The relationship among probabilistic, deterministic and potential skills in predicting the ENSO for the past 161 years

Ting Liu[1] · Youmin Tang[1,2] · Dejian Yang[3] · Yanjie Cheng[4] · Xunshu Song[1] · Zhaolu Hou[5] · Zheqi Shen[1] · Yanqiu Gao[1] · Yanling Wu[1] · Xiaojing Li[1] · Banglin Zhang[6]

Abstract

Here, we explored in depth the relationship among the deterministic prediction skill, the probabilistic prediction skill and the potential predictability. This was achieved by theoretical analyses and, in particular, by an analysis of long-term ensemble ENSO hindcast over 161 years from 1856 to 2016. First, a nonlinear monotonic relationship between the deterministic prediction skill and the probabilistic prediction skill, derived by theoretical analysis, was examined and validated using the ensemble hindcast. Further, the co-variability between the potential predictability and the deterministic prediction skill was explored in both perfect model assumption and actual model scenario. On these bases, we investigated the relationship between the potential predictability and probabilistic prediction skill from both the practice of ENSO forecast and theoretical perspective. The results of the study indicate that there are nonlinear monotonic relationships among these three kinds of measures. The potential predictability is considered to be a good indicator for the actual prediction skill in terms of both the deterministic measures and the probabilistic framework. The relationships identified here exhibit considerable significant practical sense to conduct predictability researches, which provide an inexpensive and moderate approach for inquiring prediction uncertainties without the requirement of costly ensemble experiments.

Keywords ENSO ensemble forecast · Predictability · Nonlinear monotonic relationship

1 Introduction

Usually, it involves two branches in the field of predictability researches on the El Niño-Southern Oscillation (ENSO) prediction. The first one is to investigate the ENSO potential predictability, which will reveal what the upper limit of the prediction skill might be and how much room would leave for the improvement of ENSO prediction systems (Tang et al. 2005, 2008; Cheng et al. 2010a; Kumar and Hu 2014; Kumar and Chen 2015). The common measures of the potential predictability can be categorized into variance-based metric and information-based metric, both without using the observations. Specifically, the signal-to-total variance ratio (STR) and signal-to-noise ratio (SNR) are two variance-based measures that are extensively employed (Kumar and Hoerling 2000; Peng et al. 2011; Hu and Huang 2012; Kumar et al. 2016), whereas several commonly used information-based measures include the predictive power (PP), predictive information (PI), relative entropy (RE), and mutual information (MI) (Tang et al. 2008). The studies that use these methods have revealed that the ENSO

✉ Youmin Tang
   ytang@unbc.ca

[1] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China

[2] Environmental Science and Engineering, University of Northern British Columbia, Prince George, BC, Canada

[3] College of Oceanography, Hohai University, Nanjing, China

[4] National Climate Center, China Meteorological Administration, Beijing, China

[5] State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

[6] Institute of Tropical and Marine Meteorology, Guangdong Provincial Key Laboratory of Regional Numerical Weather Prediction, CMA, Guangzhou 510641, China

predictability is mainly dominated by signal component (Tang et al. 2008; Cheng et al. 2011; Kumar and Hu 2014).

The second one is to improve the actual prediction skill through the model development (Zebiak and Cane 1987; Tang and Hsieh 2002; Zhang et al. 2013), data assimilation (Chen et al. 2004; Zheng et al. 2007; Deng et al. 2010; Zheng and Zhu 2010; Zhu et al. 2014; Tang et al. 2016), and ensemble prediction etc. (Tang et al. 2006, 2008; Zheng et al. 2009a; Cheng et al. 2010a; Hou et al. 2018). Many efforts have been paid and significant progresses have been made toward this goal in last decades, as summarized in a recent review paper by Tang et al. (2018). Typically, the actual prediction skill can be measured in a deterministic manner and a probabilistic way. The former is widely measured using ensemble mean that is able to filter out the unpredictable feature and provide a nearly unbiased estimate for the future state of the climate element, whereas the latter intends to provide an estimated likelihood for the future state of climate variable, which can be commonly verified using Brier skill score (BSS).

The connection between the actual prediction skill and the potential predictability has been an interesting issue in the study of ENSO predictability. Several efforts have been devoted for investigating their relationship in the framework of the deterministic measures. For instance, it was found that the potential predictability metrics are good indicators in quantifying the deterministic actual skill in many ENSO models (Kumar and Hoerling 2000; Kumar et al. 2001; Tang et al. 2008; Cheng et al. 2011; Kumar et al. 2001; Kumar and Chen 2015). Recently, the linkage between probabilistic actual skill and deterministic actual skill has been addressed. Cheng et al. (2010a) found that the deterministic correlation skill is nonlinearly related with BSS probabilistic skill in the ENSO prediction. Yang et al. (2016, 2018) further theoretically derived their relationship and found that this relationship originated from the effect of the resolution term, while the reliability did not make contribution.

An interesting issue that has not been explored yet is the possible relationship between the probabilistic prediction skill and the potential predictability. Such a relationship, if existed and found, would be considerably interesting and important, since it offers practical confidence and estimate how much reliable the issued prediction is. Given the potential predictability-deterministic prediction skill relationship and the connection between probabilistic predictability and deterministic prediction skill, one can expect such a relationship existing.

In this paper, we will focus on investigating the relationship among the deterministic skill, probabilistic skill and potential predictability based on a long-term ensemble hindcast production of ENSO. Emphasis is placed on examining the potential predictability and the probabilistic prediction skill from both practical hindcast experiments and theoretical analysis.

In Sect. 2, a concise description of the coupled model and the predictability measures are presented. Section 3 provides detailed information on the construction of the ensemble prediction and evaluates the forecast skill. In Sect. 4, we emphatically explore the relationships among the deterministic skill, the probabilistic skill and the potential predictability, in particular, the relationship between the probabilistic skill and the potential predictability. The main conclusion and discussion are summarized and followed in the final section.

2 Model and predictability metrics

2.1 The LDEO5 model

In this study, we employ the latest edition of the Zebiak–Cane (ZC) model, which is also named as LDEO5 model (Chen et al. 2004). The ZC model is an intermediate coupled model with a linear reduced-gravity ocean model and a Gill-type atmospheric model which driven by the anomalous heating combined sea surface temperature anomaly (SSTA) and low-level moisture convergence (Zebiak 1986). ZC model can reproduces certain key feature of the ENSO phenomenon and is the first coupled model used for ENSO prediction. It has high efficiency in calculation and been extensively applied to investigate the ENSO predictability. The domain of this model spans the tropical Pacific Ocean (124°E–80°W and 28.75°S–28.75°N), with a temporal resolution of 10 days. To initialize the long-term retrospection prediction, we assimilate the monthly reconstructed Kaplan SST V2 datasets (Kaplan et al. 1998) from 1856 to 2016 via a coupled nudging scheme (Chen et al. 2004). We also employ two model output statistic (MOS) procedures as present by Chen et al. (2004) at each integration step to rectify the systematic errors of the model.

2.2 Measures for ensemble prediction

1. Deterministic prediction skill

We employ the anomaly correlation coefficient (ACC) to measure the deterministic prediction skill. For a given lead time of prediction, the ACC is defined below:

$$r = \frac{\sum_{i=1}^{M}(x^o - \overline{x^o})(x^f - \overline{x^f})}{\sqrt{\sum_{i=1}^{M}(x^o - \overline{x^o})^2}\sqrt{\sum_{i=1}^{M}(x^f - \overline{x^f})^2}} \qquad (1)$$

where x is the variable of interest, o and f indicate the observation and forecast, respectively. $\overline{x^o}$ and $\overline{x^f}$ denote the time mean of the observation and forecast, respectively. M denotes the total number of the initial conditions

(161 years × 12 months = 1932). Prior to performing the calculation, anomalization is applied to the prediction and observation via eliminating seasonal cycle.

2. Probabilistic prediction skill

Here we consider three categorical events: above-normal, below-normal and near-normal, and which are identified by the one-third percentiles of the climatic distribution. The Brier Score (BS), a commonly used probabilistic measure, is used. The BS can be formulated as:

$$BS = \frac{1}{M} \sum_{i=1}^{M} (x_i - o_i)^2 \tag{2}$$

here M denotes the number of prediction cases, x_i is the forecast probabilistic of an event, and o_i is the relevant observed outcome. The value of 0 for o_i indicates that the event does not occur and 1 otherwise. Using the binning method (Atger 2004), the probabilistic space has been equally divided into N boxes (taken 10 here). The BS is further decomposed into three terms, that can be referred to as resolution (BS_{RES}), reliability (BS_{REL}), and uncertainty (BS_{UNC}), which is defined as follows (Wilks 2011):

$$BS = \frac{1}{M} \sum_{n=1}^{N} M_n(\overline{x}_n - \overline{o}_n)^2 - \frac{1}{M} \sum_{n=1}^{N} M_n(\overline{o}_n - \overline{o})^2$$
$$+ \overline{o}(1 - \overline{o}) \equiv BS_{REL} - BS_{RES} + BS_{UNC} \tag{3}$$

where M_n and $\overline{x}_n$ are the number and mean of all x_is classifying into nth box and the $\overline{o}_n$ denotes the corresponding observed probabilistic. $\overline{o}$ indicates the mean of $\overline{o}_n$ and indicates the observed climatological probabilistic. The BS includes three items. The term of reliability is a metric of measuring the consistency between the observed probabilistic and the forecast probabilistic. The resolution refers to quantify the difference between the climatological probabilistic and the observed frequencies. The uncertainty term depicts the "a priori" information original from the climatological, which is equivalent to the BS of climatological prediction (i.e., $BS_{UNC} = BS_{CLIM}$).

The BSS is defined as follows (Wilks 2011):

$$BSS = 1 - \frac{BS}{BS_{CLIM}} \tag{4}$$

Given that $BS_{UNC} = BS_{CLIM}$, the BSS is further expressed as (Kharin and Zwiers 2003):

$$BSS = 1 - \frac{BS}{BS_{UNC}} = \frac{BS_{RES}}{BS_{UNC}} - \frac{BS_{REL}}{BS_{UNC}} = BSS_{RES} - BSS_{REL} \tag{5}$$

Compared with BS, BSS is positively oriented. In this study, we employed BSS as the overall probabilistic skill measure; further, the "standardized" reliability and resolution of the BS refer to as the reliability and resolution terms of the BSS (Kharin and Zwiers 2003).

3. Potential predictability

To quantify the potential predictability, we employed the STR measure and an information-based metric (MI). Both of the two measures do not involve with the observation, which are different from the actual prediction skill.

The variabilities of the signal and noise for the ensemble prediction can be measured as the variance of the ensemble mean and ensemble spread of all the initial conditions (Tang et al. 2013), and can be expressed as follows:

$$Var(S) = \frac{1}{M} \sum_{i=1}^{M} (\overline{x}_i - \overline{\overline{x}})^2 \tag{6}$$

$$Var(N) = \frac{1}{MK} \sum_{j=1}^{M} \sum_{k=1}^{K} (x_{j,k} - \overline{x}_j)^2 \tag{7}$$

where $x_{j,k}$ denotes the kth member of the ensemble prediction for the jth initial condition. K is the ensemble size and M indicates the total number of the initial conditions. Given the influence of the sampling error on measuring the signal variance, the more reasonable estimation of the signal variance can be given as follows (Rowell 1998):

$$Var(S) = \frac{1}{M} \sum_{i=1}^{M} (\overline{x}_i - \overline{\overline{x}})^2 - \frac{1}{K} Var(N) \tag{8}$$

Further, the STR is defined as:

$$STR = \frac{Var(S)}{Var(S) + Var(N)} \tag{9}$$

MI is one of the information-based metrics that can measure the overall potential predictability of the prediction system. According to the information theory, MI is expressed as (DelSole 2004):

$$MI = \iint p(\mu, i) ln \left[\frac{p(\mu, i)}{p(\mu)p(i)} \right] d\mu di \tag{10}$$

where μ denotes the predicted future state of a climate variable, with a climatological distribution of $p(\mu)$. $p(i)$ indicates the probabilistic distribution of the initial condition i. $p(\mu, i)$ is the joint probabilistic distribution between μ and i. When the climatological and forecast distribution are Gaussian, which are believed to be applicable in the climate prediction, the simplified MI as follows (DelSole and Tippett 2007):

$$MI = \frac{1}{2}\left(ln\sigma_\mu^2 - \frac{1}{M}\sum_{i=1}^{M} ln\sigma_{\mu|i}^2\right) \qquad (11)$$

where σ_μ^2 and $\sigma_{\mu|i}^2$ indicate the climatological variance and the ensemble variance of ith initial condition, respectively. Here we followed the lead-time-dependent method used by Yang et al. (2012) to estimate σ_μ^2. Specifically, the total members of all the initial conditions (sample size: $M \times K$) at each lead time are employed to calculate σ_μ^2. $\sigma_{\mu|i}^2$ is estimated based on all the members of the ith initial condition.

In the perfect model framework, we assumed that an arbitrarily predicted ensemble member can be recognized as the "observation". The square root of STR is often defined as the potential correlation, which can measure the perfect correlation skill, denoted by R as follows (Kumar 2009; Tippett et al. 2010; Tang et al. 2013):

$$R = \sqrt{STR} \qquad (12)$$

For the Gaussian variable and forecast variance is constant, there is a theoretical relationship between MI and R as flowing (DelSole 2004):

$$MI = -\frac{1}{2}ln(1 - R^2) = -\frac{1}{2}ln(1 - STR) \qquad (13)$$

For more details, see relevant literatures (Kleeman and Moore 1997; DelSole 2004; Tang et al. 2013).

3 Construction of the ensemble prediction

Generally, uncertainties for the ENSO prediction derived from the uncertainties in initial condition and the deficiencies in model formulation. The practice of the ensemble prediction is a useful strategy for sampling and evaluating these uncertainties (Wilks and Vannitsem 2018). From the perspective of optimal error growth, we combined two kinds of the optimal perturbation schemes for the initial condition and the stochastic atmospheric noise that has not been well considered in the framework of model.

The initial condition perturbation is constructed by the singular vector (SV) approach (Lorenz 1965), which can capture the fastest growing initial errors in the ENSO evolution (Xue et al. 1997; Fan et al. 2000; Cheng et al. 2009, 2010b). Following Cheng et al. (2009), we carry out the SV analysis for every monthly initial conditions with the optimal period of 12 months from 1856 to 2016. Consistent with the previous research (Cheng et al. 2009), the spatial pattern of the first SV is not sensitive to the initial condition in the LDEO5 model, and the fastest perturbation growth occurs at around 12 month lead (not shown). The averaged first SV pattern over the past 161 years of 12 month lead

is featured by an east–west dipole pattern over the tropical Pacific region (Fig. 1a), and the associated zonal SST gradient can favor the eastward warm Kelvin wave to transport the warmer water into the eastern Pacific Ocean, and resulting in an El Niño anomalies pattern as in Fig. 1b.

The increasing researches have indicated that the stochastic atmospheric events significantly contributed to the ENSO variability (Penland and Sardeshmukh 1995; Kleeman and Moore 1997; Thompson and Battisti 2000; Zavala et al. 2005; Gebbie et al. 2007; Lian et al. 2014; Chen et al. 2015). Further, the model used in this study is free of stochastic processes. An effective method that considering the impact of stochastic uncertainties on the predictions in such noise-free models would be to use the stochastic optimals (SOs, Kleeman and Moore 1997) analysis. Compared with the SVs that can only consider the uncertainties related to the initial condition, the SOs analysis characterizes the uncertainties that are triggered by the stochastic processes during the entire prediction (Tang et al. 2008). More detail of SOs can be found in the related references (Kleeman and Moore 1997; Tang et al. 2008). In this study, we only consider atmospheric stochastic processes, and construct the SOs by perturbing the wind stress during the prediction integration.
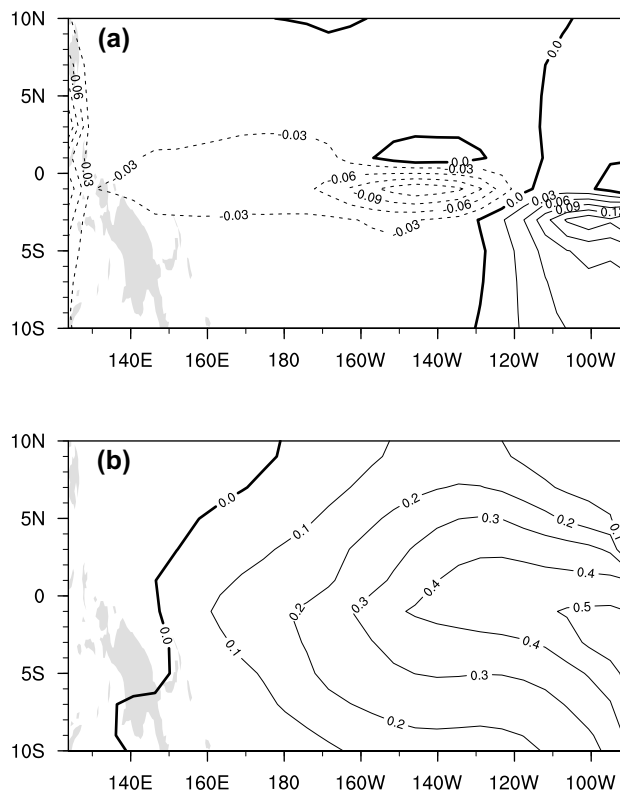

Fig. 1 **a** The first singular vector of SST anomalies (°C) averaged in the past 161 years. **b** The corresponding final pattern of SST anomalies (°C) for **a**
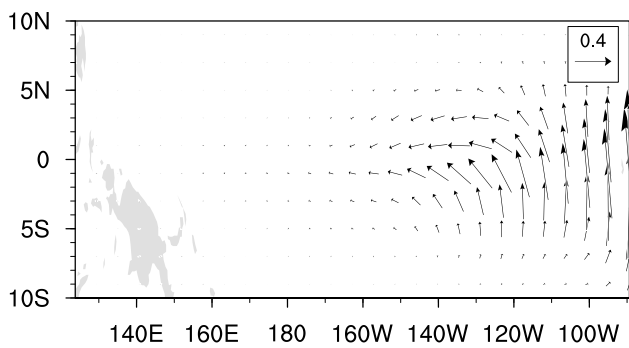
Fig. 2 The leading mode of the first stochastic optimal winds (m s^{-1}) over the past 161 year

Figure 2 depicts the leading EOF mode of the first SOs of all predictions, representing the fastest error growth of SST due to stochastic processes incorporated into wind stress. Figure 2 shows the strong convergence in the east of 140°W and the offshore wind near 100°W, both favoring upwelling in the eastern Pacific, and leading to the La Niño-like SST anomalies, and vice versa.

Given that involving two or more perturbation modes does not improve the prediction skill (Cheng et al. 2011), the ensemble hindcast here is constructed by jointly perturbing the initial conditions with the leading SV of SST and the leading SO of wind stress. The hindcast is performed on the first day of each calendar month for from 1856 to 2016, each lasting 12 months, with an ensemble size of 100. As depicted in Fig. 3, this joint perturbation scheme provides a skillful and reliable ENSO forecast. The correlation skill of Niño 3.4 SSTA index (defined as the averaged SSTA over the 5°S–5°N, 120°–170°W) between the ensemble mean and the observation (Fig. 3a, red line) is much better than the persistence forecast (Fig. 3a, blue line). Further, the correlation skill is over 0.7 (0.5) at 6-month (12-month lead) for the period of the past 161 years, which can catch up, even better than, that of the current state-of-the-art coupled models. If an ensemble prediction system reasonably considers possible uncertainties, the averaged ensemble spread (SPRD) will be close to its root mean square error ($RMSE_{EM}$, Toth et al. 2003). The latter should be comparable with, or even better than, the RMSE of the unperturbed single forecast ($RMSE_{CTL}$, Toth and Kalnay 1997). The aforementioned relationship holds well in our ensemble forecast system as shown in Fig. 3b.
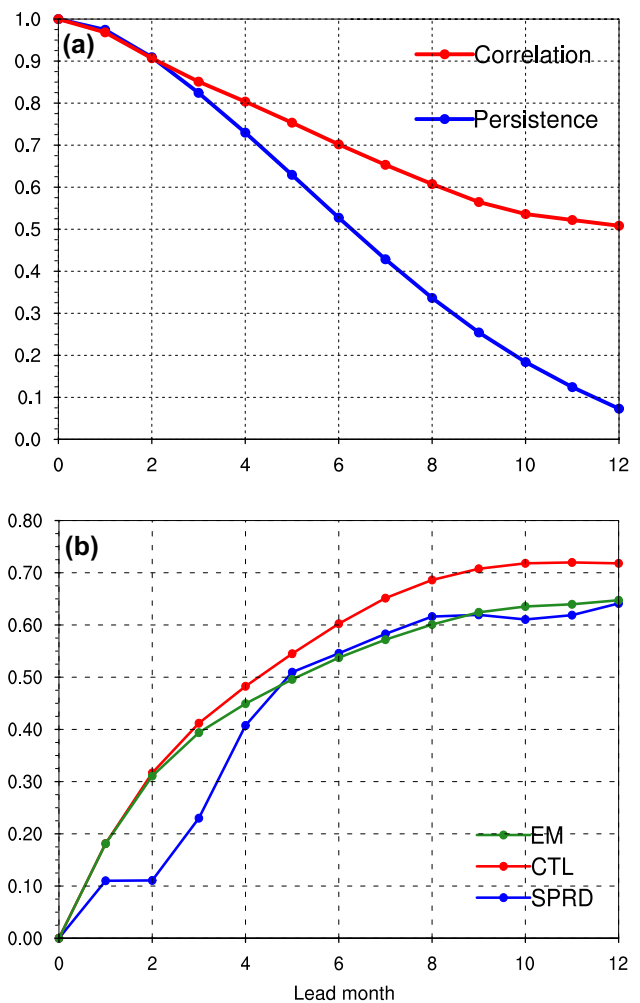

Fig. 3 **a** Correlation skill (red) and persistence skill (blue) of the Niño 3.4 index from ensemble mean along with lead time. **b** Root mean square error of Niño 3.4 index from ensemble mean (green), control run (red) and ensemble spread (blue) along with lead time

4 Relationship among the deterministic skill, the probabilistic skill and the potential skill
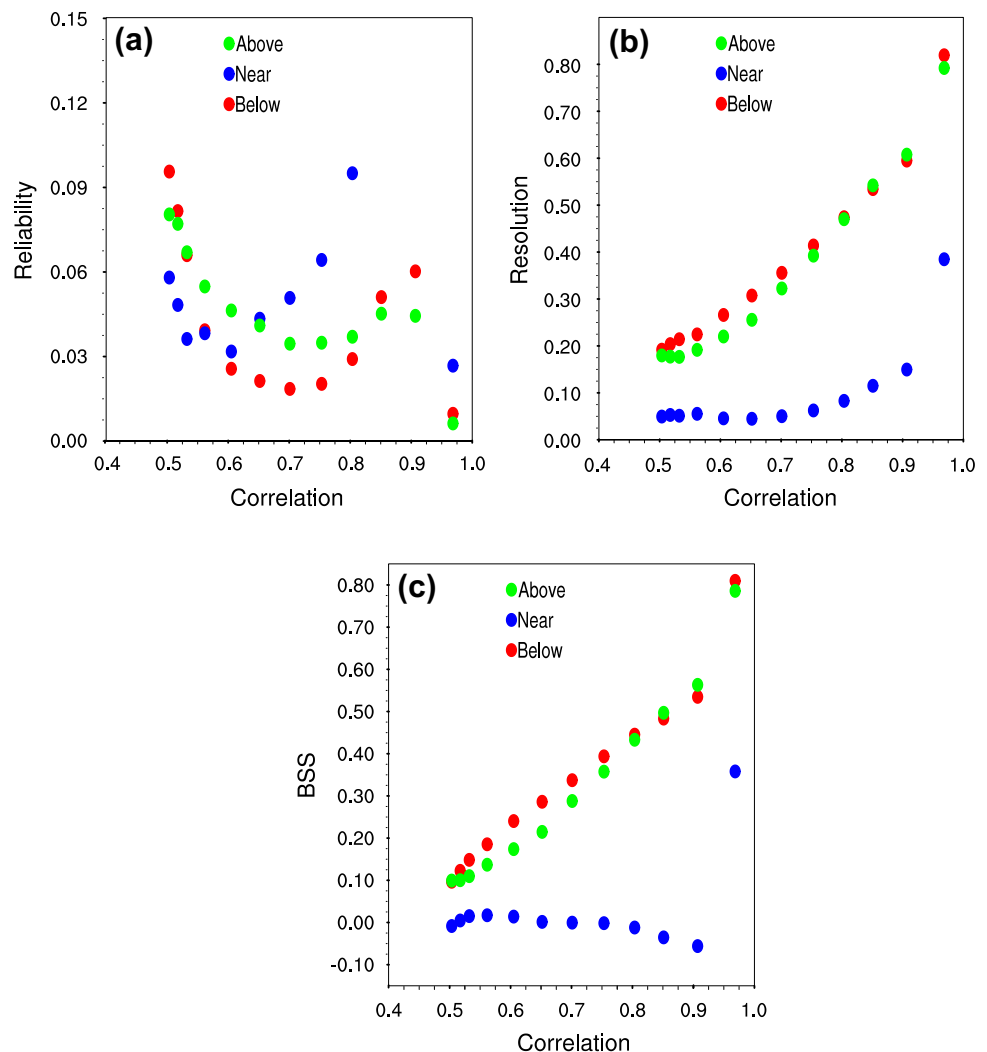
4.1 Relationship between the deterministic skill and the probabilistic skill

Previous studies have reported that monotonic nonlinear relationships are observed between the probabilistic skill (BSS) and the deterministic skill (r) in the practice of the seasonal climate prediction with respect to the precipitation (Wang et al. 2009; Yang et al. 2016), surface air temperature (Alessandri et al. 2011), zonal wind (Yang et al. 2016), and geopotential height (Yang et al. 2018). Furthermore, these monotonic nonlinear relationships benefit from the contributions of the covariability between the correlation and the resolution skill (Alessandri et al. 2011; Yang et al.

2016, 2018). For ENSO ensemble prediction experiment, this principle is also operating (Fig. 4). As depicted in the Fig. 4a, the distribution in the scatterplot of the reliability versus the correlation is irregular and it's difficult to draw a relationship. In contrast, there is a distinct monotonic nonlinear relationship in the scatterplot of the resolution versus the correlation (Fig. 4b). In each lead month, the correlations exhibits covariability with the resolution skills for all three categorical ENSO events. The scatter pattern also highly obeys the theoretical relationship derived by Yang et al. (2018). Since the resolution dominates the BSS, a good relationship between BSS and the correlation is also observed as shown in Fig. 4c, although it is not as much as that in Fig. 4b due to the degradation of reliability.

In short, the monotonic nonlinear deterministic skill-probabilistic skill relationship which was found in previous studies (Yang et al. 2018), also holds well in the ENSO ensemble prediction system. This implies that the improvement of the ENSO resolution prediction also enhances its deterministic skill, and vice versa.

4.2 Relationship between the potential predictability skill and the deterministic prediction skill

Previous studies argued that the potential predictability, measured by MI, exhibits a good monotonic relationship with the deterministic skill under the prefect model scenario (Tang et al. 2008; Cheng et al. 2010a, b). This is also true in this case. Shown in Fig. 5a is the potential correlation R against the MI, depicting an evident monotonic nonlinear relationship (the blue dots). Compared with the theoretical value (the red dots), it reveals that the theoretical relationship in Eq. (13) can be approximate represented in our ENSO prediction In fact, the equalities in the above equation hold is and only if the forecast variance is constant. However, this is not strictly hold in our ensemble prediction (not shown), which results in the difference between the real MI and the corresponding theoretical value. Interestingly, such a relation also exists between the actual deterministic skill (r) and the MI as depicted in Fig. 5b. Namely, a larger MI
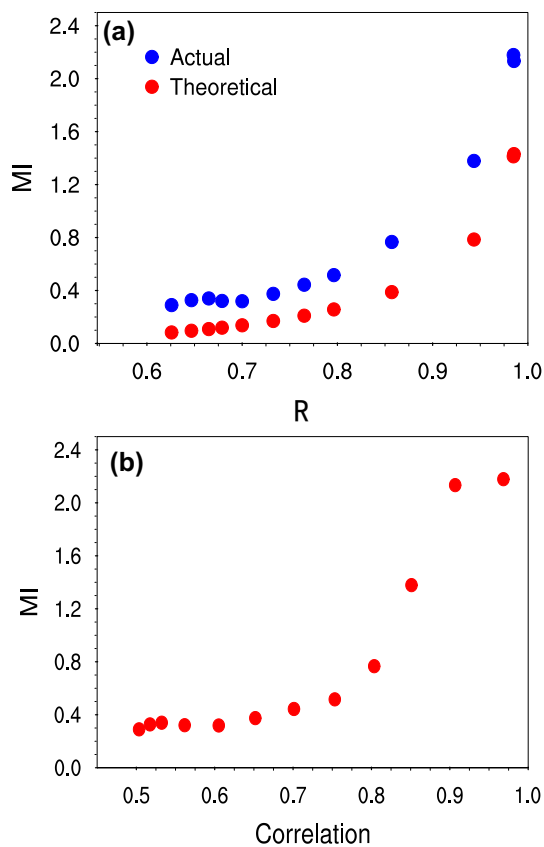
Fig. 4 Scatterplots of **a** reliability versus correlation; **b** resolution versus correlation; **c** Brier skill score (BSS) versus correlation of the Niño 3.4 index along with lead time for below-normal (red), above-normal (green) and near-normal (blue) events

Fig. 5 Scatterplots of **a** potential correlation (R) versus MI along with lead time. The blue and red dots in a indicated the model and theoretical values, respectively. **b** Same as a, but for the real model scenario

corresponds to a higher deterministic correlation skill both in perfect and actual model scenarios, which approximately follow the theoretical relationship as expressed in Eq. (13).

4.3 Relationship between the potential predictability skill and the probabilistic prediction skill

The aforementioned discussion indicates that a monotonic nonlinear relationship exists between the probabilistic skill (resolution) and the deterministic skill. Also, the deterministic skill exhibits a nonlinear monotonic relationship with the potential predictability skill (MI). On this basis, the expectation of the connection between the probabilistic skill and the potential predictability skill comes naturally. As discussed in Yang et al. (2012), the actual correlation (r) should be equal to the potential correlation (R) under the perfect model scenario. In our ENSO prediction system, the two exhibits co-variability with the correlation coefficient is 0.987, indicating that the potential correlation is a useful indicator for the actual correlation. In addition, Yang et al. (2018) also proved mathematically such a relationship existing between

the actual correlation and the resolution. Therefore, after substituting the correlation term of Eqs. (21)–(23) in Appendix A with the Eq. (13) of this study, we obtain the theoretical resolution as a function of the MI for the perfect model scenario in the following manner (for more details, refer to Appendix A):

$$P(O_{AN} = 1|\mu) = \phi\left(\frac{\phi^{-1}(1/3) + \sqrt{1 - e^{-2MI}}X}{e^{-2MI}}\right) \quad (14)$$

$$P(O_{BN} = 1|\mu) = \phi\left(\frac{\phi^{-1}(1/3) - \sqrt{1 - e^{-2MI}}X}{e^{-2MI}}\right) \quad (15)$$

$$P(O_{NN} = 1|\mu) = 1 - P(O_{AN} = 1|\mu) - P(O_{BN} = 1|\mu) \quad (16)$$

$$BSS_{RES} = \frac{9}{2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}X^2\right)\left[P(O = 1|\mu) - \frac{1}{3}\right]^2 dX \quad (17)$$

where $X = \frac{\mu - \mu_c}{\sigma_\mu}$, μ is the ensemble mean prediction; further, μ_c and σ_μ are the corresponding climatological mean and standard deviation, respectively. $P(O = 1|\mu)$ denotes the probabilistic of each categorical event. The cumulative distribution function and the its inverse function are indicated as $\phi(.)$ and $\phi^{-1}(.)$, respectively.

Figure 6 depicts the numerical integration (the lines) of Eq. (17), thereby denoting that the resolution is a monotonic nonlinear function of MI both for the above-normal (Fig. 6a), below-normal (Fig. 6b) and the near-normal events (Fig. 6c). Compared with the model MI-resolution relationship under the perfect model scenario (the dots in the Fig. 6), theoretical relationship is more well hold for the above-normal and below-normal events than that for the near-normal events. The possible reason is that there are more reliable probability prediction skills for the above-normal and below-normal events in this ensemble prediction system. Figure 7 presents the observed frequency against the forecast probability for three categorical events. For the near-normal events (Fig. 7c), the curves are more deviated from the diagonal line (perfect line) than the above-normal (Fig. 7a) and below-normal (Fig. 7b) prediction's, illustrating the significant "overconfidence" character in the probability prediction for the near-normal events. This indicate that this ensemble prediction system can give more reliable probability prediction skill for the above-normal and below-normal events. Even so, there is also some bias between the theoretical value and the model result from lead 1–3 months for the above-normal and below-normal events to a certain extent. Due to the inadequate development of the spread, the ensemble prediction system underestimated the possible uncertainties, which results in the low probability prediction
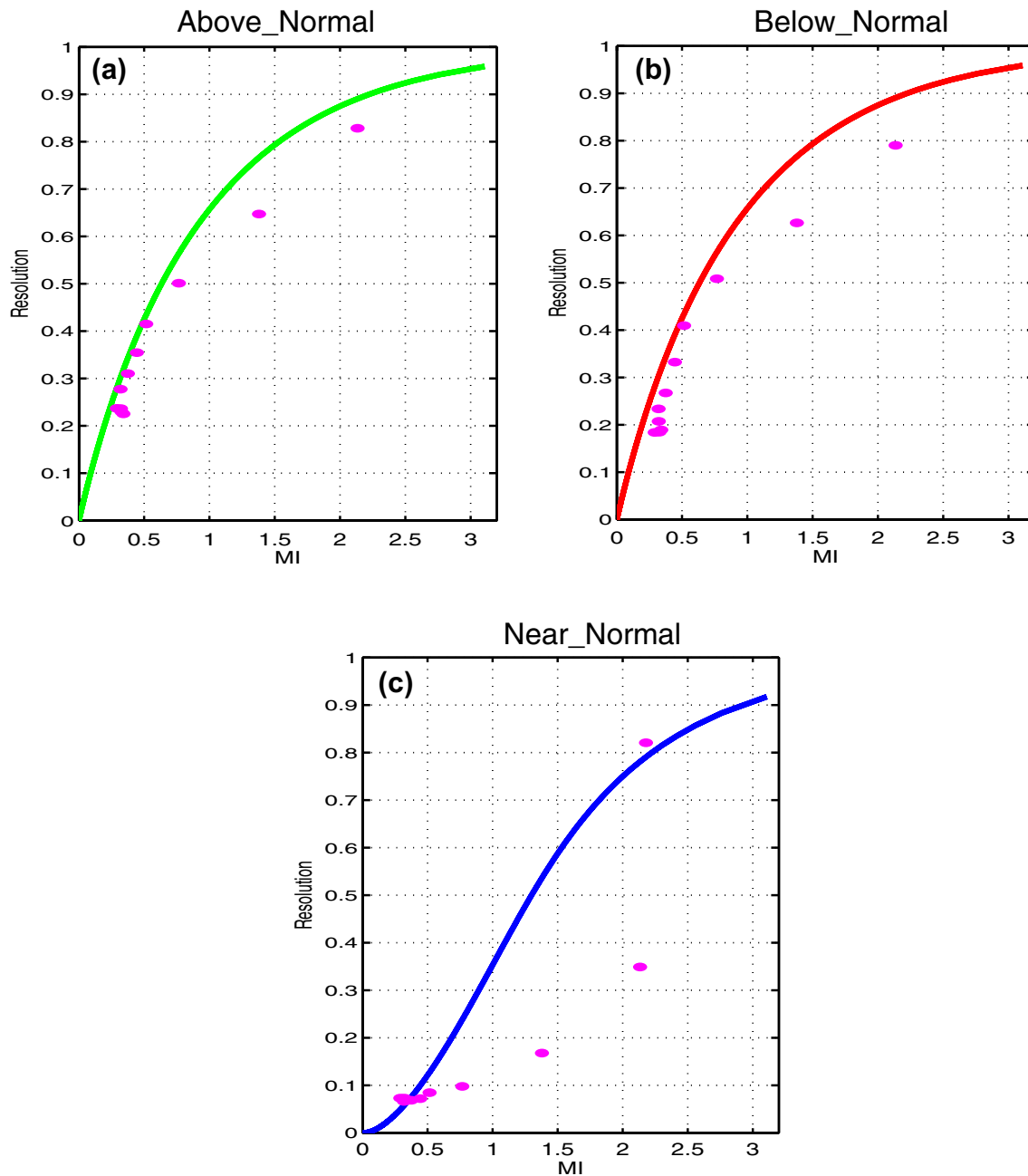
Fig. 6 Theoretical relationship (the lines) between the MI and resolution and corresponding the model predictions (the dots) under the perfect model scenario for **a** above-normal events, **b** below-normal events and **c** near-normal events

skill and accounting for the above mentioned bias. The insufficient spread in the lead 1–3 months (as shown in the Fig. 3b) may be limited by the linear perturbation method as used here, and it still exists in previous research (Cheng et al. 2010a, b). We can employ some no-linear methods to overcome this deficiency in the future.

In addition, although the Eq. (17) describes the monotonic relationship between the MI and the resolution based under a perfect model assumption, it seems that this theoretical resolution-MI relationship is also approximately validated in real ENSO ensemble prediction (Fig. 8). A salient feature is that there seemingly exists a nonlinear relationship between MI and the resolution. And that the relationship is nearly monotonic, namely, a larger MI corresponds a higher resolution and the variation of the resolution with the MI changes more rapidly when the MI is at small value. However, the biases between the model result (the dots) and the theoretical value (the line) are certainly more visible than that under the
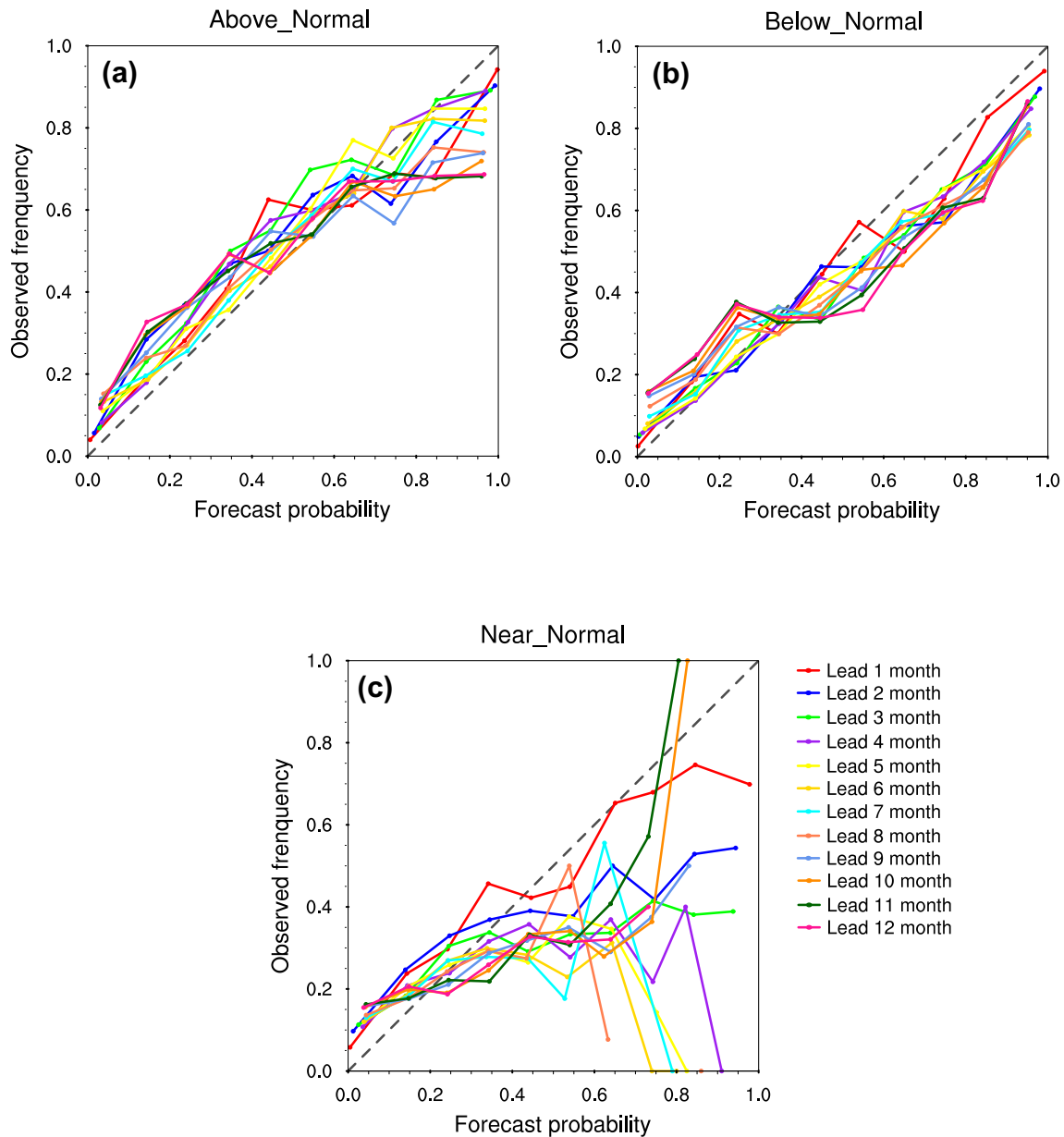
Fig. 7 Reliability diagram for **a** above-normal events, **b** below-normal events and **c** near-normal events the from lead 1–12 months

perfect model scenario (Fig. 6). The possible reason lie in that the model uncertainties also affect the prediction result and lead to the more imperfect bias between the theoretical value and the model results. Of particular note is with the development of the ensemble spread, the real model result nearly lay on the theoretical curve of resolution-MI relationship and come near to the theoretical value, suggesting that the reliable ensemble prediction system, which contain the possible uncertainties may can cover the shortage of the model formulation to some extent.

In spite of these insignificant imperfections, the theoretical nonlinear monotonic potential predictability–probability

prediction skill relationship can still be well verified in our ENSO ensemble prediction. As the deterministic prediction skill also exhibits covariability with the potential predictability skill in our ENSO ensemble forecast system (Fig. 5b), then indicating that a high potential predictability MI always corresponds to more accurate actual prediction skill in both the deterministic manner and probabilistic way.
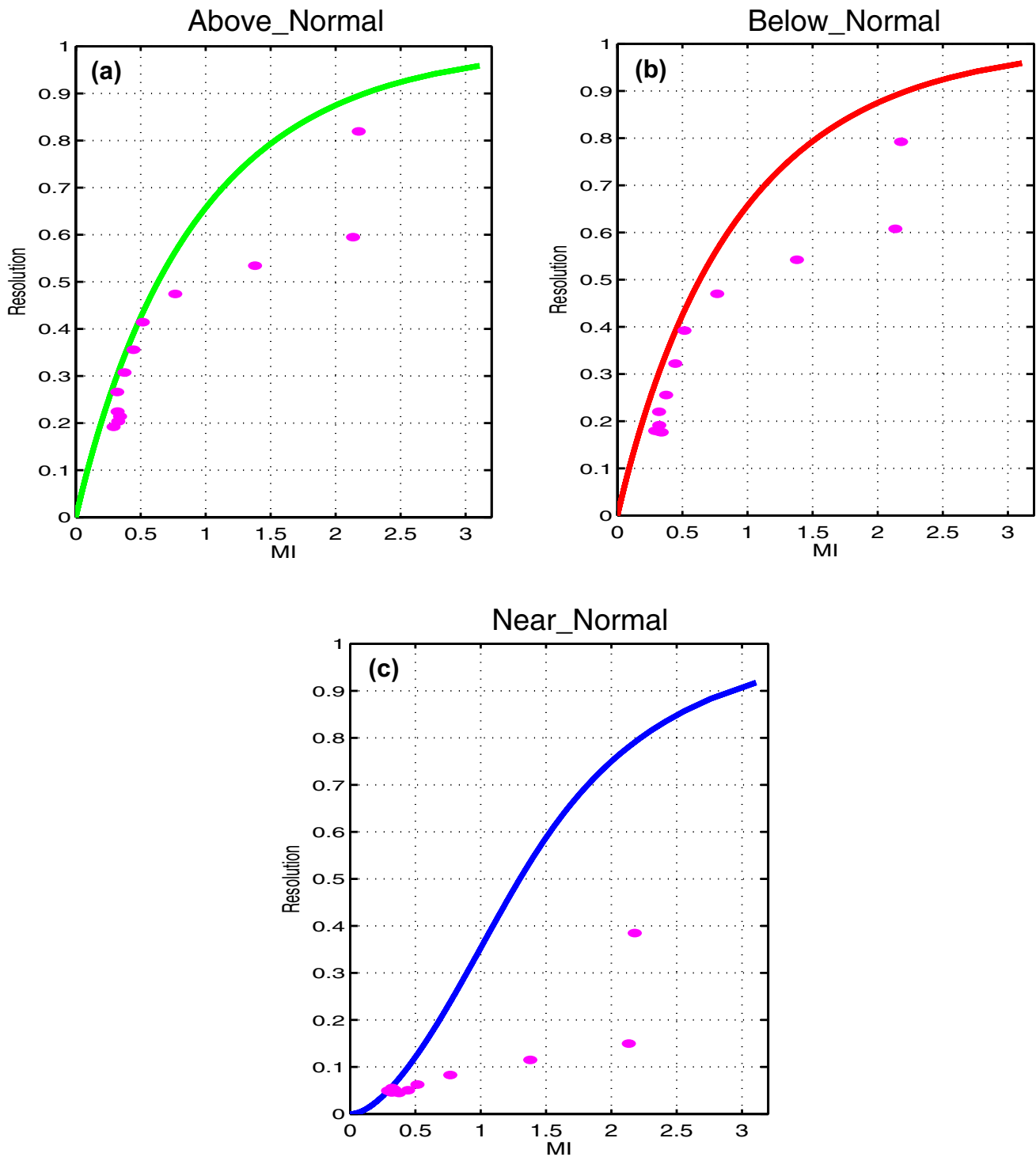
Fig. 8 Same as Fig. 6, but for the real model prediction

5 Summary and discussion

The practice of evaluating and understanding the ENSO predictability involves two perspectives: actual prediction skill and the potential predictability. The former one can be investigated either from the deterministic viewpoint or probabilistic angle. Further, the inherent relationship between these two actual prediction skills has been validated from both the theoretical derivation and the practice of the seasonal climate prediction (Wang et al. 2009;

Alessandri et al. 2011; Yang et al. 2016, 2018). In addition, the potential predictability measures are considered to be potential candidate indicators for the actual deterministic skill in a perfect model scenario (Tang et al. 2008; Cheng et al. 2010a, b). However, it still lacks systematic research on the issue about the coherent relationship among the deterministic skill, the probabilistic skill and the potential predictability in the ENSO prediction. In this study, we investigated the relationship among them based on the ENSO ensemble hindcast prediction over the 161 year, and especially discussed the connection between the probabilistic skill and the potential predictability skill from the practical application and theoretical identification, which has never been involved.

First, we constructed the ensemble forecast system with the combination of the optimal perturbations of the initial condition and the model stochastic physical processes. Based on this forecast system, we performed long-term hindcast over the past 161 year (1856–2016). The evaluation of the prediction indicated that this joint ensemble construction strategy could provide the skillful long-term ENSO prediction at lead 12 months. The further analysis demonstrated that the deterministic skill exhibited a nonlinear monotonic relationship with its probabilistic counterpart, in which the resolution property made the dominant contribution. These further confirm the theoretical result reported by Yang et al. (2018) with respect to the practice of the ENSO prediction and to imply that the enhancement of the ENSO resolution skill can also correspond to the improvement of the correlation skill.

In addition, our analysis also demonstrated that there was a nonlinear monotonic relationship between the deterministic prediction skill and the information-based potential predictability skill (MI) whether under the perfect model scenario or actual model prediction. These relationships could be approximately explained by a Gaussian-based theoretical equation, implying that MI has potential to indicate the actual prediction skill in this system.

Given that both the probabilistic skill and the potential predictability skill exhibit covariability with its deterministic compatriot in our ENSO ensemble forecast system, it's naturally motivated us to further investigate the possible relationship between the probabilistic prediction skill and potential predictability skill. The result revealed the existence of a nonlinear monotonic link between the two in our ENSO ensemble forecast. Specifically, the high potential predictability skill always corresponds with high probabilistic skill. Under the assumptions that the forecast probabilistic density function is Gaussian, we derived a theoretical expression for the BSS_{RES}-MI relationship for the perfect model scenario. This theoretical demonstration was also practically well confirmed in our ENSO ensemble prediction.

All in all, the nonlinear monotonic relationships among the deterministic skill, the probabilistic skill and the potential predictability that are discussed in this study implies that we can quantify the model prediction skill (deterministic correlation or probabilistic resolution) by estimating the MI while issuing the prediction, which offers useful information for the end-users of prediction. According to the resolution-correlation relationship, it can be inferred that a possible effective way to improve the deterministic skill of ENSO prediction is to perform the multi-model ensemble prediction (Tippett and Barnston 2008), which has the potential to improve the resolution (Yang et al. 2016, 2018).

In short, the theoretical relationships among the deterministic skill, probabilistic skill and the potential predictability can approximately hold for the ENSO ensemble forecast. However, there are still some imperfections, especially with regard to the short lead time, where the ensemble spread is not sufficiently developed. This may be related to the linear-based perturbations we employed here. In further, we will focus on improving the ENSO actual skill through a combination of multiple-model ensemble prediction approach and some nonlinear-based perturbation methods, such as stochastic model-error perturbations (Zheng et al. 2009b) and Conditional Nonlinear Optimal Perturbation (Duan and Mu 2004) and so on.

Appendix A

As Yang et al. (2016, 2018) discussed, the resolution could be understood as the statistical dependence between the event occurrence and the probabilistic forecast. On the infinite sample size condition, the formal resolution is expressed as (Palmer et al. 2000; Yang et al. 2016, 2018):

$$BSS_{RES} = \frac{9}{2} \int_0^1 f_p(p)[P(O = 1|p) - P(O = 1)]^2 dp \qquad (18)$$

where p denotes the forecast probabilistic of the considered event and O represents the corresponding observational fact

of this event, and 1 and 0 are to indicate occurrence and nonoccurrence, respectively. $P(\ )$ and $P(\ |\ )$ indicate unconditional and conditional probabilities separately. $f_p(p)$ refers to as the probabilistic density function (PDF) of p.

If the forecast PDF obeys a Gaussian distribution with a homogeneous forecast variance [the seasonal average variables always satisfied this assumption (Kumar et al. 2000; Tang et al. 2008; Yang et al. 2012, 2016, 2018)], then the p is eventually a function of the forecast mean μ only. Then, Eq. (18) can be rewritten as:

$$BSS_{RES} = \frac{9}{2} \int_{-\infty}^{\infty} f_\mu(\mu)[P(O = 1|\mu) - P(O = 1)]^2 d\mu \quad (19)$$

where $f_\mu(\mu)$ is the PDF of μ. This equation implies that the resolution is governed by the statistical dependence between the forecast (μ) and observation (χ).

Generally, the seasonal mean forecast (μ) and observation (χ) variables would approximately satisfy a joint Gaussian distribution. And the conditional PDF of χ given μ ($f_{\chi|\mu}(\chi|\mu)$) is a Gaussian PDF with mean ($E(\chi|\mu)$) is $\chi_c + r\frac{\sigma_\chi}{\sigma_\mu}(\mu - \mu_c)$) and variance ($Var(\chi|\mu)$) is $(1 - r^2)\sigma_\chi^2$)). Where $\mu_c(\sigma_\mu^2)$ represents the climatological mean (variance) of μ and $x_c(\sigma_\chi^2)$ denotes the climatological mean (variance) of χ, and r is the correlation between μ and χ.

Then $P(O = 1|\mu)$ in Eq. (19) can be obtained by integration of $f_{\chi|\mu}(\chi|\mu)$ from the χ_l to χ_r. Then it further would be written as the form of the cumulative distribution function of Standard Gaussian distribution:

$$P(O = 1|\mu) = \phi[(\chi_r - E(\chi|\mu))/\sqrt{Var(\chi|\mu)}] \\ - \phi[(\chi_l - E(\chi|\mu))/\sqrt{Var(\chi|\mu)}] \quad (20)$$

For the below-normal event, $\chi_l = -\infty$ and $\chi_r = \chi_c + \sigma_\chi \phi^{-1}(1/3)$; for the above-normal event, $\chi_l = \chi_c + \sigma_\chi \phi^{-1}(2/3) = \chi_c - \sigma_\chi \phi^{-1}(1/3)$ and $\chi_r = +\infty$. Here ϕ^{-1} is the inverse function of ϕ. That is, we have

$$P(O_{BN} = 1|\mu) = \phi\left(\frac{\chi_c + \sigma_\chi \phi^{-1}(1/3) - \chi_c - r\frac{\sigma_\chi}{\sigma_\mu}(\mu - \mu_c)}{\sqrt{(1 - r^2)}\sigma_\chi}\right) \\ - 0 = \phi\left(\frac{\phi^{-1}(1/3) - r\frac{(\mu - \mu_c)}{\sigma_\mu}}{\sqrt{(1 - r^2)}}\right) \quad (21)$$

$$P(O_{AN} = 1|\mu) = 1 - \phi\left(\frac{\chi_c - \sigma_\chi \phi^{-1}(1/3) - \chi_c - r\frac{\sigma_\chi}{\sigma_\mu}(\mu - \mu_c)}{\sqrt{(1 - r^2)}\sigma_\chi}\right) \\ = \phi\left(\frac{\phi^{-1}(1/3) + r\frac{(\mu - \mu_c)}{\sigma_\mu}}{\sqrt{(1 - r^2)}}\right) \quad (22)$$

$$P(O_{NN} = 1|\mu) = 1 - P(O_{AN} = 1|\mu) - P(O_{BN} = 1|\mu) \quad (23)$$

Given under the perfect model scenario, the actual correlation (r) should be equal to potential correlation (R, Yang et al. 2012; Kumar et al. 2014), and replacing the correlation item of the Eqs. (21–23) based on the relationship between the MI and R, we have:

$$P(O_{BN} = 1|\mu) = \phi\left(\frac{\phi^{-1}(1/3) - \sqrt{1 - e^{-2MI}}\frac{(\mu - \mu_c)}{\sigma_\mu}}{e^{-2MI}}\right) \quad (24)$$

$$P(O_{AN} = 1|\mu) = \phi\left(\frac{\phi^{-1}(1/3) + \sqrt{1 - e^{-2MI}}\frac{(\mu - \mu_c)}{\sigma_\mu}}{e^{-2MI}}\right) \quad (25)$$

$$P(O_{NN} = 1|\mu) = 1 - P(O_{AN} = 1|\mu) - P(O_{BN} = 1|\mu) \quad (26)$$

When substituting the explicit Gaussian PDF of $f_\mu(\mu)$ and considering the fact that $P(O = 1) = 1/3$, the expression of resolution is:

$$BSS_{RES} = \frac{9}{2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\mu - \mu_c}{\sigma_\mu}\right)^2\right][P(O = 1|\mu) \\ - P(O = 1)]^2 d\left(\frac{\mu - \mu_c}{\sigma_\mu}\right) \quad (27)$$

References

Alessandri A, Borrelli A, Navarra A, Arribas A, Déqué M, Rogel P, Weisheimer A (2011) Evaluation of probabilistic quality and value of the ENSEMBELS multimodel seasonal forecasts: comparison with DEMETER. Mon Weather Rev 139:581–607

Atger F (2004) Estimation of the reliability of ensemble-based probabilistic forecasts. Q J R Meteorol Soc 130:627–646

Chen D, Cane MA, Kaplan A, Zebiak SE, Huang D (2004) Predictability of El Niño over the past 148 years. Nature 428:733–736

Chen D, Lian T, Fu C, Cane MA, Tang YM, Murtugudde R, Song XS, Wu QY, Zhou L (2015) Strong influence of westerly wind bursts on El Niño diversity. Nat Geosci 8:339–345

Cheng YJ, Tang YM, Zhou XB, Jackson P, Chen D (2009) Further analysis of singular vector and ENSO predictability from 1856 to 2003—Part I: singular vector and control factors. Clim Dyn 35:807–826

Cheng YJ, Tang YM, Jackson P, Chen D, Deng ZW (2010a) Ensemble construction and verification of the probabilistic ENSO prediction in the LDEO5 model. J Clim 23:5476–5479

Cheng YJ, Tang YM, Jackson P, Chen D, Zhou XB, Deng ZW (2010b) Further analysis of singular vector and ENSO predictability from 1856 to 2003—Part II: singular value and predictability. Clim Dyn 35:827–840

Cheng YJ, Tang YM, Chen D (2011) Relationship between predictability and forecast skill of ENSO on various time scales. J Geophys Res 116(C12). https://doi.org/10.1029/2011JC007249

DelSole T (2004) Predictability and information theory. Part I: measures of predictability. J Atmos Sci 61:2425–2440

DelSole T, Tippett MK (2007) Predictability: recent insights from information theory. Rev Geophys 45:4002

Deng ZW, Tang YM, Wang G (2010) Assimilation of Argo temperature and salinity profiles using a bias-aware localized EnKF system for the Pacific Ocean. Ocean Model 35:187–205

Duan WS, Mu M (2004) Conditional nonlinear optimal perturbations as the optimal precursors for El Nino–Southern Oscillation events. J Geophys Res Atmos 109(D23). https://doi.org/10.1029/2004jd004756

Fan Y, Allen M, Anderson D, Balmaseda M (2000) How predictability depends on the nature of uncertainty in initial conditions in a coupled model of ENSO. J Clim 13:3298–3313

Gebbie G, Eisenman I, Wittenberg A, Tziperman E (2007) Modulation of westerly wind bursts by sea surface temperature: a semistochastic feedback for ENSO. J Atmos Sci 64:3281–3295

Hou ZL, Li JP, Ding RQ, Feng J, Duan WS (2018) The application of nonlinear local Lyapunov vectors to the Zebiak–Cane model and their performance in ensemble prediction. Clim Dyn 51:283–304

Hu Z, Huang B (2012) The predictive skill and the most predictable pattern in the tropical atlantic: the effect of ENSO. Mon Weather Rev 135:1786–1806

Kaplan A, Cane M, Kushnir Y, Clement A, Blumenthal M, Rajagopalan B (1998) Analysis of global sea surface temperature 1856–1991. J Geophys Res 103:18567–18589

Kharin V, Zwiers F (2003) Improved seasonal probability forecasts. J Clim 16:1684–1701

Kleeman R, Moore A (1997) A theory for the limitation of ENSO predictability due to stochastic atmospheric transients. J Atmos Sci 54:753–767

Kumar A (2009) Finite samples and uncertainty estimates for skill measures for seasonal prediction. Mon Weather Rev 137:2622–2631

Kumar A, Chen M (2015) Inherent predictability, requirements on the ensemble size, and complementarity. Mon Weather Rev 143:3192–3203

Kumar A, Hoerling M (2000) Analysis of a conceptual model of seasonal climate variability and implications for seasonal prediction. Bull Am Meteorol Soc 81:255–264

Kumar A, Hu Z (2014) How variable is the uncertainty in ENSO sea surface temperature prediction? J Clim 27:2779–2788

Kumar A, Barnston A, Peng P, Hoerling M, Goddard L (2000) Changes in the spread of the variability of the seasonal mean atmospheric states associated with ENSO. J Clim 13:3139–3151

Kumar A, Barnston A, Hoerling M (2001) Seasonal predictions, probabilistic verifications, and ensemble size. J Clim 14:671–1676

Kumar A, Peng P, Chen M (2014) Is there a relationship between potential and actual skill? Mon Weather Rev 142:2220–2227

Kumar A, Hu Z, Jha B, Peng P (2016) Estimating ENSO predictability based on multi-model hindcasts. Clim Dyn 48:1–13

Lian T, Chen D, Tang YM (2014) Effects of westerly wind bursts on El Niño: a new perspective. Geophys Res Lett 41:3522–3527

Lorenz EN (1965) A study of the predictability of a 28-variable atmospheric model. Tellus 3:321–333

Palmer T, Branković Č, Richardson D (2000) A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. Q J R Meteorol Soc 126:2013–2033

Peng P, Kumar A, Wang W (2011) An analysis of seasonal predictability in coupled model forecasts. Clim Dyn 36:637–648

Penland C, Sardeshmukh PD (1995) The optimal growth of tropical sea surface temperature anomalies. J Clim 8:1999–2024

Rowell DP (1998) Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. J Clim 11(2):109–120

Tang YM, Hsieh W (2002) Hybrid coupled models of the tropical Pacific—II ENSO prediction. Clim Dyn 19:343–353

Tang YM, Kleeman R, Moore A (2005) On the reliability of ENSO dynamical predictions. J Atmos Sci 62:1770–1791

Tang YM, Kleeman R, Miller S (2006) ENSO predictability of a fully coupled GCM model using singular vector analysis. J Clim 19:3361–3377

Tang YM, Kleeman R, Moore A (2008) Comparison of information-based measures of forecast uncertainty in ensemble ENSO prediction. J Clim 21:230–247

Tang YM, Chen D, Yang D (2013) Methods of estimating uncertainty of climate prediction and climate change projection, in Climate Change-Realities, Impacts Over Ice Cap, Sea Level and Risks. InTech, Rijeka

Tang YM, Shen ZQ, Gao YQ (2016) An introduction to ensemble-based data assimilation method in the earth sciences, in nonlinear systems-design, analysis, estimation and control. InTech, Rijeka

Tang YM, Zhang RH, Liu T, Duan WS, Yang DJ, Zheng F, Ren HL, Lian T, Gao C, Chen D, Mu M (2018) Progress in ENSO prediction and predictability study. Natl Sci Rev 5:826–839

Thompson C, Battisti D (2000) A linear stochastic dynamical model of ENSO. Part I: Model development. J Clim 13:2818–2832

Tippett MK, Barnston A (2008) Skill of multimodel ENSO probability forecasts. Mon Weather Rev 136:3933–3946

Tippett MK, Barnston A, Delsole T (2010) Comments on "finite samples and uncertainty estimates for skill measures for seasonal prediction". Mon Weather Rev 138:1487–1493

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

Toth Z, Talagrand O, Candille G, Zhu Y (2003) Probability and ensemble forecasts. In: Jolliffe IT, Stephenson DB (eds) Forecast verification: a practitioner's guide in atmospheric science. Wiley, Hoboken, pp 137–163

Wang B, Lee J-Y, Kang I-S, Shukla J, Park C-K, Kumar A, Schemm J, Cocke S, Kug J-S, Luo J-J (2009) Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). Clim Dyn 33:93–117

Wilks DS (2011) Statistical methods in the atmospheric sciences, Int. Geophys. Ser., vol. 100, 3rd edn. Academic, San Diego

Wilks DS, Vannitsem S (2018) Uncertain forecasts from deterministic dynamics. In: Vannitsem et al (eds) Statistical postprocessing of ensemble forecasts, pp 1–13. https://doi.org/10.1016/B978-0-12-812372-0.00001-7

Xue Y, Cane MA, Zebiak S (1997) Predictability of a coupled model of ENSO using singular vector analysis, Part I: optimal growth in seasonal background and ENSO cycles. Mon Weather Rev 125:2043–2056

Yang DJ, Tang YM, Zhang YC, Yang XQ (2012) Information-based potential predictability of the Asian summer monsoon in a coupled model. J Geophys Res Atmos 117(D3). https://doi.org/10.1029/2011JD016775

Yang DJ, Yang XQ, Xie Q, Zhang YC, Ren XJ, Tang YM (2016) Probabilistic versus deterministic skill in predicting the western North Pacific-East Asian summer monsoon variability with multimodel ensembles. J Geophys Res Atmos 121:1079–1103

Yang DJ, Yang XQ, Ye D, Sun X, Fang J, Chu C, Feng T, Jiang Y, Liang J, Ren XJ, Zhang Y, Tang YM (2018) On the relationship between probabilistic and deterministic skills in dynamical seasonal climate prediction. J Geophys Res Atmos 123:5261–5283

Zavala J, Zhang C, Moore A, Kleeman R (2005) The linear response of ENSO to the Madden–Julian oscillation. J Clim 18:2441–2459

Zebiak S (1986) Atmospheric convergence feedback in a simple model for El Niño. Mon Weather Rev 114:1263–1271

Zebiak S, Cane M (1987) A model El Niño-Southern oscillation. Mon Weather Rev 115:2262–2278

Zhang RH, Zheng F, Zhu J, Wang Z (2013) A successful real-time forecast of the 2010–11 La Niña event. Sci Rep 3:1108

Zheng F, Zhu J (2010) Coupled assimilation for an intermediated coupled ENSO prediction model. Ocean Dyn 60:1061–1073

Zheng F, Zhu J, Zhang RH (2007) Impact of altimetry data on ENSO ensemble initializations and predictions. Geophys Res Lett 34(13). https://doi.org/10.1029/2007gl030451

Zheng F, Zhu J, Wang H, Zhang RH (2009a) Ensemble hindcasts of ENSO events over the past 120 years using a large number of ensembles. Adv Atmos Sci 26:359–372

Zheng F, Wang H, Zhu J (2009b) ENSO ensemble prediction: initial error perturbations vs. model error perturbations. Chin Sci Bull 54:2516–2523

Zhu J, Huang B, Zhang R, Hu Z, Kumar A, Balmaseda M, Marx L, Kinter J III (2014) Salinity anomaly as a trigger for ENSO events. Sci Rep 4:6821