SCIENCE CHINA
Earth Sciences

• RESEARCH PAPER •

January 2014 Vol.57 No.1: 1–10
doi: 10.1007/s11430-014-4991-4

A SVD-based ensemble projection algorithm for calculating the conditional nonlinear optimal perturbation

CHEN Lei[1,2], DUAN WanSuo[1] & XU Hui[1]

[1] *LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China;*
[2] *University of Chinese Academy of Sciences, Beijing 100049, China*

Received February 11, 2014; accepted June 9, 2014

Conditional nonlinear optimal perturbation (CNOP) is an extension of the linear singular vector technique in the nonlinear regime. It represents the initial perturbation that is subjected to a given physical constraint, and results in the largest nonlinear evolution at the prediction time. CNOP-type errors play an important role in the predictability of weather and climate. Generally, when calculating CNOP in a complicated numerical model, we need the gradient of the objective function with respect to the initial perturbations to provide the descent direction for searching the phase space. The adjoint technique is widely used to calculate the gradient of the objective function. However, it is difficult and cumbersome to construct the adjoint model of a complicated numerical model, which imposes a limitation on the application of CNOP. Based on previous research, this study proposes a new ensemble projection algorithm based on singular vector decomposition (SVD). The new algorithm avoids the localization procedure of previous ensemble projection algorithms, and overcomes the uncertainty caused by choosing the localization radius empirically. The new algorithm is applied to calculate the CNOP in an intermediate forecasting model. The results show that the CNOP obtained by the new ensemble-based algorithm can effectively approximate that calculated by the adjoint algorithm, and retains the general spatial characteristics of the latter. Hence, the new SVD-based ensemble projection algorithm proposed in this study is an effective method of approximating the CNOP.

Singular vector decomposition, Ensemble projection algorithm, ENSO, Conditional nonlinear optimal perturbation

Citation: Chen L, Duan W S, Xu H. 2014. A SVD-based ensemble projection algorithm for calculating the conditional nonlinear optimal perturbation. Science China: Earth Sciences, 57: 1−6, doi: 10.1007/s11430-014-4991-4

Determining the fastest growing initial error is a key issue in the numerical forecasting of weather and climate, and is effectively accomplished by an optimization method. Since Lorenz (1965) first introduced the concept of linear singular vectors (LSVs) into meteorology to investigate the problem of the fastest growing initial error, they have been widely applied to the study of predictability in weather and climate (Palmer et al., 1998; Buizza et al., 1999). However, LSVs represent the direction of fastest growth for initial perturbations in a linear model. In the case of a nonlinear model, LSV is a reasonable approximation of the fastest growing initial perturbation when the amplitude of perturbation is sufficiently small and the forecast time is short. That is, LSVs cannot determine the fastest growing initial perturbations in the nonlinear model,nor can they reveal the effect of nonlinear physical processes imposed on the evolution of the initial error.

To overcome these limitations of LSVs, Mu et al. (2003) proposed the concept of conditional nonlinear optimal perturbation (CNOP). CNOP is characterized by the initial perturbation that satisfies given constraint conditions and results in the largest nonlinear evolution at the prediction time. Physically, CNOP represents the initial error that in-

*Corresponding author (email: xuh@lasg.iap.ac.cn)

duces the largest forecast at the prediction time. Mu et al. (2003, 2007), Duan et al. (2004, 2009, 2012), and Yu et al. (2009, 2012) used CNOP to investigate problems concerning the optimal precursor of El Niño, the fastest growing initial error, spring predictability barrier, and so on. Their results revealed the impact of nonlinear physical processes on the predictability of El Niño, and supplied useful information for improving the accuracy of forecasting of this phenomenon. CNOP has also been applied to the determination of sensitive areas in adaptive observations of typhoons (Qin et al., 2011, 2013) and the transition of climatology in grassland ecosystems (Sun et al., 2011). In brief, CNOP is gradually increasing in importance in the study of the predictability of weather and climate.

The application of CNOP first requires the computation of the optimal perturbation. That is, mathematically, we must find the global maximum value of a nonlinear functional. An appropriate nonlinear optimization algorithm is then necessary. There are two main types of optimization algorithm: traditional algorithms that calculate the gradient of the objective function (such as Conjugate Gradient, LBFGS, and SPG2), and intelligent algorithms that do not require gradient information (such as particle swarm optimization and genetic algorithms). However, the number of degrees of freedom in a complicated atmosphere-ocean coupled numerical model is generally around 10^6–10^7, and the cost of computing the CNOP using intelligent algorithms is almost unacceptable. Of the traditional optimization algorithms, the adjoint method is the most efficient means of numerically calculating the gradient. Nevertheless, many numerical models have no corresponding adjoint model. In particular, coding the adjoint model of a complex numerical model is a large, tedious, time-consuming job, which limits the further application of CNOP.

To overcome the above difficulties in computing the CNOP, Wang et al. (2010) made use of an ensemble projection algorithm (the Monte Carlo method) to approximate the tangent linear matrix, calculated the gradient with respect to initial perturbations, and then computed the CNOP. The algorithm is adjoint-free in terms of calculating the gradient, and has relatively high computational efficiency, which enhances the applicability of CNOP. A localization technique, which is a key component of ensemble-based algorithms, is used to assess spurious correlations between observation stations and model grids. During this process, the covariance of the localization Schur radii should be determined artificially and empirically. Based on the algorithm proposed by Wang et al. (2010), we propose a new singular vector decomposition (SVD)-based ensemble projection(EP) algorithm. The algorithm constructs the CNOP using the main modes of historical time series of related variables in numerical models. Thus, the issue of searching for the CNOP is transformed into that of determining the optimal weight coefficient combination of the main modes of SVD. This transformation not only considers the physical meaning of the CNOP, but also avoids the empirical choice of localization radius in Wang et al. (2010).At the same time, our proposed method reduces the degree of freedom of the optimization problem, because the control variables are the weight coefficients of the main modes of SVD instead of physical variables.

The remainder of this paper is organized as follows. In Section 1, the concept of CNOP and the EP algorithm proposed by Wang et al. (2010) are briefly described. The proposed SVD-based EP algorithm is introduced in Section 2. In Section 3, the new algorithm is applied to an El Niño–Southern Oscillation (ENSO) forecasting model of medium complexity, the Zebiak-Cane (ZC) model (Zebiak et al., 1987), and the validity of the new algorithm is investigated by comparing its results with those given by the adjoint and LSV methods. Finally, we discuss our results and summarize our conclusions in Section 4.

1 CNOP method and EP algorithm

1.1 CNOP method

CNOP represents an initial perturbation subjected to a given physical constraint, and results in the largest nonlinear evolution at the prediction time. CNOP $x_{0\delta}$ is the solution of the following optimization problem

$$J(x_{0\delta}) = \max_{\|x_0\|_\delta \le \delta} \left\| \mathbf{M}(X_0 + x_0, t_0, t) - \mathbf{M}(X_0, t_0, t) \right\|^2, \qquad (1)$$

where X_0 is the initial field of the reference state, x_0 is the perturbation of X_0, t_0 and t are the initial optimization time and terminal time, respectively, δ is the constraint radius of the initial perturbation, and $\boldsymbol{M}$ is a nonlinear propagation operator. The CNOP can be obtained using a nonlinear optimization algorithm. Generally, optimization algorithms, such as LBFGS, SQP, and SPG2, are used to find the minimum value of the objective function. To conveniently apply existing optimization algorithms, let $J_1(x_0)=-J(x_0)$ and we can solve the equivalent optimization problem

$$J_1(x_{0\delta}) = \min_{\|x_0\|_\delta \le \delta} J_1(x_0). \qquad (2)$$

For traditional nonlinear optimization algorithms, the key is to obtain the gradient of the objective function with respect to the initial perturbation x_0. For the above objective function, the gradient with respect to x_0 is $-2\mathbf{H}^T(\mathbf{M}(X_0 + x_0, t_0, t) - \mathbf{M}(X_0, t_0, t))$, where $\boldsymbol{H}^{\mathrm{T}}$ is the adjoint model of the tangent linear model $\boldsymbol{H}$. Constructing the adjoint of a complex numerical model is very time-consuming, limiting the applicability of CNOP. To overcome this difficulty, Wang et al. (2010) developed an EP-based algorithm to compute CNOP. Their method is adjoint-free when calculating the gradient, and this means that the CNOP method can be employed conveniently.

1.2 EP algorithm

The primary idea of the EP algorithm of Wang et al. (2010) is to approximate the tangent linear matrix ***H*** in the gradient formula using an ensemble method, and then compute the CNOP. The tangent linear matrix in the algorithm is approximated as

$$\mathbf{H} \approx p_y (p_x^T p_x)^{-1} p_x^T, \tag{3}$$

where $p_x = (x_0^1, x_0^2, \cdots, x_0^n)$ denotes the n sample perturbations at the initial optimization time, and p_y=(y^1, y^2, $\cdots$, y^n) denotes the prediction increments at the prediction time corresponding to p_x. Houtekamer et al. (2001) pointed out that, if an ensemble is composed of far fewer samples than the number of physical variables in the model (grid numbers multiplied by number of variables), there exist spurious long-distance correlations between the initial perturbations and the corresponding prediction increments on the model grid. To ameliorate the spurious correlations, a localization technique is adopted.

Wang et al. (2010) approximate the tangent linear matrix as $\mathbf{H} \approx \rho \circ p_y (p_x^T p_x)^{-1} p_x^T$, where $\circ$ denotes the Schur product of two matrices. The elements of matrix ρ are $\rho_{i,j} = C_0(d_{i,j}^h / d_0^h) \times C_0(d_{i,j}^v / d_0^v)$, where $d_{i,j}^h$, d_0^h, $d_{i,j}^v$, and d_0^v represent the horizontal distance, horizontal localization radius, vertical distance, and vertical localization radius, respectively. C_0 is a monotonously decreasing smooth filter function given by

$$C_0(r) = \begin{cases} -\dfrac{r^5}{4} + \dfrac{r^4}{2} + \dfrac{5r^3}{8} - \dfrac{5r^2}{3} + 1, & 0 \le r \le 1, \\ \dfrac{r^5}{12} - \dfrac{r^4}{2} + \dfrac{5r^3}{8} + \dfrac{5r^2}{3} - 5r + 4 - \dfrac{2r^{-1}}{3}, & 1 \le r \le 2, \\ 0, & r \ge 2, \end{cases} \tag{4}$$

where r is either $d_{i,j}^h / d_0^h$ or $d_{i,j}^v / d_0^v$. It can be seen from eq. (3) that the approximation of the tangent linear matrix ***H*** obtained by the EP algorithm depends on the choice of localization radius. However, the size of the latter depends, to some extent, on the quality and quantity of samples. These factors mean that the approximation of the tangent linear matrix is affected by artificial experiences. In view of the above limitations, this paper proposes a new EP algorithm based on SVD. Using the new algorithm, the CNOP not only reflects the dynamical characteristics of the numerical models, but is also independent of the choice of localization radius.

2 SVD-based EP algorithm

A forced, dissipative dynamical system tends toward a low-dimensional attractor after long evolution (Teman et al., 1991; Osborne et al., 1993; Foias et al., 1997). Such a system can be expressed as

$$X = \sum_{i=1}^{\infty} a_i(t) u_i, \tag{5}$$

where X is the state space of some physical variables in the model and u_i is the primary function of the spatial modes in Hilbert space. As $i \to \infty$, $a_i(t) \to 0$ with probability μ.

The projection of a continuous system into a discrete numerical model can be expressed as

$$\mathbf{M}(X^*, t^*, t) = \sum_{i=1}^{N} \sigma_i u_i v_i^T, \tag{6}$$

where N is the degree of freedom of a numerical model, σ_i are singular values arranged from highest to lowest, u_i is the spatial mode corresponding to σ_i, and v_i is the time series of u_i. There exists an integer m ($m \le N$) such that when $i \ge m$, $|\sigma_i| < \varepsilon$, (ε is a small positive number), and then the original N-dimensional system can be truncated to an m-dimensional approximate system.

To investigate the degree of approximation in the m-dimensional truncated system, we use a primary inequality from probability theory

$$P\left\{\left| x - \sum_{i=1}^{m} a_i u_i \right| > \varepsilon \right\} = \left\{ P\left| \sum_{i=m+1}^{N} a_i u_i \right| > \varepsilon \right\} < \frac{\sum_{i=m+1}^{N} \sigma_i^2}{N\varepsilon^2}. \tag{7}$$

This gives the approximate information loss caused by the reduction of dimension of the original system. Strictly speaking, the formula is only applicable in the case of infinite dimensional ($N \to \infty$) systems. However, eq. (7) shows that if the spatial modes u_i are chosen such that $|\sigma_i|$ monotonously decreases quickly enough as i increases, the m spatial modes can be used as the bases to construct the approximate state space of the whole system. To solve this problem, SVD statistically provides a standard method that reduces the dimension of the system by effectively choosing spatial modes.

If m spatial modes are chosen and combined linearly to approximate the state vector of the discrete system, the objective function with regard to the optimization problem (2) is then transformed into the form

$$\begin{cases} J_1(x_{0\delta}) = \min\limits_{\|x_0\|_\delta \le \delta} (-\left\| \mathbf{M}(X_0 + x_0, t_0, t) - \mathbf{M}(X_0, t_0, t) \right\|^2), \\ x_0 = \sum\limits_{i=1}^{m} a_i u_i, \end{cases} \tag{8}$$

where m is the truncated number of chosen bases and a_i is the weight coefficient corresponding to the chosen base u_i. This searches for the optimal combination of weight coefficients for the chosen bases.

Let $F(a)=\left\|\mathbf{M}\left(X_0+\sum_{i=1}^{m}a_iu_i,t_0,t\right)-\mathbf{M}(X_0,t_0,t)\right\|$. Then, the optimization problem (8) is transformed into

$$J_1(a_\delta)=\min_{\|a\|_\delta\le\delta}(-F(a)^2). \tag{9}$$

The gradient of the objective function with respect to the coefficients is

$$\frac{\partial(-F(a)^2)}{\partial a}=-2L^TF(a). \tag{10}$$

The tangent linear matrix L of the model with respect to the coefficients satisfies

$$\begin{cases}(\Delta y_1,\cdots,\Delta y_i,\cdots,\Delta y_m)\approx L\mathrm{diag}(\Delta a_{1,1},\cdots,\Delta a_{i,i},\cdots,\Delta a_{m,m}),\\ \Delta y_i=\mathbf{M}(X_0+\Delta a_{i,j}u_j,t_0,t)-\mathbf{M}(X_0,t_0,t).\end{cases} \tag{11}$$

To compute the inverse efficiently, we set $\Delta\bar{a}_{i,j}=\delta_{i,j}$. That is to say, the gradient is approximated by the differences $(\Delta y_1,\cdots,\Delta y_i,\cdots,\Delta y_m)$ and $(\Delta a_{1,1},\cdots,\Delta a_{i,i},\cdots,\Delta a_{m,m})$. The computational cost of calculating the gradient in this manner depends on the number of chosen bases, and is generally greater than that using the adjoint method. However, to some extent, the bases can be chosen to reduce the computational cost.

Based on the above idea for calculating the gradient, the SPG2 algorithm proposed by Birgin et al. (2000) is used to calculate the CNOP by finding the optimal combination of weight coefficients for the chosen bases.

3 Application of SVD-based EP algorithm to calculation of CNOP

In this section, we consider the optimal precursor of an ENSO event to investigate the effectiveness of our SVD-based EP algorithm for calculating CNOP. That is, we seek the pattern of initial sea surface temperature anomalies (SSTA) and thermocline height anomalies (THA) that cause the largest evolution in SSTA in the tropical Pacific, thus inducing an ENSO event. To find the optimal precursor of ENSO, the SVD-based EP and adjoint methods are utilized to calculate the CNOP of the climatological state. The validity of the CNOP calculated by SVD-based EP is checked by comparing the precursors and evolution of SSTA obtained by the two methods. The LSV method is also used to investigate the optimal precursor of ENSO reported in previous studies (Blumenthal, 1991; Thompson et al., 1995; Moore et al., 1996). These three methods are applied to the medium-complexity ZC model.

3.1 Model and numerical experiment Scheme

The ZC model successfully predicted the El Niño event of 1986–1987, and has since been extensively applied to the research of the predictability and dynamics of ENSO (Mantua et al., 1995; Thompson, 1998; Cai et al., 2003). Recently, the ZC model was used to investigate the dynamics of error evolution in ENSO (Mu et al., 2003, 2007; Duan et al., 2004, 2009, 2012; Yu et al., 2009, 2012) using the CNOP method. In these studies, CNOP is calculated by the adjoint method, which limits its application in other complex numerical models. To extend the applicability of CNOP, we attempt to calculate CNOP in the ZC model using the SVD-based EP algorithm.

The ZC model is composed of three modules. The atmosphere module covers the area of 101.25°E–73.125°W and 29°S–29°N, the ocean dynamics extend over 124°E–80°W and 28.75°S–28.75°N, and the SSTA equation covers 129.375°E–84.375°W and 19°S–19°N. The resolution of the ocean dynamical module is 2°×0.5°, and the atmosphere module and SSTA equation have a resolution of 5.625°×2°. We initially consider SSTA and THA over the area 129.375°E–84.375°W and 19°S–19°N. The related variables form 1080-dimensional vectors.

For the optimal precursor of ENSO, consider the following problem

$$J(x_{0\delta})=\max_{\|x_0\|_\delta\le\delta}\|T_\tau\|^2, \tag{12}$$

where $x_0=(w_1T_0,w_2H_0)$ is the non-dimensional SSTA and THA ($w_1=(2.0°\mathrm{C})^{-1}$ and $w_2=(50\ \mathrm{m})^{-1}$ are the reciprocals of the respective characteristic scales). $\|x_0\|_\delta\le\delta$ is the constraint condition, and this is set as $\|x_0\|_\delta\le1.0$. $\|x_0\|_\delta=\sqrt{\sum_{i,j}[(w_1T_{0\,i,j})^2+(w_2H_{0\,i,j})^2]}$, where $(i,\ j)$ represents the indices of grid cells located in the given area, and $T_{0\,i,j}$, $H_{0\,i,j}$ represent the initial SSTA and THA in cell (i, j), respectively. The evolution of SSTA is measured as $\|T_\tau\|=\sqrt{\sum_{i,j}(w_1T_{\tau\,i,j})^2}$, where $T_{\tau\,i,j}$ is the evolution of SSTA at prediction time τ. The optimal initial fields $x_{0\delta}$ can be found such that the objective function gives the largest evolution at the prediction time.

Before applying the new EP algorithm to calculate the CNOP, we verify whether the long-time integrated behavior of the model can be approximately represented by a finite number of stable modes. We set all initial anomalies of to zero, apply wind forcing to the model for the first 5 months, and run the model freely for 200 years. This gives a group of 200-year samples. SVD is carried out over the scaled samples, and a series of spatial modes u_i (i=1,⋯, m) are obtained. These modes are mutually orthogonal with unit norms. Hence, the constraint condition for the perturbations of the initial fields can be expressed as $\|a\|_\delta\le1.0$, where a represents the weight coefficient vector of the linearly com-

bined spatial modes that approximate CNOP. On this basis, we calculate the gradient of the objective function with respect to the weight coefficients, and then obtain the CNOP.

Figure 1 shows the distribution of the squares of the first 100 singular values. It can be seen that the singular value deceases rapidly as the mode number increases. According to the discussions in Section 3, a small number of modes can be selected to approximate the whole sample space.

3.2 Numerical results

To verify the effectiveness of our SVD-based EP algorithm, the proposed algorithm is applied to a 9-month optimization period with different initial months and numbers of bases. The results are compared with those obtained by other methods.

3.2.1 Comparison between maximumvalues of the objective function

We first compare the norm of the SSTA evolution (that is, the maximum value of the objective function) given by the CNOPs of the new EP algorithm and adjoint method.

Figure 2 shows the ratio of the optimal value J^* of the

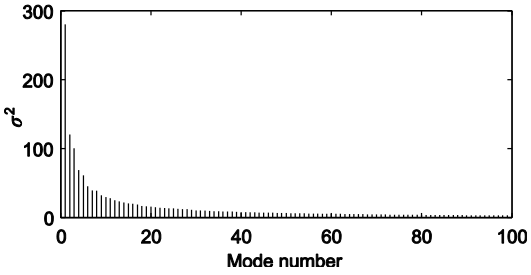

Figure 1 Squares of the first 100 singular values obtained by SVD in the long-term integration series of the ZC model (σ is the singular value).

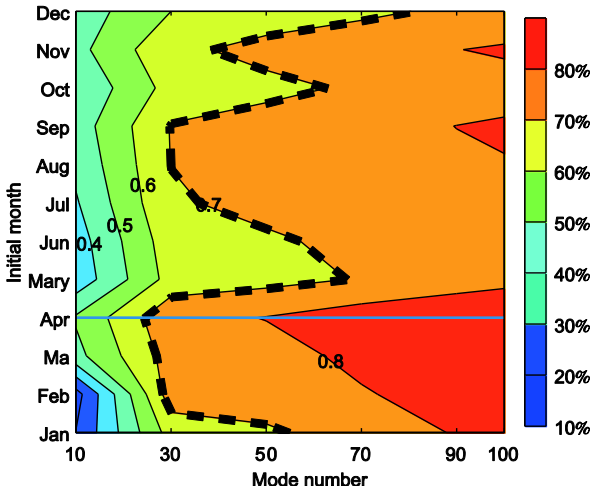

Figure 2 Ratio of the objective function values obtained by the new EP algorithm and the adjoint algorithm. Horizontal axis represents the mode number and vertical axis represents the initial optimization month. Thick dashed line denotes the isoline of 70%. The blue solid line represents the case of April as the initial optimization month.

objective objection calculated by the new EP algorithm to J, which is the optimal value given by the adjoint method, for different numbers of bases. The larger the ratio, the better the approximation. The horizontal axis is the number of bases used in the new algorithm, and the vertical axis shows the initial optimization month. The thick dashed line denotes the 70% isoline.

It can be seen from Figure 2 that, for all initial optimization months, the maximum value of the objective function in the new algorithm becomes closer to that obtained by the adjoint method as the number of modes increases. When this number exceeds 80, the ratio of the former to the latter is no less than 70%. For different initial optimization months, as the number of the chosen bases increases from 10 to 30, the approximation effect is clearly improved. Nonetheless, when the number continues to increase from 30 to 100, the speed of improvement becomes slow, and the computational cost increases rapidly. The horizontal blue line denotes the best approximation effect for an initial optimization month of April. When the base number increases from 10 to 30, the degree of approximation increases from 40% to 70%. However, as the number of bases rises toward 100, the degree of approximation remains below 90%. This suggests that, by choosing a reasonable number of bases, the CNOP obtained by the new EP algorithm can lead to relatively large developments in SSTA at the prediction time. The more bases are chosen, the better the norm of the SSTA evolution approximates that given by the adjoint method.

It can also be seen from Figure 2 that, for different initial optimization months, different numbers of bases are needed to achieve the same approximation effect. Taking the 70% effect for example, 30 modes are needed when the initial optimization is in February, March, April, August, or September, 40–70 modes are needed for January, May, June, July, October, and November, and 80–90 modes are required for December. In the following analysis, we examine the cases of April (30 modes) and October (70 modes) to investigate the similarities and differences in the maximum objective function values, the precursor signals of ENSO and nonlinear SSTA evolutions, given by several algorithms.

The LSV method is often used to study the precursors of ENSO. Thus, we compare the numerical results obtained by the new algorithm with those given by LSV, and then compare the degree of approximation of the optimal precursors obtained by these two methods to those produced by the adjoint method. Figure 3 shows that, from an initial optimization month of April, the norms of the SSTA evolutions at the prediction time vary with the different constraint radii. Comparing the nonlinear evolution of the optimal precursors, we can see that thenorm of the SSTA evolution is given by the precursor obtained by the LSV method are closer to those of the adjoint method for smaller constraint radii. The results of the new EP algorithm are better than those

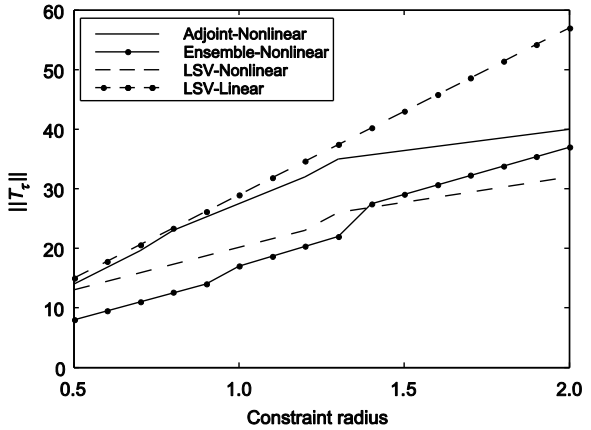

Figure 3 Norms of the SSTA evolutions at the prediction time based on SVD-based EP, adjoint method, and LSV for initial optimization time of April. Horizontal axis is the constraint radius. Black solid line denotes the results based on adjoint method, black solid line marked by black dots denotes the results based on SVD-based EP algorithm (with 30 bases), black dashed line shows nonlinear evolution of LSV, and black dashed line marded by black dots corresponds to the linear evolution of LSV.

given by LSV at larger constraint radii. This is because of the increasing nonlinearity in the ZC model. Thus, the validity of the maximum objective function value obtained by the new EP algorithm is no worse than that of LSV. It is worth pointing out that the number of chosen bases in the new algorithm is 30, and the approximation can be improved by increasing the base number. Additionally, when calculating CNOP with the new EP algorithm, the tangent linear model and adjoint model are not needed, which extends the scope of applications and increases the flexibility of the CNOP.

3.2.2 Comparison between spatial patterns of optimal precursors and SSTA wvolutions

The maximum objective function values output by the new EP algorithm, adjoint method, and LSV can be used to assess the effectiveness of the new algorithm. In this section, the spatial pattern of the ENSO precursors and corresponding SSTA evolutions based on the three methods are compared to assess the validity of the new algorithm.

Figures 4 and 5 depict the optimal precursors and corresponding 9-month SSTA evolutions for a constraint radius of 1.0 and initial optimization months of April and October, respectively. We can see that all three precursors evolve into an El Niño event. The intensity of the El Niño event predicted by the adjoint method is the strongest, followed by that of LSV, and the El Niño obtained by the new EP algorithm is the weakest. The results agree with those in Figure 3, which demonstrates that the new algorithm can be used to determine the optimal precursors of ENSO, and these optimal precursors can evolve into an El Niño event.

It can also be seen from the figures that the precursors given by the new EP algorithm retain the main large-scale characteristics of those from the adjoint method. Both methods give spatial patterns of the SSTA components that show a zonal dipole with positive anomalies in the eastern tropical Pacific and negative anomalies in the central tropical Pacific, and THA components that show a uniform deepening across the whole equatorial Pacific.

However, there are some differences between the THA components of the optimal precursors obtained by the new method and the adjoint method. The THA given by the adjoint method has more small-scale signals. The new algorithm gives relatively smoother results and larger anomalies at higher latitudes in the tropical Pacific. This may be because the CNOP is constructed by the selected modes with significant explained variance. This truncation makes the new algorithm adjoint-free, but has the limitation of filtering out some contributive signals with low explained variance and transferring this energy to the selected modes with high explained variance. However, the results of the new EP algorithm catch the large-scale features of the optimal precursors in the ZC model, and outline the main structures of the CNOP.

To verify our conjecture and further explore the similarities and differences in the CNOPs, Figure 6 compares the projection coefficients of CNOPs on the selected bases for the new algorithm and the adjoint method. The projection coefficients of CNOPs on the selected bases have similar distributions, but the CNOP of the new algorithm possesses coefficients with larger absolute values than those of the adjoint method. Thus, the CNOP of the new algorithm indeed occupies more energy across the chosen bases, and the CNOP of the adjoint method allocates this energy to the modes truncated by the new EP algorithm. The lower this energy, the better the approximation of the new ensemble method.

Therefore, the coefficient distribution of the CNOP from the adjoint method is investigated with respect to the bases discarded by the new algorithm. Figure 7 represents the coefficient distribution of the CNOP based on the adjoint method on all 1080 modes for initial optimization months of April and October. It can be seen that the coefficients of the selected modes (30 for April, 70 for October) are much larger than those of the abandoned modes, and the selected modes have much more energy than the abandoned modes.

We now compare the numerical results given by the two methods when all 12 months are used as the initial optimization time (see Table 1). It can be seen that whichever month is taken as the initial optimization time, most of the energy of the CNOP based on the adjoint method is distributed on the selected modes with larger explained variances, and less energy is allocated on the discarded modes. This leads to relatively high similarity between the spatial patterns of the CNOPs obtained by the new algorithm and the adjoint method. Accordingly, the maximum value of the objective function based on the former is an effective approximation to that based on the latter, which conforms to the above analysis.

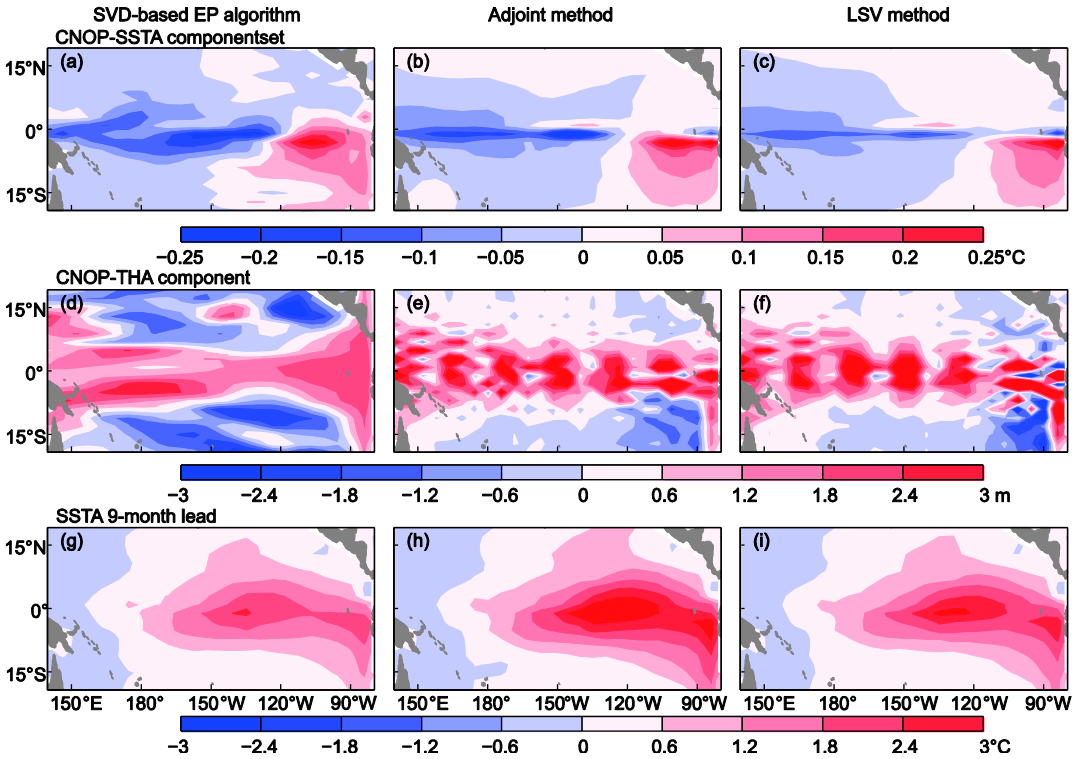

Figure 4 CNOP patterns and corresponding SSTA evolutions obtained by the SVD-based EP method (left column), adjoint method (middle column), and LSV method (right column) with April as the initial optimization time. (a), (b) and (c) SSTA component of CNOP; (d), (e) and (f) THA component of CNOP; (g), (h) and (i) Nonlinear evolution of SSTA after 9 months.

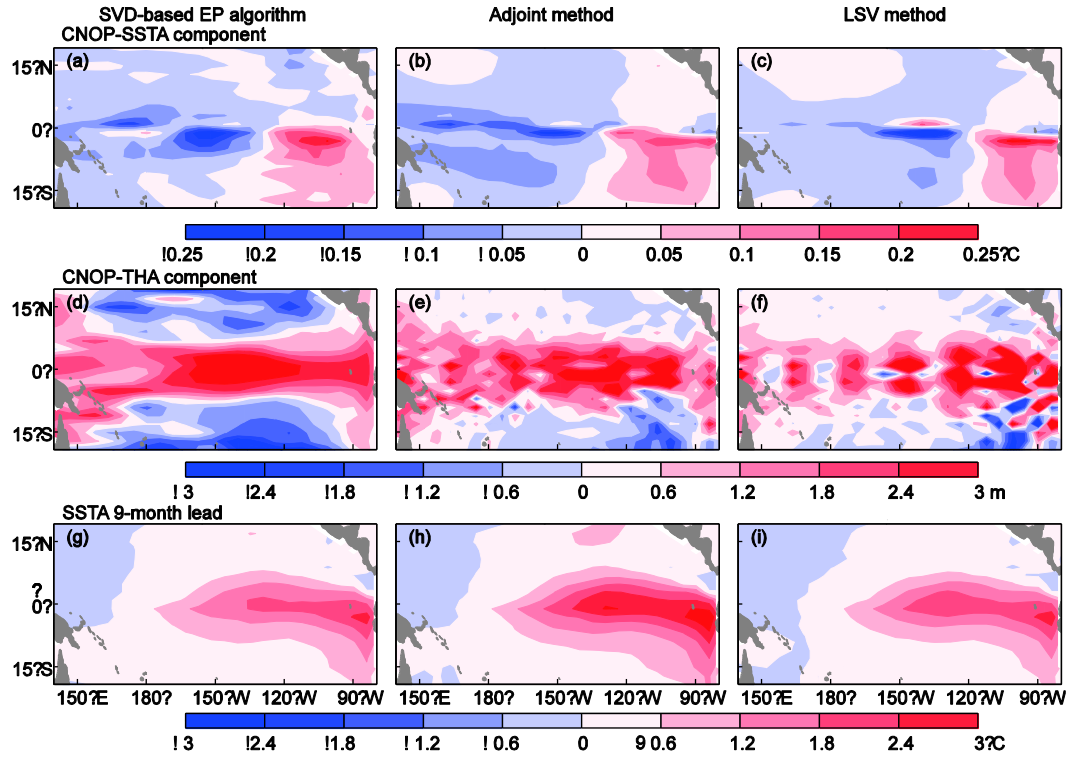

Figure 5 As in Figure 4, but with initialoptimization month of October.

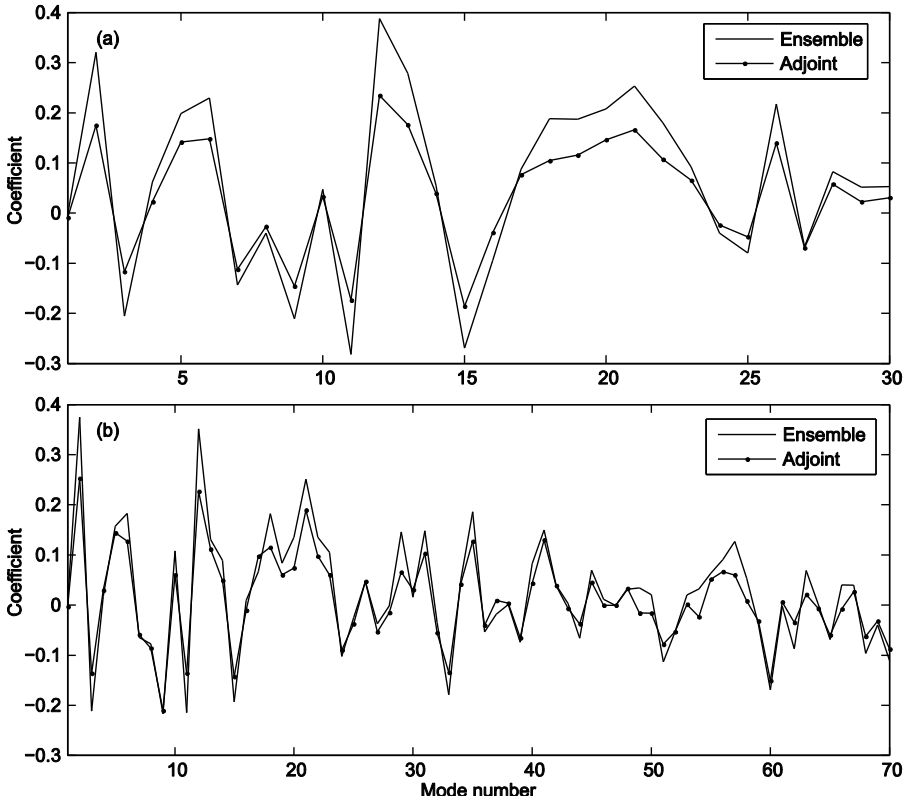

Figure 6 Projection coefficients of CNOP on the chosen modes for the SV-based EP method (solid line) and the adjoint method (solid line marked by black dots). (a) Initial optimization monthof April; (b) initial optimization month of October.

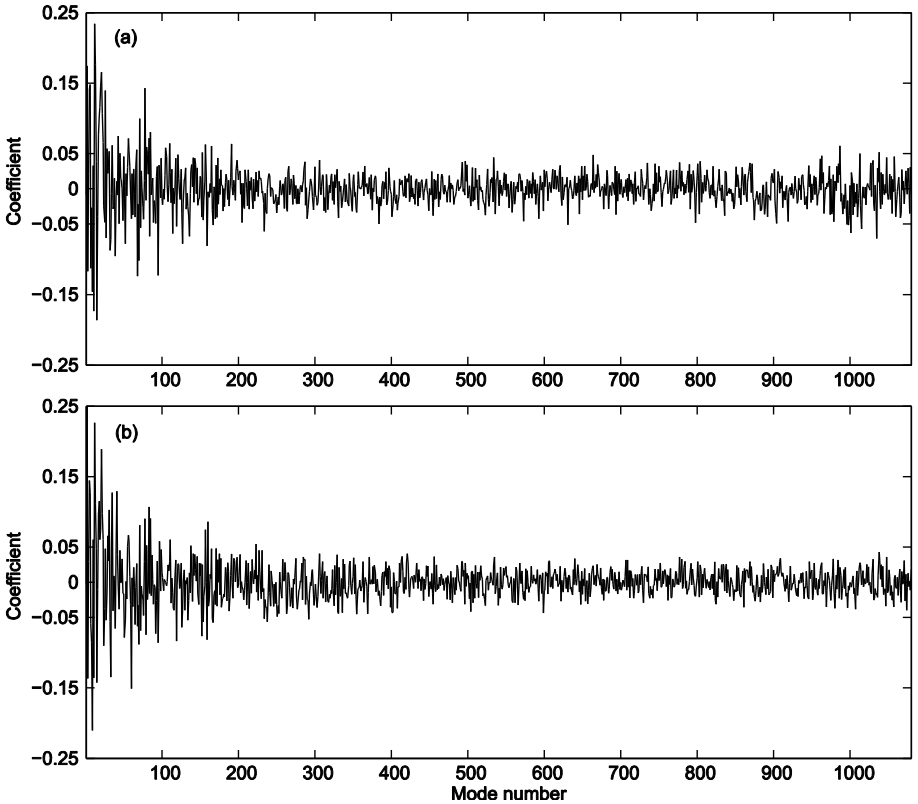

Figure 7 Projection coefficients of the CNOP based on the adjoint method for all 1080 modes. (a) Initial optimization monthof April; (b) initial optimization month of October.

Table 1 Degree of similarity between the results based on SVD-based EP method and adjoint method [a)]

Initial time	Number of chosen bases	Ratios	Similarity coefficients	J^*/J
Jan	70	72%	72%	75%
Feb	30	68%	68%	71%
Mar	30	68%	68%	72%
Apr	30	63%	63%	73%
May	70	69%	69%	73%
Jun	60	69%	69%	70%
Jul	50	70%	70%	73%
Aug	30	66%	66%	70%
Sep	30	65%	65%	70%
Oct	70	73%	73%	71%
Nov	40	64%	64%	71%
Dec	90	73%	73%	72%

a) Initial optimization time; Number of bases chosen for the new algorithm; Ratio of the energy of the CNOP obtained by the adjoint method on the chosen bases to the total energy on all 1080 modes; Similarity coefficients of CNOPs based on SVD-based EP method and adjoint method; Ratio of the maximum values of the objective functions calculated by two methods.

Considering the maximum values of the objective functions, optimal precursors, and SSTA evolutions at the prediction time, this section compares the numerical experimental results of the new SVD-based EP algorithm, adjoint method, and LSV method, and estimates the effectiveness of the new approach. The results show that the CNOP given by the new algorithm is a good approximation of the CNOP from the adjoint method, indicating that the SVD-based EP algorithm produces a valid calculation of the CNOP. It should be pointed out that the new algorithm uses a difference method to compute the gradients directly, and has a longer calculation time than the adjoint method. However, considering the difficulties and time required to codean adjoint model, using the new algorithm to calculate CNOP is an acceptable and feasible method that extends the applicability of the CNOP.

4 Discusssions and conclusions

This study proposed a new SVD-based EP algorithm. The new algorithm avoids the localization procedure in previous EP algorithms, and overcomes the uncertainty caused by empirically choosing the localization radius. The main idea of the new algorithm is to conduct SVD on the long historical data of the system, and determine a group of bases that reflects the dynamics of the system. Thus, physical variables are treated as the combination of selected bases, and the optimization is transformed into one concerning the combination coefficients. For a forced dissipative dynamical system, a small number of bases can reflect the dynamics of the system. Hence, the dimension of the system can be reduced, and the CNOP can be obtained by the optimization process.

In an example regarding the precursors of ENSO, the SVD-based ensemble projection algorithm was used to calculate the CNOP of the medium-complexity ZC model. The numerical experimental results show that, for different initial optimization months, the maximum values of the objective function obtained by the new algorithm gradually became closer to those obtained by the adjoint method as the number of chosen modes increased. For different initial months, the new algorithm requires different numbers of modes to achieve a reasonable degree of approximation. For example, the 70% approximation level may require from 30 to 90 bases. By comparing the spatial patterns of the CNOPs and corresponding SSTA evolutions, we showed that the new algorithm retains the main large-scale features of the precursors given by the adjoint method. For the SSTA component of CNOP, the patterns show a zonal dipole with positive anomalies in the eastern tropical Pacific and negative anomalies in the central tropical Pacific. The THA component exhibits a uniform deepening across the whole equatorial Pacific. These initial anomalies finally evolve into an El Nino event. These results illustrate that the new algorithm effectively calculates the CNOP, and its adjoint-free advantage extends the applicability of the CNOP to more complex ocean-atmosphere coupled models for predictions of weather and climate.

The new algorithm selects some bases with larger explained variances when constructing the CNOP. This reduces the dimensionality of the system, leading to an adjoint-free method of calculating the CNOP. However, some contributive signals with small explained variances may be filtered out and their energies transferred to the chosen bases with larger explained variances. This leads to differences between the results given by the two methods. Practically, as long as the computational cost remains acceptable, we can increase the number of selected modes so that the transferred energies of the discarded modes occupy a smaller proportion of the total energy. In this way, the degree of approximation of the new algorithm to the adjoint method can be improved.

The application of the new algorithm is directly related to the nature of the investigated dynamical system. If the attractor of the system is of low dimension, the approximation of the new algorithm to the adjoint method is more effective. Hence, for some rapidly changing phenomena, if the attractor is low dimensional and the system converges quickly to a stationary state, the new algorithm is applicable. But other rapid processes may not have a structure with simple attractors, in which case more bases are needed to construct the CNOP and the computational load increases. Moreover, in this case, the approximation of the new algorithm to the adjoint method is not satisfactory. Thus, it may be difficult to apply the new method to systems that contain such rapid processes.

The number of modes chosen for the SVD-based EP algorithm influences its computational cost. How many mod-

es are needed is related not only to the degree of complexity of the attractors, but also to characteristics of the numerical models, such as the resolution and discretization scheme. However, the number of modes required for the new method is not simply proportional to the number of variables as the resolution of the model is enhanced. For example, a high-resolution model that is four times finer than a low-resolution model in each direction will have 64 times as many variables. For an equal percentage of explained variances, the number of modes required in the high-resolution model may be more than that in the low-resolution model, but the former needs far fewer than 64 times that of the latter. That is to say, as the resolution of the model is enhanced, though the number of required bases may also increase, the amplitude of this increase is much less than that of the variables.

In particular, if a numerical model is a relatively perfect description of a continuous system, then modes with large explained variance are determined by the continuous system itself, and do not depend on the model resolution. In this case, we can choose these bases to construct the CNOP. For this type of high-resolution model, the increase in computational cost of the new algorithm is limited, and the new method is suitable for calculating the CNOP.

It is important to select appropriate bases in our algorithm, and a series of tests are required to produce these bases when using large samples. Hence, choosing samples more effectively for specific physical problems is an important issue for future studies. Additionally, when calculating the gradient with respect to the initial fields, the number of nonlinear models is equal to the number of chosen modes, which leads to a relatively large computational cost. Therefore, choosing as few modes as possible and using parallel computation will significantly reduce time consumption.

This work was jointly sponsored by the National Natural Science Foundation of China (Grant Nos. 41176013, 41230420 and 41006007).

Birgin E G, Martínez J M, Raydan M. 2000. Nonmonotone spectral projected gradient methods on convex sets. SIAM J Opt, 10: 1196–1211

Blumenthal M B.1991. Predictability of a coupled ocean-atmosphere model. J Clim, 4: 766–784

Buizza R, Montani A. 1999. Targeting observations using singular vectors. J Atmos Sci, 56: 2965–2985

Cai M, Kalnay E, Toth Z. 2003. Bred vectors of the Zebiak-Cane model and their potential application to ENSO predictions. J Clim, 16: 40–56

Duan W S, Mu M, Wang B. 2004. Conditional nonlinear optimal perturbation as the optimal precursors for El Niño-Southern Oscillation events. J Geophys Res, 109: 1984–2012

Duan W S, Xue F, Mu M. 2009. Investigating a nonlinear characteristic of ENSO events by conditional nonlinear optimal perturbation. Atmos Res, 94: 10–18

Duan W S, Yu Y S, Xu H, et al. 2012. Behaviors of nonlinearities modulating El Niño events induced by optimal precursory disturbance. Climate Dyn, 40: 1399–1413

Foias C, Teman R. 1997. Structure of the set of stationary solution of the Novier-Stokes equations. Commun Pure Appl Math, 30: 149–164

Houtekamer P L, Mitchell H L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon Weather Rev, 129: 123–137

Lorenz E N. 1965. A study of the predictability of a 28-variable atmospheric model. Tellus, 17: 321–333

Mantua N J, Battisti D S. 1995. Aperiodic variability in the Zebiak-Cane coupled ocean-atmosphere model: Air-sea interactions in the western equatorial Pacific. J Clim, 8: 2897–2927

Moore A M, Kleeman R. 1996. The dynamics of error growth and predictability in a coupled model of ENSO. Quart J Roy Meteor Soc, 122: 1405–1446

Mu M, Duan W S, Wang B. 2003. Conditional nonlinear optimal perturbation and its applications. Nonlinear Process Geophys, 10: 493–501

Mu M, Duan W S. 2003. A new approach to studying ENSO predictability: Conditional nonlinear optimal perturbation. Chin Sci Bull, 48: 1045–1047

Mu M, Xu H, Duan W S. 2007. A kind of initial errors related to "spring predictability barrier" for El Niño events in Zebiak-Cane model. Geophys Res Lett, 34: L03709, doi: 10.1029/2006GL027412

OsborneAR, PastorelloA. 1993. Simultaneous occurrence of low-dimensional chaos and colored random noise in nonlinear physical systems. Phys Lett A, 181: 159–171

Palmer T N, GelaroR, Barkmeijer J, et al. 1998. Singular vectors, metrics, and adaptive observations. J Atmos Sci, 55: 633–653

Qin X H, Duan W S, Mu M. 2013. Conditions under which CNOP sensitivity is valid for tropical cyclone adaptive observations. Quart J Roy Meteor Soc, 139: 1544–1554

Qin X H, Mu M. 2011. Influence of conditional nonlinear optimal perturbations sensitivity on typhoon track forecasts. Quart J Roy Meteor Soc, 138: 185–197

Sun G D, Mu M. 2011. Nonlinearly combined impacts of initial perturbation from human activities and parameter perturbation from climate change on the grassland ecosystem. Nonlinear Process Geophys, 18: 883–893

Teman R. 1991. Approximation of attractors, large eddy simulations and multiscale methods. Proc R Soc Lond A, 434: 23–29

Thompson C J, Battisti D S. 1995. A linear stochastic dynamical model of ENSO. Part I: Model development. J Clim, 8: 2897–2927

Thompson C J. 1998. Initial conditions for optimal growth in a coupled ocean-atmosphere model of ENSO. J Atmos Sci, 55: 537–557

Wang B, Tan X W. 2010. Conditional nonlinear optimal perturbations: Adjoint-free calculation method and preliminary test. Mon Weather Rev, 138: 1043–1049

Yu Y S, Duan W S, Xu H. 2009. Dynamics of nonlinear error growth and season-dependent predictability of El Niño events in the Zebiak-Cane model. Quart J Roy Meteor Soc, 135: 2146–2160

Yu Y S, Mu M, Duan W S, et al. 2012. Contribution of the location and spatial pattern of initial error to uncertainties in El Nino predictions. J Geophys Res, 117: 1–13

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