

1397
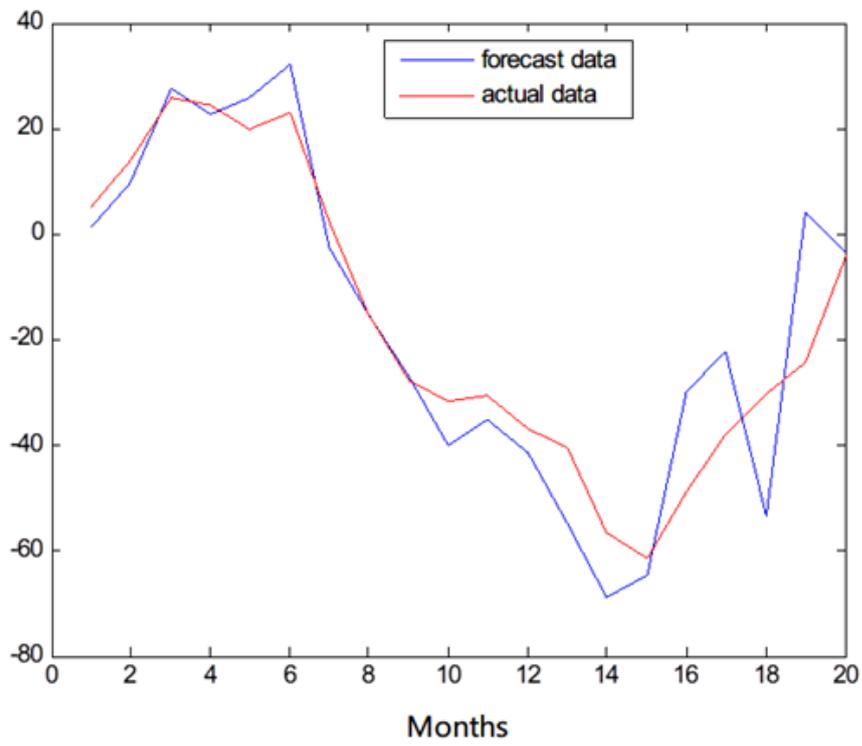
1398

1399   (a)
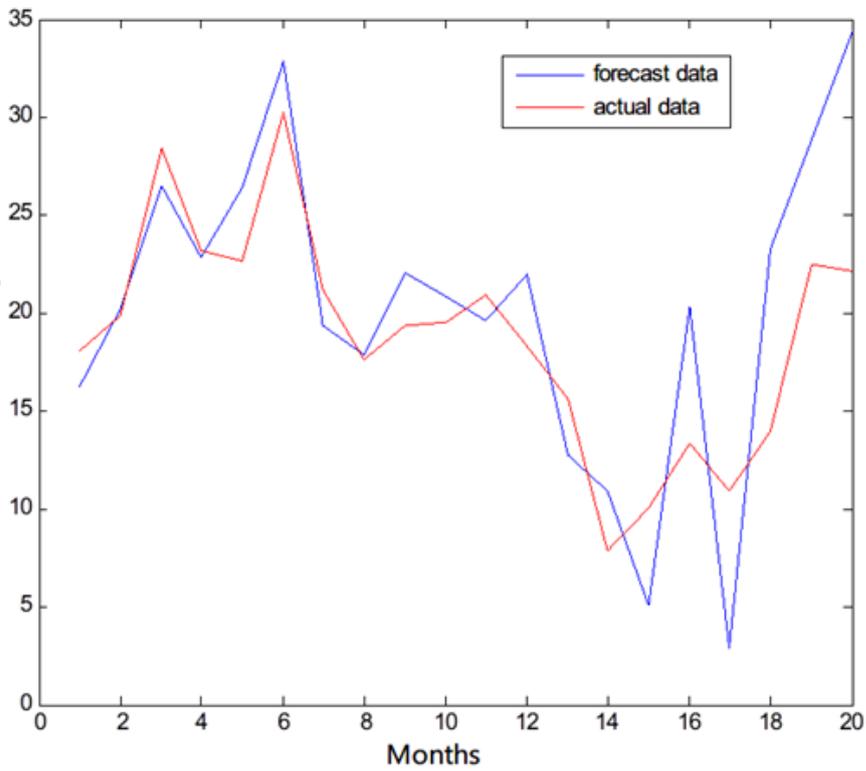
1400

1401   (b)

Fig. 4. Long-term step-by-step forecast results of the first time coefficient series $T_1 \mathcal{T}_1$ (a)and the

second time coefficient series $T_2 \mathcal{T}_2$ (b)of the SSTA field by the improved model

1404
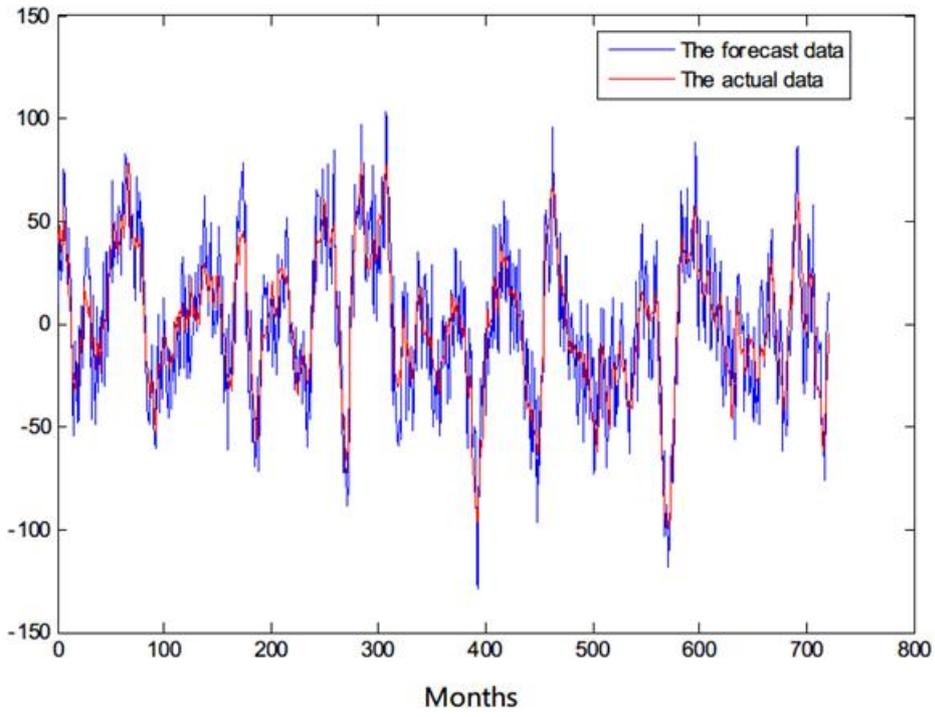
1405

1406   (a)
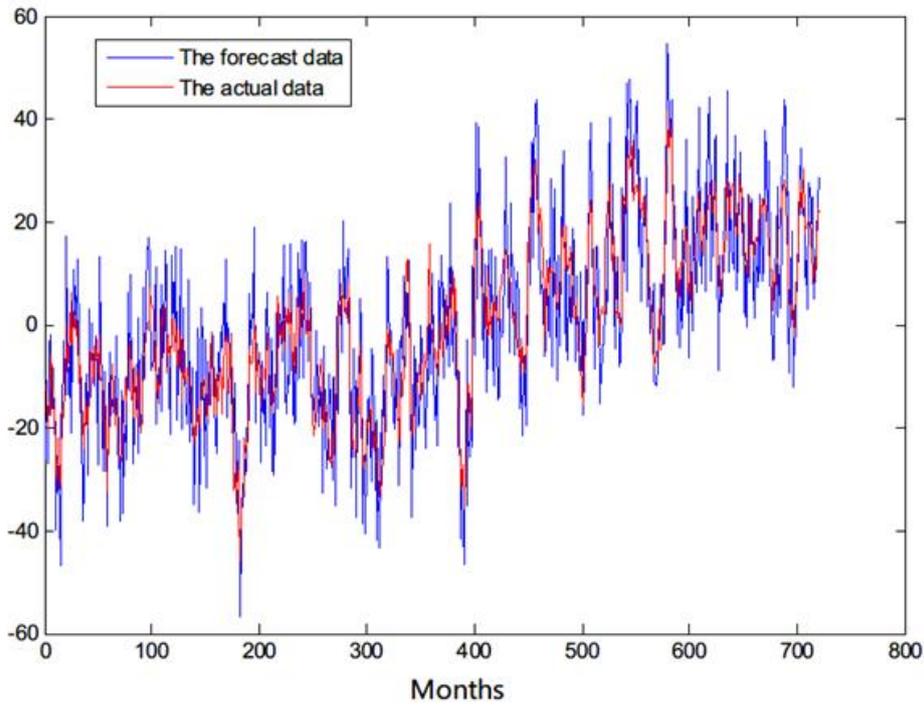
1407

1408
(b)

1409 Fig. 5. The cross-validated retroactive hindcast results of the first time coefficient series T_1 —(a)and

1410 the second time coefficient series T_2 —(b)of the SSTA field by the improved model

1411

1412

1413

1414

1415

1416

1417

1418
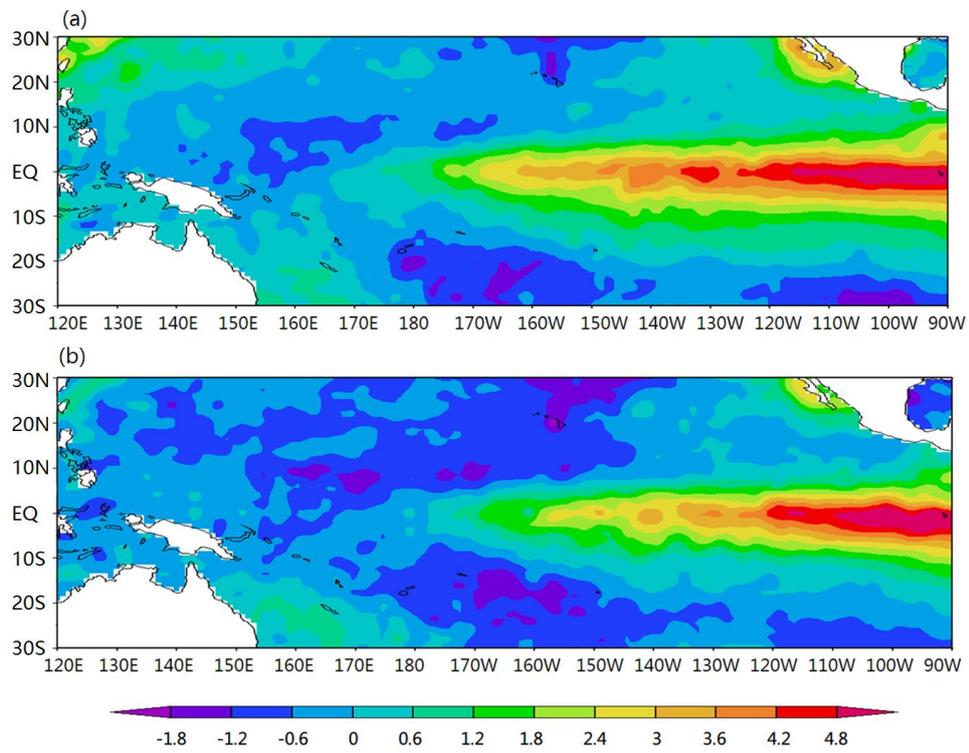
1419

1420   Fig.6. The forecast SSTA field(a) and the actual SSTA field (b)of an El Ni ño event (Dec.1997)

1421

1422

1423

1424

1425

1426

1427

1428
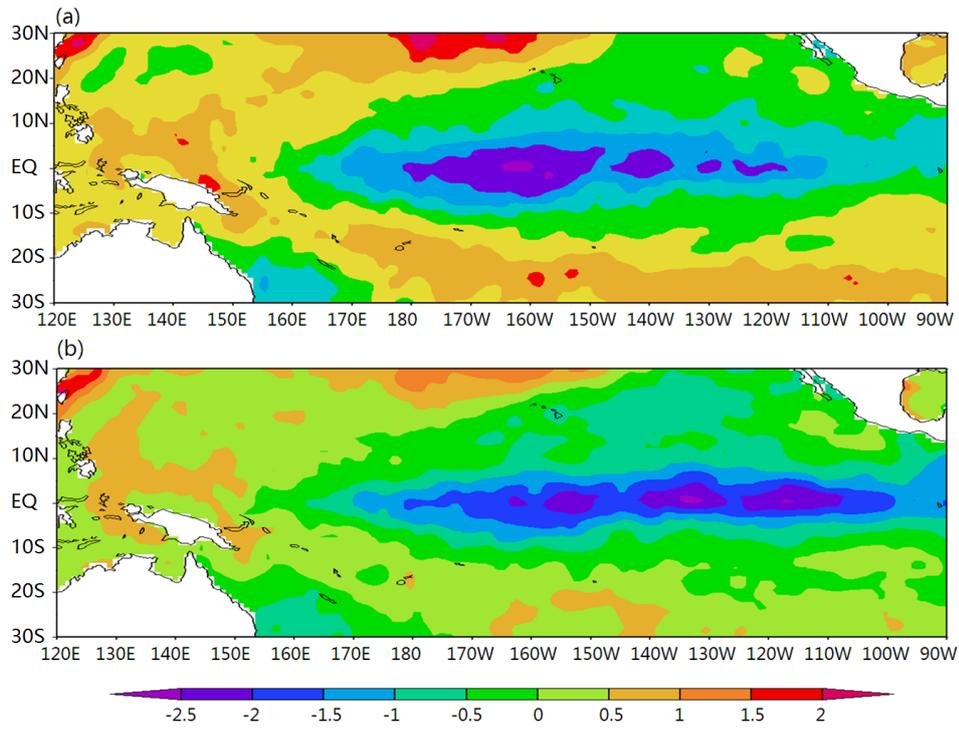
1432   Fig.7. The forecast SSTA field(a) and the actual SSTA field (b)of a La Niña event (Dec.1999)

1433

1434

1435

1436

1437

1438

1439

1442
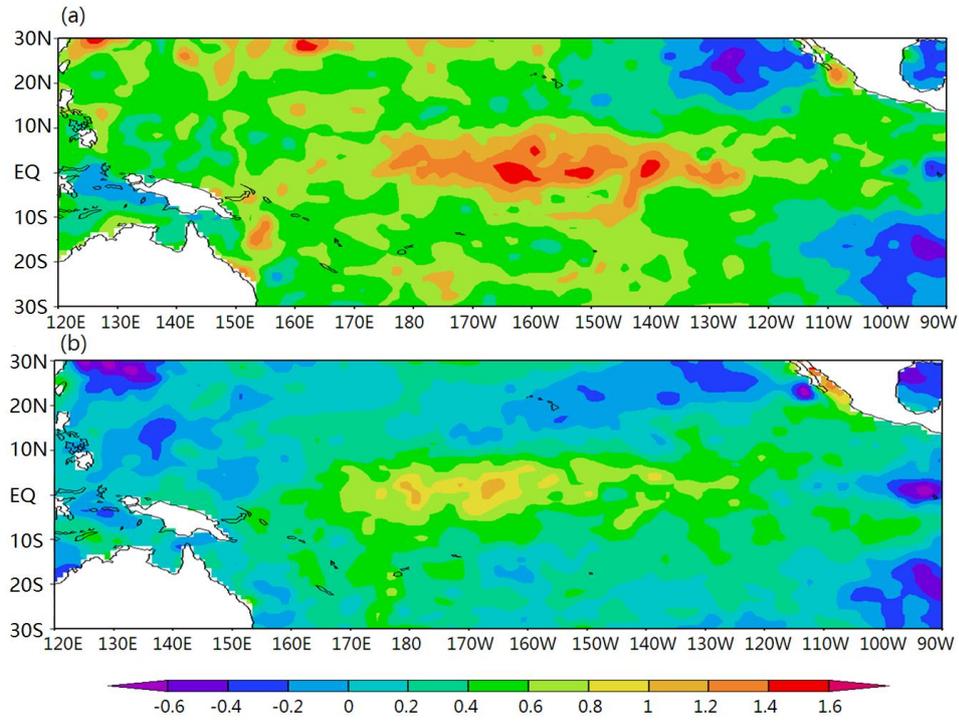
Fig.8. The forecast SSTA field(a) and the actual SSTA field (b)of neutral event (Nov.2002)

1444
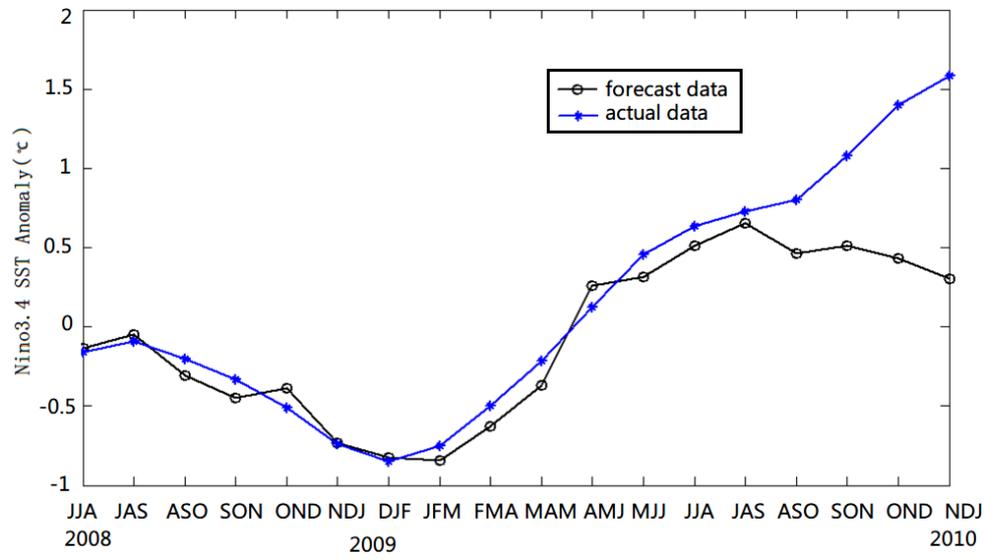
Fig.9. The improved dynamical-statistical model prediction of the ENSO index
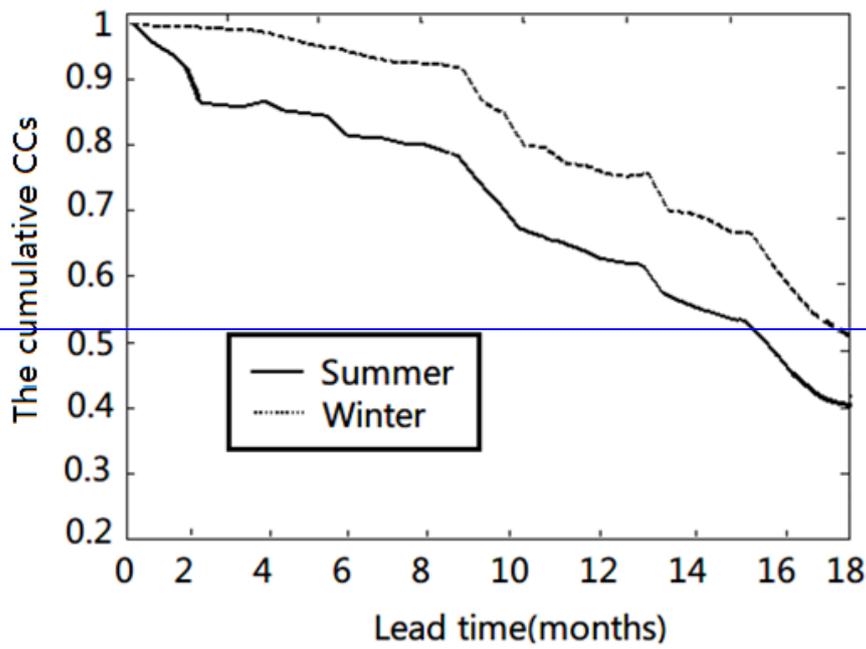
1453

1454   (a)
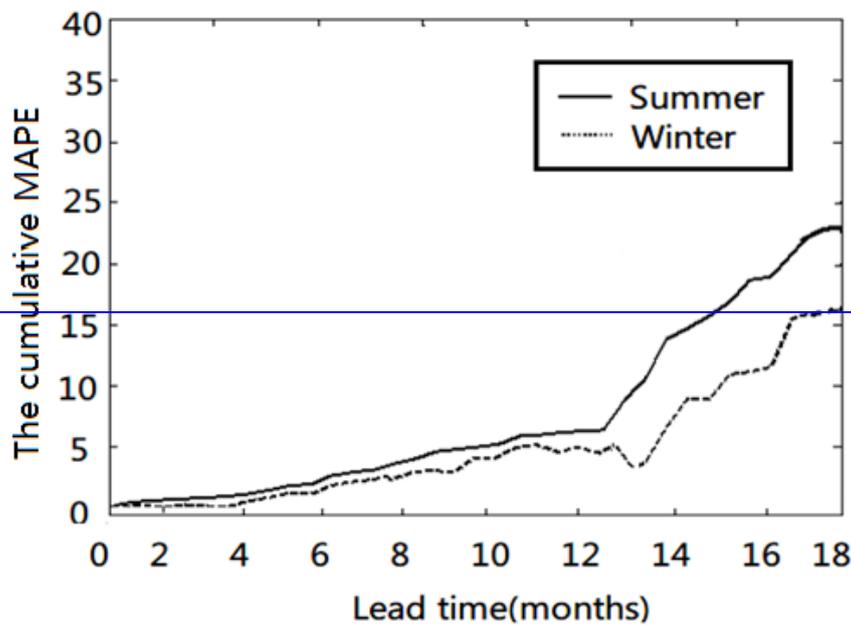
1455

1456   (b)

1457 Fig.10. The cumulative correlation coefficients(CCs) (a) and cumulative mean absolute percentage

1458 error(MAPE) (b) changing with time of different lead times
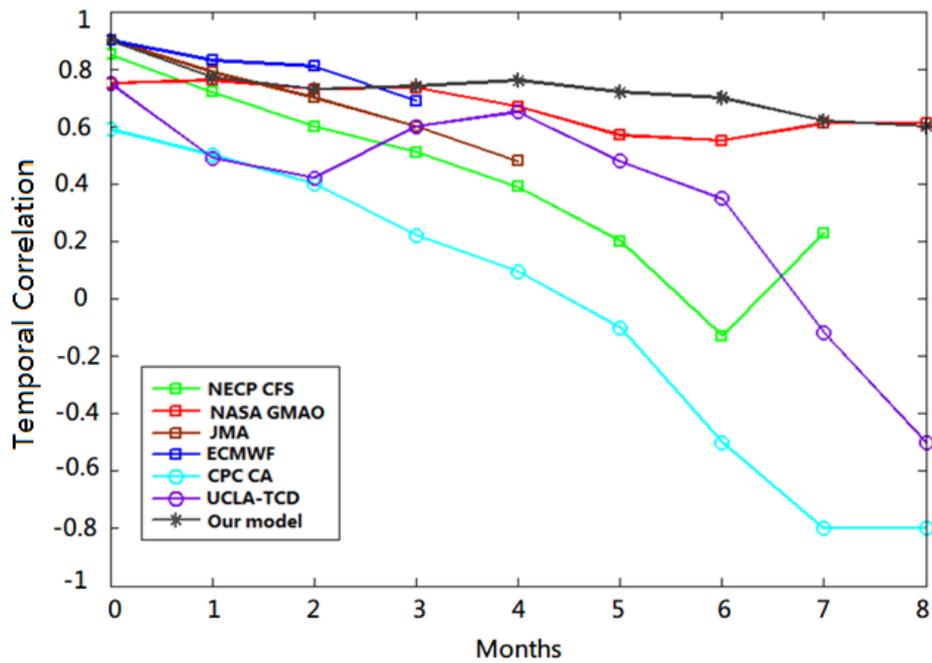
1461 Fig. 11 10. Temporal correlation between model forecasts and observations for all seasons combined, as

1462 a function of lead time. Each line highlights one model.

1463

1464

1465

1466

1467

1468

1469

1470

1471

1472

1473

1474

1475

1476

1477
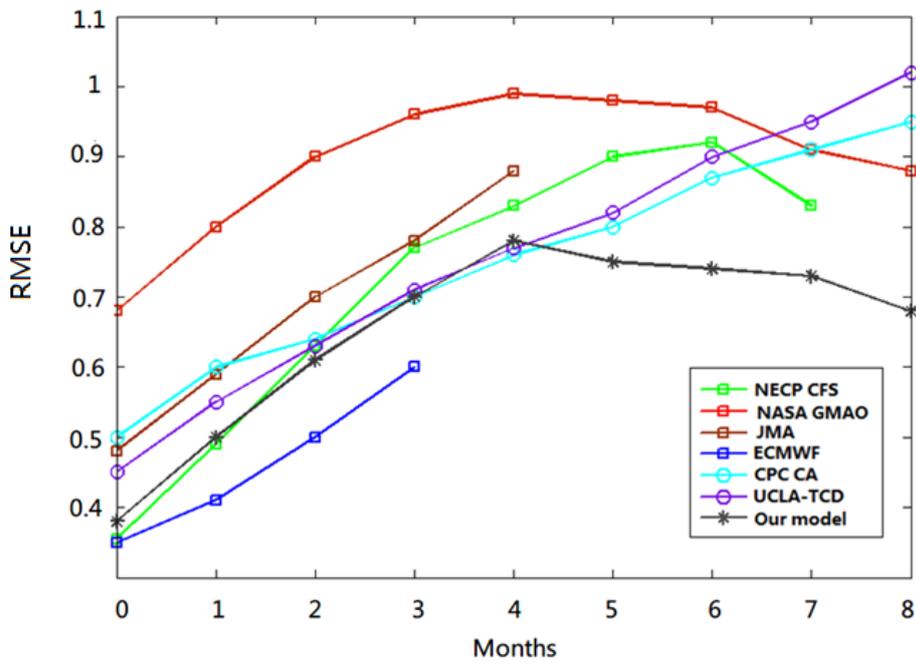
1478

Fig . ~~12~~11. RMSE in standardized units, as a function of lead time for all seasons combined. Each line

highlights one model.

1481

1482

1483

1484

1485

1486

1487

1488

1489

1490

1491

1492

1493

Table:

Table 1. The correlation analysis between the front two time series T_1, T_2 and nine impact factors

| factors	u_1	u_2	PNA	DMI	SOI	PDOI	EAWMI	OLR	SSH
T_1	0.3161	0.5684	0.4386	-0.3457	0.7734	0.4081	0.6284	0.3287	0.3363
T_2	0.2118	0.4181	0.2560	-0.2345	0.5232	0.3065	0.4825	0.1816	0.2169

Table1. The forecast results of the models of different variables

The model	The forecast skill of 60 cross-validated retroactive hindcasts experiments of the ENSO index for all seasons combined at lead times of 8 months	
	the temporal correlation	the root mean square error
One variable（T_1）	0.5051	0.8075
Two variables（T_1, T_2）	0.5613	0.7679
Three variables（T_1, T_2, SOI）	0.6027	0.7275
Four variables（$T_1, T_2, SOI, EAWMI$）	0.6344	0.6728
Five variables（$T_1, T_2, SOI, EAWMI, u_1$）	0.5923	0.7344
Six variables（$T_1, T_2, SOI, EAWMI, u_1, PNA$）	0.5528	0.7806

1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531

1532 **Table2.**The ~~correlation coefficient~~(CC_~~)~~ and ~~Mean absolute percentage error~~(MAPE~~)~~ of long-term

1533 fitting test when the retrospective order p is different

p		4	5	6	7	8	9	10
The forecast results of long-term fitting test	CC	0.75	0.73	0.81	0.74	0.70	0.72	0.68
	MAPE	18.42%	19.36%	14.56%	20.39%	25.31%	24.18%	27.33%
p		11	12	13	14	15	16	
The forecast results of long-term fitting test	CC	0.68	0.70	0.65	0.62	0.60	0.62	
	MAPE	28.10%	26.58%	30.91%	33.14%	34.97%	33.56%	

1534

1535

1536

1537

1538

1539

1540

1541

1542

1543

1544

1545

1546

1547

1548

1549 **Table3.** The forecast results of T_1 and T_2 in different examples within 6 and 12 months

Forecast events	The results within 6-months		The results within 12-months	
	CC	MAPE	CC	MAPE
The average of 18 El Niño examples of T_1	0.824	8.45%	0.719	12.67%
The average of 22 La Niña examples of T_1	0.846	7.68%	0.740	11.28%
The average of 20 Neutral examples of T_1	0.885	6.23%	0.789	9.85%
The average of total 60 examples of T_1	0.850	7.41%	0.748	10.95%
The average of 18 El Niño examples of T_2	0.811	8.79%	0.703	13.28%
The average of 22 La Niña examples of T_2	0.833	7.35%	0.731	11.96%
The average of 20 Neutral examples of T_2	0.896	6.68%	0.795	10.08%

The average of total 60 examples of T_2	0.842	7.64%	0.740	11.71%	

1550

1551

1552

1553

1554

1555

1556

1557

1558

1559

1560

1561 **Table. 4**. The TC and the MAPE between model forecasts and observations within 12 months for
1562 Nov.–Jan., Dec.–Feb. and Jan.–Mar. as lead time of winter, for Feb.–Apr., Mar.–May and Apr.–June as
1563 lead time of spring, for May-July, June-August and July-Sep. as lead time of summer and for
1564 August-Oct., Sep.-Nov. and Oct.-Dec. as lead time of autumn.

Forecast events	Lead time of all seasons combined		Lead time of summer (MJJ-JJA-JAS)		Lead time of autumn (ASO-SON-OND)		Lead time of winter (NDJ-DJF-JFM)		Lead time of spring (FMA-MAM-AMJ)	
	TC	MAPE	TC	MAPE	TC	MAPE	TC	MAPE	TC	MAPE
The average of 18 El Niño examples	0.604	9.70%	0.569	10.33%	0.632	8.85%	0.677	8.02%	0.538	11.6%
The average of 22 La Niña examples	0.625	8.97%	0.581	9.82%	0.645	8.41%	0.695	7.83%	0.579	9.82%

The average of 20 Neutral examples	0.798	5.96%	0.752	6.86%	0.831	5.31%	0.844	4.60%	0.765	7.07%
The average of total 60 examples	0.712	7.62%	0.633	8.51%	0.786	6.88%	0.776	6.52%	0.653	8.03%

Table. 4. Temporal correlation(CC) and the mean absolute percentage error (MAPE) between model forecasts and observations within 12 months for Nov.-Jan, Dec.-Feb, and Jan.-Mar as lead time of winter and for May-July, June-August and July-Sep. as lead time of summer.

Forecast events	Lead time of all seasons combined		Lead time of summer (MJJ JJA JAS)		Lead time of winter (NDJ DJF JFM)	
	CC	MAPE	CC	MAPE	CC	MAPE
The average of 18 El Niño examples	0.604	9.70%	0.569	10.33%	0.677	8.02%
The average of 22 La Niña examples	0.625	8.97%	0.581	9.82%	0.695	7.83%
The average of 20 Neutral examples	0.798	5.96%	0.752	6.86%	0.844	4.60%
The average of total 60 examples	0.712	7.62%	0.633	8.51%	0.776	6.52%

Table5. The forecast results of the different data periods

Forecast events	The data periods（Jan. 1951-Dec.2010）Lead time of all seasons combined		The data periods（Jan. 1931-Dec.2010）Lead time of all seasons combined		The data periods（Jan. 1941-Dec.2010）Lead time of all seasons combined		The data periods（Jan. 1961-Dec.2010）Lead time of all seasons combined		The data periods（Jan. 1971-Dec.2010）Lead time of all seasons combined	
	TC	MAPE	TC	MAPE	TC	MAPE	TC	MAPE	TC	MAPE
The average of 18 El Niño examples	0.604	9.70%	0.683	9.02%	0.642	9.35%	0.572	10.15%	0.551	10.44%
The average of 22 La Niña examples	0.625	8.97%	0.701	8.33%	0.675	8.55%	0.589	9.42%	0.567	9.82%
The average of 20 Neutral examples	0.798	5.96%	0.845	5.12%	0.821	5.56%	0.746	6.21%	0.721	6.58%
The average of total 60 examples	0.712	7.62%	0.771	7.14%	0.740	7.38%	0.680	7.96%	0.652	8.15%

Table5. The correlation coefficients among four factors

Correlation coefficients	T_1	T_2	SOI	EAWMI

T_1		0.419	0.401	0.337
T_2	0.419		0.424	0.356
SOI	0.401	0.424		0.408
EAWMI	0.337	0.356	0.408	

1585

1586