ABSTRACT

Title of Dissertation / Thesis: BRED VECTORS IN THE NASA NSIPP
GLOBAL COUPLED MODEL AND THEIR
APPLICATION TO COUPLED ENSEMBLE
PREDICTIONS AND DATA ASSIMILATION.

Shu-Chih Yang, Doctor of Philosophy, 2005

Dissertation / Thesis Directed By: Professor Eugenia Kalnay, Department of
Meteorology
Professor Ming Cai, Florida State University

The theme of my thesis research is to perform breeding experiments with NASA/NSIPP coupled general circulation model (CGCM) in order to obtain ENSO-related growing modes for ensemble perturbations. We show for the first time that the breeding method is an effective diagnostic tool for studying the coupled ENSO-related instabilities in a coupled ocean-atmosphere general circulation model that includes physical and dynamical processes of many different time scales. We also show for the first time that it is feasible to utilize the coupled bred vectors (BV) as a way to construct perturbations for ensemble forecasts for ENSO prediction using an operational coupled climate prediction model.

The results of the thesis research show that coupled breeding can detect a precursor signal associated with ENSO events. Bred vectors are characterized by air-sea coupled features and they are very sensitive to ENSO phases and background season. This indicates that bred vectors can effectively project on the seasonal-to-interannual instabilities by growing upon the slowly varying coupled instability.

These results are robust: bred vectors obtained from both the NASA and NCEP coupled systems exhibit similarities in many fields, even in atmospheric teleconnected regions.

We show that bred vectors have a structure similar to the one-month forecast error (analysis increment). The BV growth rate and the one-month forecast error show similar low frequency variations. Both of their subsurface temperatures have large-scale variability near the depth of thermocline. Evidence shows that bred vectors capture the eastern movement of the analysis increment (one-month forecast error) along the equatorial Pacific during 1997-1998 El Niño evolution. The results suggest that one-month forecast error in NSIPP CGCM is dominated by dynamical errors whose shape can be captured by bred vectors, especially when the BV growth rate is large.

These results suggest that bred vectors should be effective coupled perturbations for ensemble ENSO predictions, compensating for the lack of coupled ENSO-related perturbations in current operational ensembles. The similarity between the bred vectors and the one month forecast errors suggests that bred vectors can capture "errors of the month" and could also be applied to improve oceanic data assimilation by providing information on the month-to-month background variability.

BRED VECTORS IN THE NASA NSIPP GLOBAL COUPLED MODEL AND
THEIR APPLICATION TO COUPLED ENSEMBLE PREDICTIONS AND DATA
ASSIMILATION

By

Shu-Chih Yang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:
Professor Eugenia Kalnay, Chair
Professor Ming Cai, Co-Chair
Professor James Carton
Professor James Farquhar
Professor Sumant Nigam
Dr. Michele Rienecker
Dr. Joaquim Ballabrera

Acknowledgements

Table of Contents

This Table of Contents is automatically generated by MS Word, linked to the
Heading formats used within the Chapter text.

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background

The interest in the ability to predict climate variability has greatly increased during the last decade, particularly for seasonal to interannual time scales, which are mainly associated with ocean-atmosphere interactions. Of all the seasonal-interannual climate variabilities, the El Niño-Southern Oscillation (ENSO) phenomenon plays the most important role in dominating interannual sea surface temperature (SST) variations in the tropical Pacific. The ENSO phenomenon is characterized by a large SST anomaly in the eastern Pacific as a consequence of the propagation of equatorial waves induced by surface wind anomalies. Feedbacks through air-sea interaction determine the ENSO evolution, duration and intensity. Moreover, influences from ENSO can contribute to other climate anomalies and have a global-wide influence including the modification of the frequency of extreme weather events (Ropelewski and Halpert 1987).

It has been shown that a simple coupled model can explain the main characteristics of ENSO evolution based on the delayed oscillator theory (Schopf and Suarez 1987; Suarez and Schopf 1988; Battisti 1988). Intermediate models coupled tropical ocean with a steady, diagnostic atmosphere (Cane et al. 1986, 1987, Zebiak and Cane, 1987, and Battisti 1988) have demonstrated their ability to describe ENSO evolution and to provide useful forecast skill up to one year. Their results suggest that

the key to ENSO prediction lies on their ability to describe SST variations in seasonal to interannual time scales. However, the simplicity of intermediate models restricts us in understanding the impact of ENSO on other climate phenomena such as atmospheric teleconnections. In order to obtain a realistic projection of future global climate anomalies associated with the influence of ENSO, it is important to advance our ability to forecast with a coupled general circulation model (CGCM) with full physical processes in both ocean and atmosphere.

There are two primary schools of thought on the factors affecting the prediction skill of the SST seasonal-interannual forecast. Both schools assume that the coupled model used for ENSO prediction is reliable, and thus model error is not a dominant factor. One school emphasizes the presence of errors in the oceanic initial condition, as pointed out by Latif et al. (1998). The other school focuses on the stochastic atmospheric variability, since the atmosphere is unpredictable after a few days, and it can still influence to some extent to the SST evolution (Vialard et al, 2003, Kleeman and Moore 1997).

Recent studies have examined the relative importance of including these two factors in the initial condition and have demonstrated that sophisticated initializations through ensemble forecast and data assimilation can have a large impact on ENSO prediction skill. The quality of oceanic initial condition has been shown to be as a strong limiting factor on ENSO forecast skill, and its impact appears in the wind-SST-thermocline relationship (Chen et al. 1995, Ji and Leetmaa 1997, Rosati et al. 1997

2

and Alves et al. 2002). Vialard et al. (2003) performed a series of ensemble forecast experiments with a state-of-the art coupled GCM using wind perturbations, SST perturbations, and stochastic physics, individually or combining all of them. Their results suggested that the uncertainties in SST determine the spread of ensemble forecasts in early months of the forecast, while perturbations of the wind stress or atmospheric internal variability are less efficient in generating SST variations. Their results are supported by Chen et al (2004) who argue that the evolution of El Niño is controlled to a large degree by self-sustaining internal dynamics, and thus the ENSO prediction depends more on initial conditions than on unpredictable atmospheric noise. Since the oceanic initial condition obtained through ocean data assimilation is imperfect, increasing ENSO prediction skill requires the development of a good strategy to include the uncertainties in initial conditions derived from ensemble forecasts in ocean data assimilation for the best use of oceanic observations.

Following the experience of numerical weather prediction, a good ensemble prediction system should have initial perturbations that *represent the actual analysis uncertainties*. The spread of ensemble then will provide some measure of the level of uncertainty in the forecasts. The uncertainties in climate prediction have large-scale features associated with the dominant climate variability, as we will see in Chapter 4. Therefore, ensemble members can be *effective* if such uncertainties carried forward by dynamical models project strongly on large scale, climate relevant features. For example, we expect a good ensemble forecast of ENSO prediction to reflect the

uncertainties in the tropical Pacific and as well as its associated teleconnected atmospheric regions with a seasonal to interannual time scale.

Given the coupled nature of ENSO, initial ensemble perturbations need to be constructed in a coupled manner in order to ensure they reflect the coupled ENSO uncertainties in the seasonal to interannual scales. Such coupled uncertainties cannot be derived within current data assimilation system in coupled GCMs since the data assimilation is usually done separately for the ocean and the atmosphere. Therefore there is currently a need for a feasible approach to estimate the background error covariance in data assimilation including the large-scale coupled ENSO uncertainties.

1.2 Ensemble forecasting for ENSO prediction

Current ensemble forecasts are based on one of two approaches for the initialization process: the "*two tier*" and the "*one tier*" configuration (i.e., double or single stage configuration). In the older and widely used two-tier system, a single forecast of SST anomalies is used to force an ensemble of atmospheric forecasts (Bengtsson et al. 1993). This one-way communication between atmosphere and ocean neglects the *coupled* nature of the initial perturbations and thus inhibits the growth of coupled perturbations. Thus, this approach does not contain coupled ENSO-related perturbations. This hardly seems optimal to for seasonal and interannual prediction. The one tier or single stage configuration of CGCM introduced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Stockdale et al. 1998) generates all the ensemble forecasting members via a coupled ocean-atmospheric

model in order to have the perturbation growing under a coupled configuration. However, the initial ensemble perturbations are only introduced in the atmosphere so that the initial perturbations are still handicapped by not being initially coupled. In the NASA seasonal to interannual prediction coupled GCM, the routine initialization for ensemble forecast is also carried through a one-tier system, and includes perturbations in both the atmosphere and ocean, but their initial perturbations are generated independently (section 2.4). Although coupled instabilities will eventually develop in these one-tier systems, they are still handicapped by *not including coupled uncertainties in the initial perturbations for ensemble predictions*. Results from Vialard et al. (2003) showed that the ensemble spread is limited because the uncertainties in initial condition are not well sampled. Therefore, there is a need for ensemble ENSO prediction systems to have initial perturbations generated in a coupled manner and to ensure the ensemble perturbations can effectively project on coupled instabilities.

(a) Methods for generating ensemble perturbations

There are two main methods used in atmospheric operational forecasting for generating effective ensemble perturbations in ensemble forecasts: the breeding method (Toth and Kalnay 1993, 1997) and singular vectors (Errico and Vukicevic 1992; Buizza and Palmer 1995, Palmer et al. 1998). These two methods have been widely tested with an atmospheric GCM model for numerical weather prediction and both have been applied in intermediate coupled ocean-atmosphere models, like the Zebiak-Cane/Battisti models to find the ENSO-related fast growing mode. In this

section, we briefly review recent studies on the initial perturbations used for ENSO prediction.

- Singular vectors

Singular vectors are those directions that will maximize the error growth in a dynamic system, given a choice of perturbation norm and optimization time. In order to obtain singular vectors, one needs to linearize a nonlinear model to construct its tangent linear and adjoint models. Initial singular vectors are the eigenvectors derived from the matrix $\mathbf{M^T M}$, where $\mathbf{M}$ is the tangent linear model and $\mathbf{M}^T$ is its adjoint model. The leading *initial* singular vector is a set of initial field perturbations that will give the largest growth with respect to the chosen norm after linearly evolving for a chosen optimization time interval. The linearly-evolving perturbation at the end of the time interval is called *final* singular vector.

Several efforts of previous studies were made to derive singular vectors from coupled models with intermediate complexity including dynamical and hybrid models. However, although the *final* SVs tend to be similar to each other (and to coupled bred vectors), there is *considerable disagreement on the initial singular vectors obtained even with the same choice of norm*. With an optimization time of 3-6 months, Penland and Sardeshmukh (1995), Chen et al. (1997) and Thompson (1998) obtained similar results with a SST norm that maximizes the final growth of SST perturbations. Their initial singular vectors show large amplitudes of SST signal located in southeastern Pacific. Figure 1.2.1(a) and (b) from Chen et al. (1997) represent the initial and final

singular vectors of this kind (optimization time is 6 months). It should be pointed out that even though the growth rate of their singular vector strongly depends on seasonal cycle and ENSO phase, their initial singular vectors are *insensitive to both*. However, also using a SST norm, Xue et al. (1994, 1997ab) and Fan et al. (2001) obtained ENSO-like initial singular vectors, emphasizing more the signals in the eastern Pacific and much more concentrated towards the equator (figure 1.2.1(c) and (d) as the initial and final singular vector, from Xue et al. 1997a), even though their final singular vectors are very similar. Finally, a very different pattern compared to all the other studies has been obtained from a series of studies done by Moore and Kleeman (1997b, 1999a, b and 2001). Their results emphasize large signals in the west to central Pacific, as shown in the examples of figure 1.2.1(e) and (f) as their initial and final singular vector (from Moore and Kleeman 2001).

When the optimization norm is chosen to include the perturbations growth associated with thermocline depth, the initial perturbations of thermocline depth are somewhat similar to each other (Xue et al. 1997, Thompson 1997, Fan et al. 2000 and Moore and Kleeman 2000), but with still quite different initial SST pattern. The results from Fan et al. (2000) and Moore and Kleeman (1996) disagree on the relative importance between SST and thermocline in the initial condition. The former used a norm of analysis error covariance and concluded that both fields are important for the perturbation growth while the latter used an energy norm that stressed the thermocline depth only.

In most of these works, the coupled models have a dynamic tropical ocean model coupled with a simple atmospheric component, either a statistical or a diagnostic model, such as the Zebiak-Cane (ZC) type of model. Therefore, differences between the structures of initial singular vectors have been attributed to different physics applied in their atmosphere component. The model determinant features can also explain the similarity between the patterns of initial thermocline depth since thermocline variations are determined by linear tropical wave dynamics, which are more or less similar in all the ocean components.

Among all the other works, a more realistic atmospheric response has been considered in works by Moore and Kleeman (1997,1999a, b) and Moore et al. (2003) such as parameterizing deep convection associated the warm pool sensitivity or utilizing a statistical relationship trained with real observations of wind stress, in order to include the stochastic forcing from the atmosphere. With these ingredients, their singular vectors now emphasize the west to central Pacific rather than the east. Their recent work (Moore and Kleeman, 2001) found that such sensitivities of SST variations responding to atmospheric forcing could be retrieved by weighting the SST norm according to the standard deviation of SST anomalies rather than using SSTs directly. We know that the real atmosphere and ocean are strongly coupled in the warm pool region. This suggests that the growing perturbations in this region should be obtained naturally by coupled dynamical processes that include not only the SST but also atmosphere wind and upper ocean condition, and not by the judicious choice of a particular norm.

Two important points are suggested by the results of those studies related to singular vectors: (1) choosing a norm with limited variables reduces the number of degrees of freedom of perturbations and (2) there is a need to apply complete physics in a coupled model like a coupled general circulation model in order to obtain a more realistic response. However, such a complex system contains many types of instabilities, and because singular vectors are obtained by linearizing the complex models, singular vectors will be dominated by its faster growing modes, such as atmospheric synoptic instabilities. It is difficult to separate slow modes such as the ENSO mode from the fast modes such as weather noise by means of the choice of a norm or an optimization time. In addition, the enormous computational cost of constructing the adjoint model makes it impossible to calculate singular vectors for a coupled GCM. Recently, Kleeman et al. (2003) proposed an alternative to avoid constructing the adjoint model. Their climatically relevant singular vectors are obtained by reducing the linear propagator to a subspace that is mainly spanned by dominant correlation empirical orthogonal function (EOF) of climate variability associated with SST variations (a reduce-state space). A large ensemble size of model integrations is performed for the convergence of the pattern of the singular vector. Nevertheless, the choice of such reduced-state space considers only the SST climate variability, which is not optimal without counting in the coupled variations from wind and subsurface thermal structure.
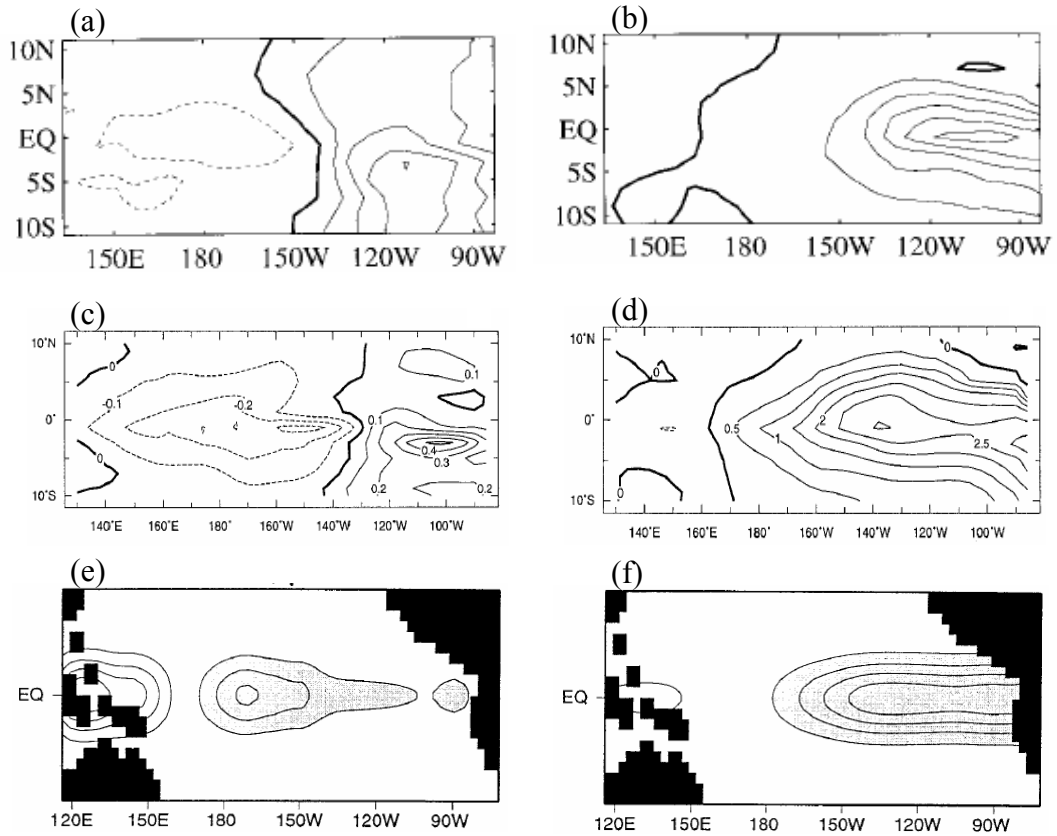
Figure 1.2.1 Singular vectors with SST norm from different coupled model.
(a) and (b) are initial and final singular vectors (after 6 months), figures are adopted from figure2(a) and(b) in Chen et al. (1997).
(c) and (d) are initial and final singular vectors (after 3 months), figures are adopted from figure 13(b) and (c) in Xue et al.(1997).
(e) and (f) are initial and final singular vectors (after 3 months), figures are adopted from figure 6(a) and (b) in Moore and Kleeman (2001).

- Bred vectors

Breeding is simply running the nonlinear model twice, with a periodic rescaling of the differences, known as bred vectors. It is a finite-time, finite-amplitude generalization of the method used to create Lyapunov vectors, which are the perturbations with a sustainable fast growth. Figure 1.2.2 is a diagram illustrating the steps for performing a breeding cycle in a nonlinear model:

(1) Introduce a random perturbation in initial (unperturbed) condition where that the initial perturbation has an amplitude equal to the size we choose.

(2) Integrate both the non-perturbed and perturbed initial conditions forward for a chosen period to obtain a control and perturbed forecasts.

(3) Rescale the difference between control and perturbed forecasts to the size of the initial perturbation and add this rescaled difference to the next corresponding initial condition.

(4) Repeat step (2) and (3) through the end of breeding experiments. The rescaled difference between two nonlinear integrations is defined as the bred vector.

In a coupled model, there are fast atmospheric instabilities associated with weather and convection, and the slower coupled instabilities we are interested in. In order to obtain the slow coupled instability, we use the slow component of the system for rescaling, and, thus in step (3), we measure the growth of the perturbations based on an oceanic variable, e.g., the SST perturbation. Then, the same rescaling factor is applied to both ocean and atmosphere perturbations. In a perfect model experiment,

there is only one control run and the initial condition is the model integration at the corresponding time. For a non-perfect model experiment, real observations are used to update the oceanic state, which will be described in Chapter 4.
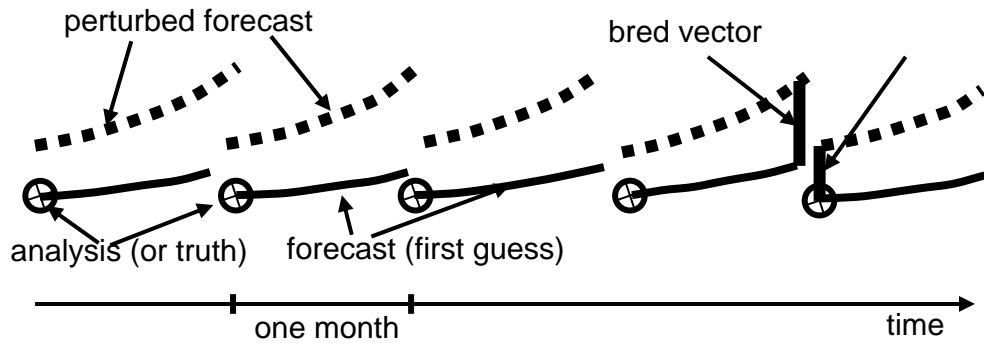
Figure 1.2.2 Schematic diagram showing a continuously evolving breeding cycle upon the unperturbed(control) model integration in a perfect model setting. The difference between the unperturbed and perturbed forecasts yields the bred vectors. The growth rate is computed as the ratio of the final to the initial size.

Toth and Kalnay (1993, 1997) suggested that the breeding method could be applied as a natural filter in a complex system that contains different types of instabilities. Their experience choosing an appropriate amplitude of the initial perturbations showed that baroclinic instability in the atmosphere can be separated from much faster convective instabilities. The idea is to take advantage of the fact that fast growing convective instabilities will saturate at an amplitude that is much smaller than the analysis error, and that this takes place at an earlier time. Toth and Kalnay (1996) suggested that the breeding method could also be used to isolate the ENSO coupled instability in a coupled ocean-atmosphere system since weather instabilities that appear in atmosphere can be treated as weather noise with an earlier

saturation of the error growth rate. Unlike singular vectors in a coupled system constructed in a reduced-space state, the breeding method can handle the full model nonlinear complexity in a natural way. Thus, breeding parameters allow the slow coupling instability to dominate the total growing instabilities in a nonlinear model, while the singular vector is looking for the linear perturbation that maximizes the final perturbation norm. Peña and Kalnay (2004) compared the ability of bred vectors and singular vectors to separate the slow growing mode from the fast growing mode in coupled Lorenz-type models with distinct time scales of instabilities representing a slow "coupled tropical ocean-atmosphere system" and an "extratropical atmosphere". They confirmed the conjecture in Toth and Kalnay (1996) that breeding is able to isolate the slow modes of the coupled system when rescaling intervals and amplitudes are chosen from physically appropriate scales and the rescaling factor is obtained from the slow component of the system. In their results, singular vectors are unable to isolate the slow because they are linear and the fast growing modes dominate the perturbation growth rate. Their results explain why the singular vectors can identify ENSO mode only when tangent linear models exclude the process of fast growing instability, such as the work done by Kleeman et al. (2003). Independent work by Boffetta et al. (1998) also demonstrated that choosing the perturbation size is a powerful tool to isolate the slow mode in a multiple time scales system.

The first practical attempt of applying the breeding method in an atmosphere-ocean coupled system was done by Cai et al. (2003) with the intermediate ZC model. Here, we illustrate their main results with two figures. They found that the growth

13

rate of bred vector is very sensitive to the ENSO phase as shown by figure 1.2.3. It shows that the growth rate is weakest at the peak time of the ENSO states (in both El Niño and La Niña) and strongest between the events. Unlike the singular vectors, the pattern of the coupled bred vector is insensitive to the choice of rescaling norm and is very sensitive to the background ENSO phase and seasonality. When the coupled bred mode was removed from the initial condition errors (figure 1.2.4(c)), it reduced substantially the growth of forecast error and essentially eliminated the "spring barrier" (compare with figure 1.2.4(b)). They also obtained a significant improvement for ensemble mean forecasts by using a pair of plus/minus bred vectors as ensemble perturbations. Their work suggests potential impact in both ensemble prediction and data assimilation for ENSO predictions.

Since the ZC model is simple and has only a diagnostic atmosphere, it could be expected that the dominant mode associated with ENSO variability would be apparent in their results. Processes with growth rates related to different time scales are not considered in this model and might be much harder to handle.

1.3 Potential applications of bred vectors in data assimilation

Kalnay and Toth (1994) pointed out that data assimilation cycles are in essence analogous to breeding cycles. The bred vector at time $i+1$ is the difference between two nonlinear model integrations initialized with a small difference at time i (Eq. 1.3.1). In this equation, M stands for the nonlinear model operator, x^i is the control run at initial time i, $\mathbf{BV}^i$ is the bred vector obtained at time i and r is the rescaling factor. The dynamical evolution of the bred vector ($M(x^i + \mathbf{BV}^i) - M(x^i)$) can be

approximated by evolving the bred vector ($\mathbf{BV}^i$) with a tangent linear model ($\mathbf{L}$) corresponding to the nonlinear model M:

$$\mathrm{BV}^{i+1} = r[M(x^i + \mathrm{BV}^i) - M(x^i)] \cong r\,\mathrm{LBV}^i. \tag{1.3.1}$$

On the other hand, data assimilation is a process looking for a best-analyzed state by statistically combining the forecast state and observation. Here, we illustrate this process assuming that observations are made of model variables (Eq. 1.3.2). This equation represents an updating process at time $i+1$, where x_a is the analysis state, x_f is the forecast state, y is the observation and $\mathbf{K}$ is a weighting matrix to rescale the difference between forecast and observation:

$$x_a^{\,i+1} = x_f^{\,i+1} + \mathbf{K}(y^{i+1} - x_f^{\,i+1}) \tag{1.3.2}$$

In Eq. 1.3.2, the difference ($x_a^{\,i+1} - x_f^{\,i+1}$) is referred to as the analysis increment. This quantity is also used as a proxy for representing the one-month forecast error.

By defining the evolution of truth state with the nonlinear model, $M(x_t)$, we can rewrite Eq (1.3.2) as an increment form as described in Eq. (1.3.3), where ε_f is the forecast error, (difference between truth and forecast state) and ε_o is the observation error, (difference between truth and observation).

$$x_a^{\,i+1} - x_f^{\,i+1} = \mathrm{K}(y^{i+1} - x_f^{\,i+1}) = \mathrm{K}[(x_t^{\,i+1} + \varepsilon_o^{\,i+1}) - (x_t^{\,i+1} + \varepsilon_f^{\,i+1})] \approx \mathrm{K}[\varepsilon_o^{\,i+1} - \mathrm{L}\varepsilon_a^{\,i}] \approx -\mathrm{KL}\varepsilon_a^{\,i}$$

$$\tag{1.3.3}$$

The forecast error, ε_f^{i+1}, at time $i+1$ can be approximated by evolving the analysis error linearly with the tangent linear model ($\varepsilon_f^{i+1} = \mathbf{L}\varepsilon_a^i$). It is reasonable to assume that observational errors are random and small. Therefore, the analysis increment can be approximated as $-KL\varepsilon_a^i$, i.e., the analysis increment essentially corrects most (but not all) the forecast error that evolved from the previous analysis error. Eqs. (1.3.1) and (1.3.3) suggest that both bred vectors and analysis increments should be dominated by the dynamical processes that produce error growth. From this, we will argue that the *analysis increment should project strongly on the bred vectors* (since bred vectors are like the fastest growing normal modes of the evolving flow, they are defined except for the sign).

This is because the forecast error is dominated by dynamic errors, and the analysis error partially "inherits" these dynamical errors because the observations are not enough to completely correct them. Therefore, in the analysis cycle there is forecast error growth during the forecast, and a partial reduction of the errors during the analysis. The process is analogous to the bred perturbation growth and rescaling that creates the bred vectors. Thus we expect that the forecast errors (which we do not know) and the analysis increments (which try to correct them using the observations) should have a strong projection on bred vectors.

Since bred vectors are able to capture the evolving dynamic errors, bred vectors could be used in a data assimilation scheme to minimize the projection on these dynamic errors in the observational increments (difference between the observations

and the forecast), a minimization along a single degree of freedom that simply requires adding the bred vector multiplied by an appropriate amplitude to the forecast (Kalnay and Toth 1994, Cai et al. 2003). A similar idea could be applied to NSIPP oceanic assimilation scheme by reducing the errors that project on the bred vector (estimated by the global difference between the observations and the forecast), making better use of the available ocean data.

Corazza et al. (2002) demonstrated that the forecast errors have indeed a very substantial projection into the subspace spanned by ten bred vectors using a quasi-geostrophic model of 10^4 degrees of freedom. This result was used to represent the shape of the "errors of the day" in order to compensate the deficiency that a time-independent background error covariance cannot detect "day-to-day" background variability. By augmenting the background error covariance with a weighted outer product of bred vectors, they found that the analysis error variance could be reduced substantially at a negligible computational cost. Their result suggests that it would be possible to improve the oceanic assimilation scheme in current NASA/NSIPP operational system by providing the information of the "coupled errors of the month" with the coupled bred vectors for the background error covariance. Therefore, we will examine the relationship between the bred vector and analysis increment in order to check whether this idea is feasible (Chapter 4).

The goal of this thesis is to explore the feasibility to create, and the potential practical applications of coupled bred vectors in a system as complex as a coupled GCM. The outline of the thesis is as follows:

Chapter 2 is a brief description of the general characteristics of the NASA Seasonal to Interannual (NSIPP) coupled GCM, including its performance on ENSO simulation and prediction. In Chapter 3 we first implement the breeding method in the NSIPP coupled GCM in order to determine if it is possible to isolate the ENSO-related growing mode from weather noise. This is a "perfect model" simulation. We examine the characteristics of coupled bred vectors in the NSIPP system that we obtain and compare them with those obtained in a similar experiment using instead the NCEP coupled GCM, also assuming a perfect model. The results presented in Chapter 3 show that the coupled bred vectors are well defined, with robust characteristics that are very similar in the NASA and in the NCEP system. Finally, in Chapter 4, we drop the perfect model assumption and implement breeding on the operational NSIPP system, which assimilates real observations. This is a much more complex system but we still find that the coupled bred vectors are strongly related to El Niño, and, very encouragingly, are similar to the analysis increments (forecast errors). This suggests that it is worthwhile to pursue their application to ensemble forecasting and data assimilation. A summary and plans for future collaboration with NSIPP operational system will be given in Chapter 5. Appendices present summaries of other studies independent from the present one that have been or will be submitted for publication.
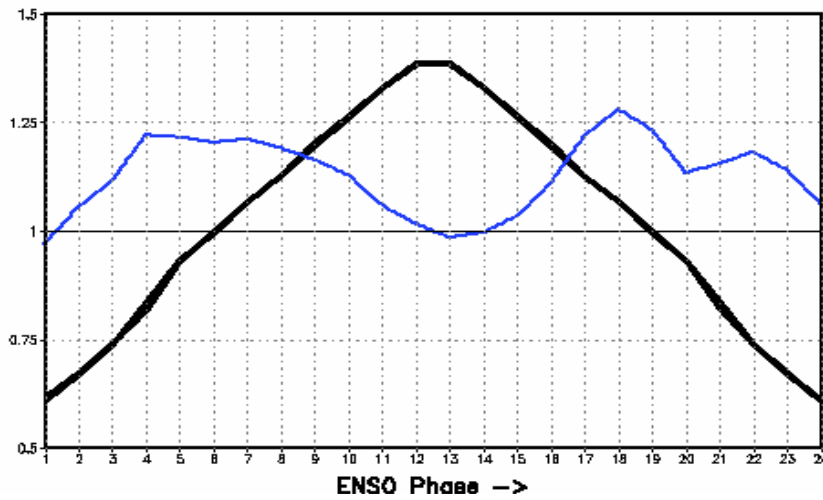
Figure 1.2.3 The growth rate of bred vector (blue) obtained with Zebiak and Cane model evolves against phases of background ENSO events (black). Figure is adopted from Cai et al. (2003).
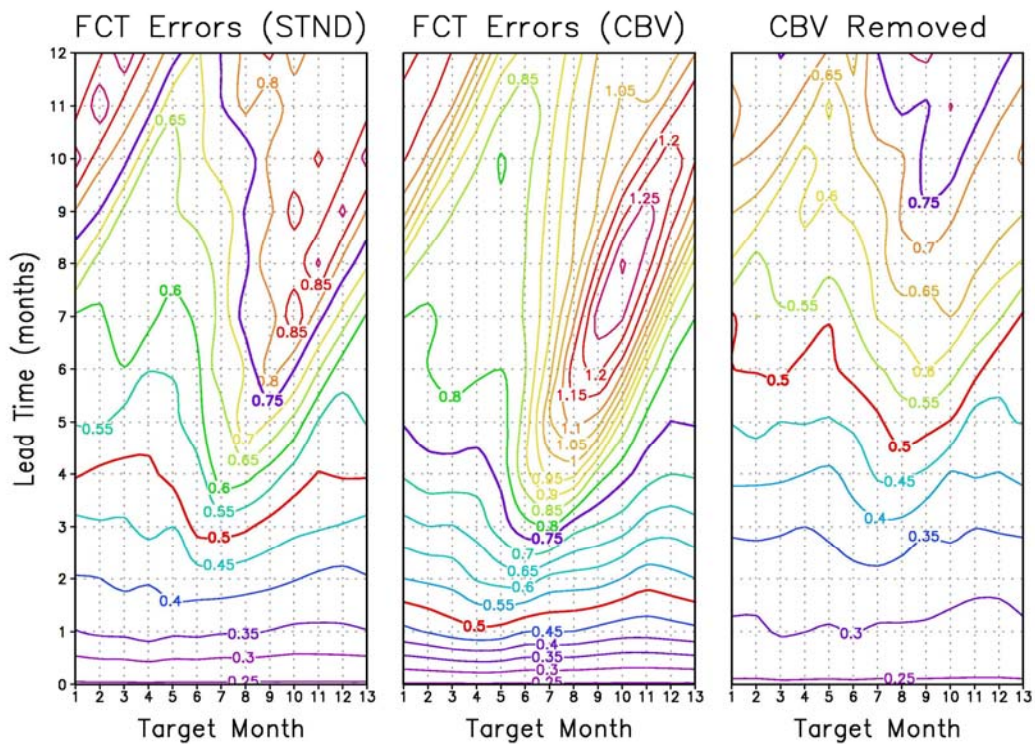
Figure 1.2.4 Forecast rmse as a function of forecast lead time and the target month. Contour interval is 0.05 in a dimensionless unit. Initial errors are simulated with (a) random fields, (b) composite bred vector, and (c) random fields used in (a) minus composite bred vector used in (b). Figure is adopted from Cai et al. (2003).

19

Chapter 2: NSIPP Coupled model description

2.1 Components in NSIPP coupled model

The NSIPP coupled model is a fully coupled global ocean-atmosphere-land system developed at NASA Goddard Space Flight Center (GSFC) (Miller et al. 2004; Vintzileos et al. 2003). It is comprised of the NSIPP-atmospheric general circulation model (AGCM, Suarez 1996, Bacmeister and Suarez 2002; Bacmeister et al. 2000), the Poseidon ocean model (OGCM, Schopf and Loughe 1995), and the Mosaic land surface model (LSM, Koster and Suarez 1992) all developed in NASA Goddard Space Flight Center.

(a) The Poseidon ocean model

The Poseidon model is a reduced-gravity quasi-isopycnal ocean model (Schopf and Loughe, 1995). The prognostic variables in this model are layer thickness, temperature, salinity, and zonal and meridional components of current. It uses finite-differences on latitude-longitude coordinate in the horizontal and in the vertical. A generalized vertical coordinate is used to represent layers typed as a turbulent well-mixed surface layer, and nearly isopycnal (constant density) deeper layers. It is designed to have zero value at the surface and increases by one between successive layer interfaces. The Poseidon model has been documented and validated in hindcast studies of El Niño (Schopf and Loughe 1995) and has been updated to include

prognostic salinity (Yang et al., 1999). The resolution used in the current operational

version is 1/3 degree in latitude, 5/8 degree in longitude, and 27 layers in vertical.

The model equations are formulated to ensure conservation of mass, momentum,

heat, salt and other tracers under any grid choice. The entrainment in the turbulence

well- mixed surface layer is parameterized according to a Kraus-Turner bulk mixed

layer model (Sterl and Kattenberg, 1994). The isopycnal region is treated in a quasi-

isopycnal fashion, in which layers do not vanish at outcrops and remain with a thin

minimum thickness at all grid points. A high order Shapiro filter is implemented to

deal with horizontal mixing. Vertical mixing and diffusion are parameterized using a

Richardson number dependent scheme of Pacanowski and Philander (1981). The

diffusion coefficients calculated according to that scheme are enhanced when needed

to simulate convective overturning in cases of gravitationally unstable density

profiles. A thermodynamic sea-ice model, following Hakkinen and Mellor (1992) is

included, with heat and freshwater exchange with the first layer of the ocean model.

(b) NSIPP atmospheric general circulation model

The NSIPP AGCM was developed at Goddard for the purpose of climate

simulation and prediction so that physical parameterizations are designed to ensure

the AGCM can represent proper climate variabilities. Therefore, model developments

have focused on better simulation of tropical windstress and keeping middle latitude

atmospheric stationary waves sufficiently unbiased in order to be able to have proper

extratropical ENSO response.

This model uses a finite-difference C grid on latitude-longitude coordinate in the horizontal and a standard sigma coordinate in the vertical. Model variables are surface pressure, zonal and meridional wind components, potential temperature and relative humidity. The resolution in the current operational version has 2 degree in latitude, 2.5 degree in longitude and 34 layers.

In this model, finite differences are second order accuracy except for advection by the rotational part of the flow. The momentum equation uses a fourth order version of enstrophy conserving scheme of Sadourny (1975). Vertical differencing follows Arakawa and Suarez (1983). The horizontal advection schemes for potential temperature and moisture are fourth-order and conserve the quantity and its square (Takacs and Suarez, 1996). Solar and radiative heating rates are parameterized as in Chou and Suarez (1999) and Chou and Suarez (1994). From the moist physics parameterizations, the cloud fraction is estimated at each level. For solar radiation calculation, the GCM levels are grouped into three regions identified with high, middle and low clouds. Turbulence is handled with the Louis et al. (1982) scheme with a smaller mixing length scale of 20 m (a typical value is 80-160 m), in order to correct the excessive annual mean stress and unrealistic seasonal variation over equatorial Pacific.

Penetrative convection originating in the boundary is parameterized using the Relaxed Arakawa-Schubert (RAS) scheme (Moorthi and Suarez, 1992) to parameterize both deep and shallow convections. This scheme has a particularly good performance in simulating atmospheric response to tropical anomalies, which is a crucial aspect in coupled prediction problems. It has been upgraded with a more complete liquid water budget in convective adjustment. The large-scale cloudiness is determined using a relative humidity-based diagnostic scheme. It has been modified to include subsidence drying in the original RAS to avoid excessive cloudiness diagnosed initially over tropics and subtropical oceans.

Comparing to an earlier version, the model levels have been increased to 34 levels, mainly to increase the resolution of the boundary layer. Important improvements for the latest version include the use the gravity-wave drag parameterization (Zhou et al. 1996) and a filtered topography. The former one produces a remarkable improvement in the simulation of the zonal flow and the latter one improves the simulated stationary planetary waves.

The NSIPP AGCM is coupled to the Mosaic Land Surface Model (LSM) of Koster and Suarez (1996), a well-established soil-vegetation-atmosphere transfer model. A general description of the model is given in Koster and Suarez (1992). The Mosaic LSM computes area-average energy and waver fluxed from the land surface in response to meteorology forcing. In the Mosaic LSM, a square area with different vegetation types is divided into several homogeneous subregions characterized by a

23

single vegetation or bare soil type. Observed vegetation distributions are used to determine such partitioning. The energy balance is calculated separately for each sub-region. Detailed descriptions of the head and energy balance in LSM are given in Koster and Suarez (1996).

(c) Features of model coupling

The NSIPP CGCM employs the Goddard Earth Modeling System (GEMS) to couple the atmosphere, ocean and land models. The ocean and atmosphere exchange information once a day. The land/ocean mask for the coupled model is defined on the ocean's latitude-longitude grid, so each grid box is either all ocean or all land. The atmosphere-to-ocean couplers interpolate from the atmospheric grid to the mass point of the underlying ocean boxes using bilinear interpolation (e.g., Vintzileos and Sadourny, 1997). In the ocean-to-atmosphere coupling, interpolation consists of averaging together the underlying ocean grid boxes. Fluxes are exchanged on a daily basis. The coupling between the land and the atmosphere is handled in a similar fashion. The CGCM runs without any flux correction.

The ocean domain extends from Antarctica to 72°N. There is a 10°-wide buffer zone at the northern boundary in which the temperature, salinity and layer thickness are relaxed to climatological fields derived from the World Ocean Atlas 1997 (Levitus, 1994). From the northern boundary to the North Pole, a slab ocean "mixed

layer" model, with sea-ice enabled, is used merely for heat exchange with the atmosphere. This approach is also used for shallow seas and the continental shelves.

2.2 Data assimilation

(a) Current operational oceanic data assimilation (optimal interpolation)

Currently, the oceanic initial conditions are obtained from univariate optimal interpolation (UOI) data assimilation scheme. In the UOI scheme, only the model temperature field is updated. Daily subsurface temperature data from TAO observing system, XBT instruments and the ARGO array are assimilated. Daily wind stress data (SSM/I) is used to force the Poseidon model. In order to preserve the property of water-mass distribution, the model salinity field is corrected following Troccoli and Haines (1999) to have the salinity increments consistent with the temperature analysis. Detail about salinity correction in temperature assimilation scheme can be found in Troccoli et al. (2003).

The data assimilation procedure is solved volume by volume (processing element) in the Poseidon model. The analysis state is obtained from

$$x_a = x_f + W\left[y_o - H(x_f)\right] \tag{2.3.1}$$

$$W = P_f H^T (H P_f H^T + R)^{-1} \tag{2.3.2}$$

In Eq. (2.2.1), x_a and x_f are state vectors $(n_x \times 1)$ containing analysis and forecast temperature, y_o is the vector containing n_d observations. H is the operator which converts model variable to observational space and $\mathbf{W}$ is the optimal weight matrix to minimize the analysis error covariance. In Eq (2.2.2), $\mathbf{P}_f$ is the forecast (background) error covariance, $\mathbf{R}$ is the observation error covariance and $\mathbf{H}$ is the linearization of observational operator H. In the UOI, H is a 3D linear interpolation operator which maps the model temperature field to the location of observation and thus $H(x_f)=\mathbf{H}x_f$.

In the practical implementation, $[\mathbf{HP}_f\mathbf{H}^T + \mathbf{R}]^{-1}[y_o - \mathbf{H}x_f]$ is first solved by finding the solution, $\mathbf{b}$, for a linear system in Eq. (2.3.3)

$$\left[\mathbf{HP}_f\mathbf{H}^\mathbf{T} + \mathbf{R}\right]\mathbf{b} = y_o - \mathbf{H}x_f \qquad (2.3.3)$$

Then $\mathbf{b}$ is used to rewrite Eq. (2.3.1) to Eq.(2.3.4)

$$x_a = x_f + \mathbf{P}_f\mathbf{H}^\mathbf{T}\mathbf{b} \qquad (2.3.4)$$

The background error covariance used in the UOI is constant in time and only depends on the distance between forecast locations and a Gaussian functional form is chosen:

$$\mathbf{P}_f(\Delta\lambda, \Delta\phi, \Delta z) = C\exp\{-(\Delta\lambda / L_\lambda)^2 - (\Delta\phi / L_\phi)^2 - (\Delta z / L_z)^2\} \qquad (2.3.5)$$

where L_λ defines the zonal decorrelation scale, L_ϕ the meridional decorrelation scale and L_z the vertical decorrelation scale. In this application, $L_\lambda = 1800$ km, $L_\phi = 400$ km in the equatorial waveguide, and $L_z = 50$ m.

The salinity correction is applied at each model grid point in two steps. At time $i+1$, we have the model temperature background profile, $T_{bg}(z_{bg})^{i+1}$ and the analyzed temperature profile, $T_a(z_a)^{i+1}$, at each grid point. First, a vertical displacement of the model temperature background profile to match the deepest analyzed temperature is made $T(z_{bg}) = T(z_a)$. The same displacement is applied to the salinity profile. Second, salinity increment is computed following the scheme in Troccoli et al. (1999) using the temperature-salinity (T-S) relationship as described in Eq. (2.3.6).

$$S(z_a)^{i+1} = S_{bg}(z_{bg})^{i+1} \qquad \text{if } |z_a - z_{bg}| \leq \Delta z$$
$$S(z_a)^{i+1} = S_{bg}(z_a)^{i+1} \qquad \text{if there is no } z_{bg} \text{ such that } T_{bg}(z_{bg}) = T_a(z_a), \text{ or } |z_a - z_{bg}| > \Delta z$$

(2.3.6)

If the difference between depths of background and analyzed temperature is larger than a specified depth tolerance (Δz), the salinity correction is made. This tolerance value is chosen as 100 m. Also, as the temperature-salinity assumption generally does not hold near the surface, the salinity will not be updated in the surface isothermal layer.

The final step in the UOI process is updating the model state by inserting the analysis increment ($x_a - x_f$) in a gradual manner (IAU, e.g., Bloom *et al.* 1996). IAU

is performed following Eq. (2.3.7) where $\frac{\partial \mathbf{x}}{\partial t} = \mathbf{F}(\mathbf{x}, t)$ represents the original model equation, $x_a(t_i)$ and $x_f(t_i)$ are the analysis and forecast at the time i. The analysis increment is added as a state-independent forcing term in model equations.

$$\frac{\partial \mathbf{x}}{\partial t} = \mathbf{F}(\mathbf{x}, t) + \frac{(\mathbf{x}_a(t_i) - \mathbf{x}_f(t_i))}{(t_{i+1} - t_i)}, t_i \leq t < t_{i+1} \tag{2.3.7}$$

The IAU is done within a 5-day window and it has been documented for being able to improve observed-minus-forecast statistics with respect to a non-incremental updating scheme (Bloom *et al.* 1996). In addition, it provides two advantages for model initialization. First, it reduces initialization shocks resulting from imbalances between the model fields due to the direct insertion of the analysis increments in intermittent data assimilation. Second, the IAU allows the model to gradually adjust the model thickness field in response to the temperature and salinity increments without violating the constraints imposed by the continuity equation.

2.3 The performance of NSIPP CGCM on ENSO simulation

In Chapter 3, we will first examine the characteristics of bred vectors in a perfect model scenario, using a lower resolution CGCM (3°x3.7°x34 layers in AGCM and 0.5°x1.25°x20 layers in OCGM). In preparation for that, we will examine in this section the model ability to perform ENSO simulations in order to understand the CGCM behavior, which will influence the characteristics of the bred vectors.

A 62-year perfect model experiment has been made with this coarser version of the NSIPP CGCM. The simulated SST and wind stress from model simulations have shown reasonable distribution in both seasonal and interannual variability (Rienecker et al. 2000).

The standard deviations of the observed SST (Reynolds et al. 2002) and the model simulated SST anomalies show that the largest SST variations are located in the Pacific east of the date line (figure 2.3.1(a) and (b)). Generally, the NSIPP CGCM can capture most of the observed variability but with much smaller amplitude off the eastern coast. Also, the interannual variability is reasonably represented by the CGCM. Figure 2.3.2(a)-(c) are the Hovmöller diagrams of zonal wind stress, SST and thermocline anomalies along the equator. It shows that the model exhibits a realistic ENSO-like variability characterized by a delayed oscillator behavior with a quasi four-year period. The covariability between these three variables also demonstrates that NSIPP CGCM is able to describe the air-sea interaction processes. In this figure, there is a strong biennial component in the thermocline anomalies. The biennial component in the Poseidon model is stronger than the observed one and this will modulate the ENSO periodicity.

In Chapter 3, we will perform breeding experiments based on this simulated run from year 2020 to year 2029.

(a) STD of observed SST anomaly
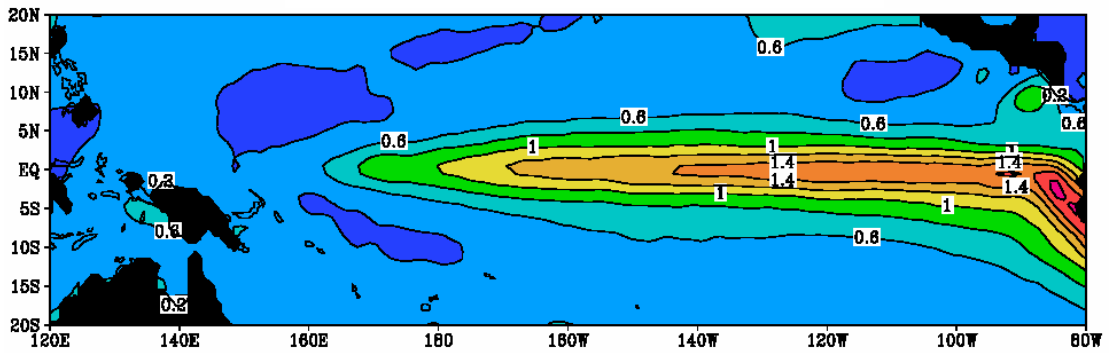
(b) STD of NSIPP SST anomaly
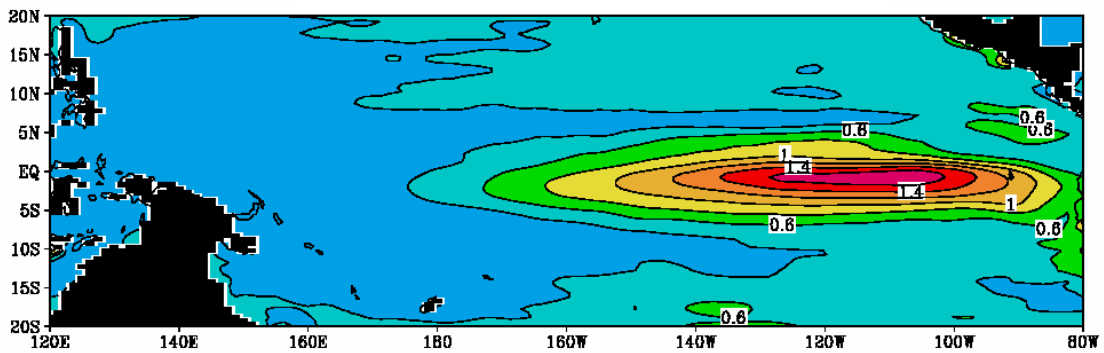
Figure 2.3.1 (a) standard deviation of observed SST[1] anomaly (°C) and (b) standard deviation of NSIPP simulated SST[2] anomaly(°C).

[1]: observed SST anomaly is calculated based on the Reynold OIv.2 SST from 1982 to 2003
[2]: simulated SST anomaly is calculated on the last 50 years of the long perfect model run
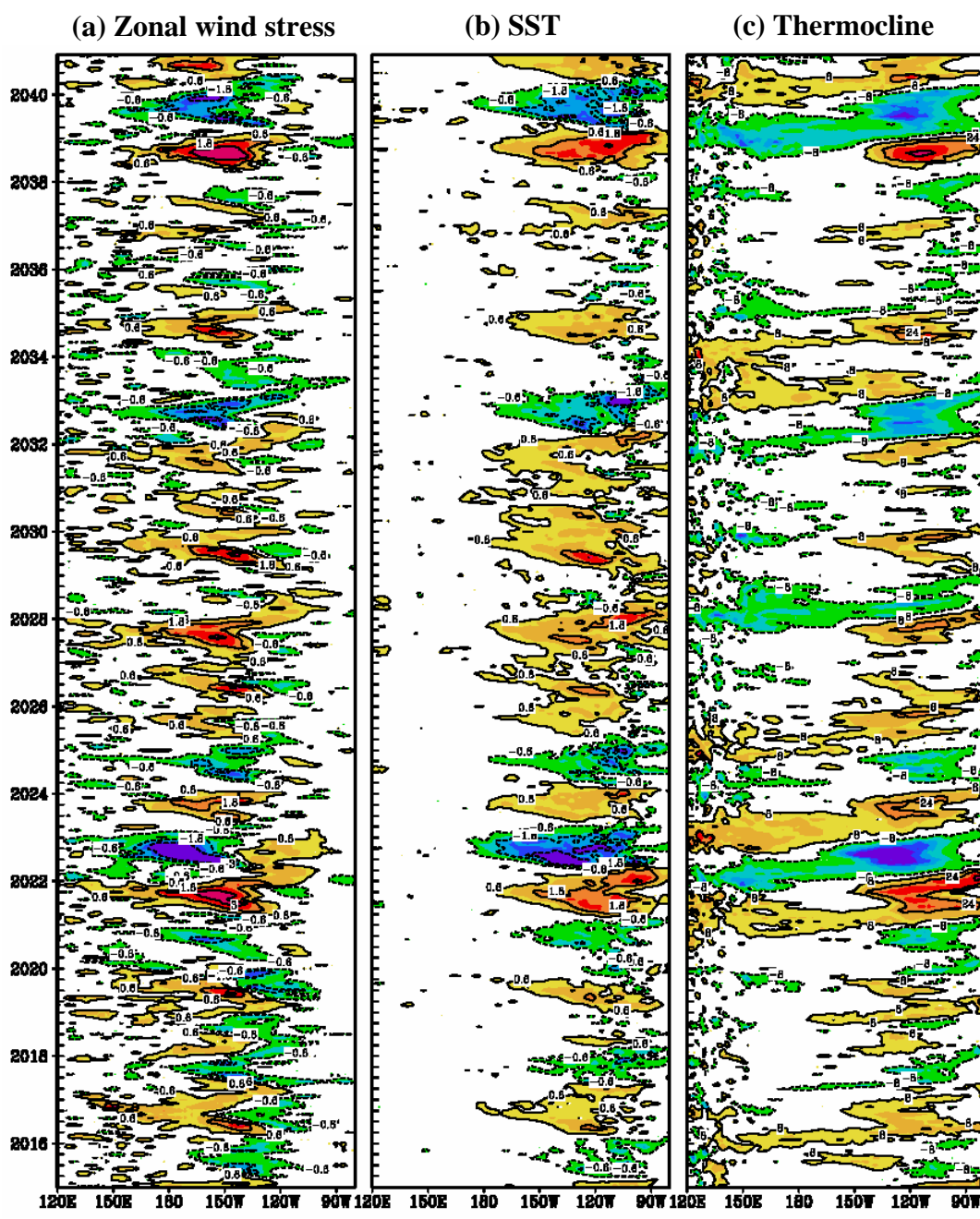
Figure 2.3.2 Hovmöller diagram of simulated anomalies at the equator (a) zonal wind stress anomaly $(100 \times N/m^2)$ (b) SST anomaly (°C) and (c) thermocline anomaly (m). The model climatology is calculated based on the last 50-year data of a perfect model run.

2.4 Ensemble forecast in NSIPP operational system

The NSIPP experiments forecast group generates 12-month ensemble forecasts using a fully coupled atmosphere-ocean-land system (Tier 1 forecasts). Their ocean is initialized with ocean states from NSIPP's assimilation of the global *in situ* temperature profiles using optimal interpolation as described in section 2.2 (Troccoli et al. 2002). The atmospheric and land states are independently initialized from NSIPP AMIP-style (Atmospheric Model Intercomparison Project) runs (Gates, 1992). NSIPP's AMIP runs are a coupled atmosphere-land simulation, where the monthly mean Reynolds sea-surface temperatures are specified as a boundary condition for the atmospheric model. In the coupling process, the AGCM starts first; therefore, the ocean sees the average daily value of the atmospheric condition (wind stress) for that day rather than directly using the AMIP-style initial condition.

Hindcast experiments were made from January 1993 with 6 member ensembles – three ocean-only perturbations, two atmosphere-only perturbations, and one with no perturbation. In our study, we will refer to the run with no perturbation as the control run. Ensemble perturbations used in the current operational system for ocean and atmospheric GCM are generated separately.

The perturbations for the ocean are obtained from differences between two analysis states. The initial analysis state is chosen randomly within 15 days of the forecasts initialization time, and is subtracted from the analysis state 3 days after the

first one. This 3-day analysis difference perturbation is rescaled randomly to a size that is no greater than 10% of the layer thickness maximum of this analysis difference. The magnitudes of the perturbations are limited in order to keep them in certain ranges with physical meaning. The perturbations are calculated and applied within the global ocean domain. The atmospheric perturbations are generated from choosing a random pair of ensemble members of an existing up-to-date AMIP run (30 model members). 10% of the difference between this randomly chosen pair is then added to a base member.

(a) ENSO prediction in the operational ensemble forecast system

A common way to diagnose the presence of an El Niño event in an ENSO prediction is to examine the forecast SST anomaly in the Niño3 region (150°E-90°W, 5°S-5°N). Current ENSO prediction with NSIPP CGCM can provide useful skill for up to nine months (http://nsipp.gsfc.nasa.gov/). However, the skill of the forecast Niño3 index[*] in the tropical eastern Pacific varies, depending on the starting month. Figure 2.4.1(a) and 2.4.1(b) are 6 months hindcast results for Niño3 index starting from April 1st and September 1st from 1993 to 2004, compared to the observed Niño3 index (Blue dash line in figure 2.4.1). As shown in figure 2.4.1(a), hindcast skills initializing from April drop quickly as the forecast month increases. Such drift is especially severe when observations show that the SST is anomalously cold (La Niña events). By contrast, forecasts initialized from September agree much better with the observed SST anomalies. In those hindcast experiments, SST anomalies have been

---

[*] Niño 3 index is the area mean of SST anomaly in Niño3 region (150°E-90°E, 5°S-5°N)

obtained by subtracting from the ensemble mean climatological forecast drift corresponding to each initialization month. The climatology/drift is calculated from forecasts over the 1993-2001 period.

The strong dependence on forecast initialization month indicates that the hindcast experiments have difficulties to predict the feedbacks between air-sea interaction that happen in the spring and summer. As suggested by previous work (Chen et al. 1995 and Cai et al 2003), the "spring predictability barrier" can be ameliorated by removing the growing error from the initial condition.

Occasionally, the hindcast results show a complete different sign of SST anomaly compared to the observations. Two factors are considered to explain such false ENSO alarm. First is the strong quasi-biennial component that has been noticed in the coupled model free run (figure 2.3.2(c)). It tends to have a strong warm event after a strong cold event and thus, creates erroneous warming. Second, it is also possible that such result comes from the interactions between the drifting mean state of the coupled model and the forecast interannual anomalies. This deficiency is strongly related to the behavior of the model mean state and as a result, forecast errors will contain large model errors.

In the next Chapter, we present the results of breeding within a "perfect model" scenario by using the long coupled integration as "nature" as described in section 2.3 and the breeding experiments were performed between the simulated year 2020-2029.
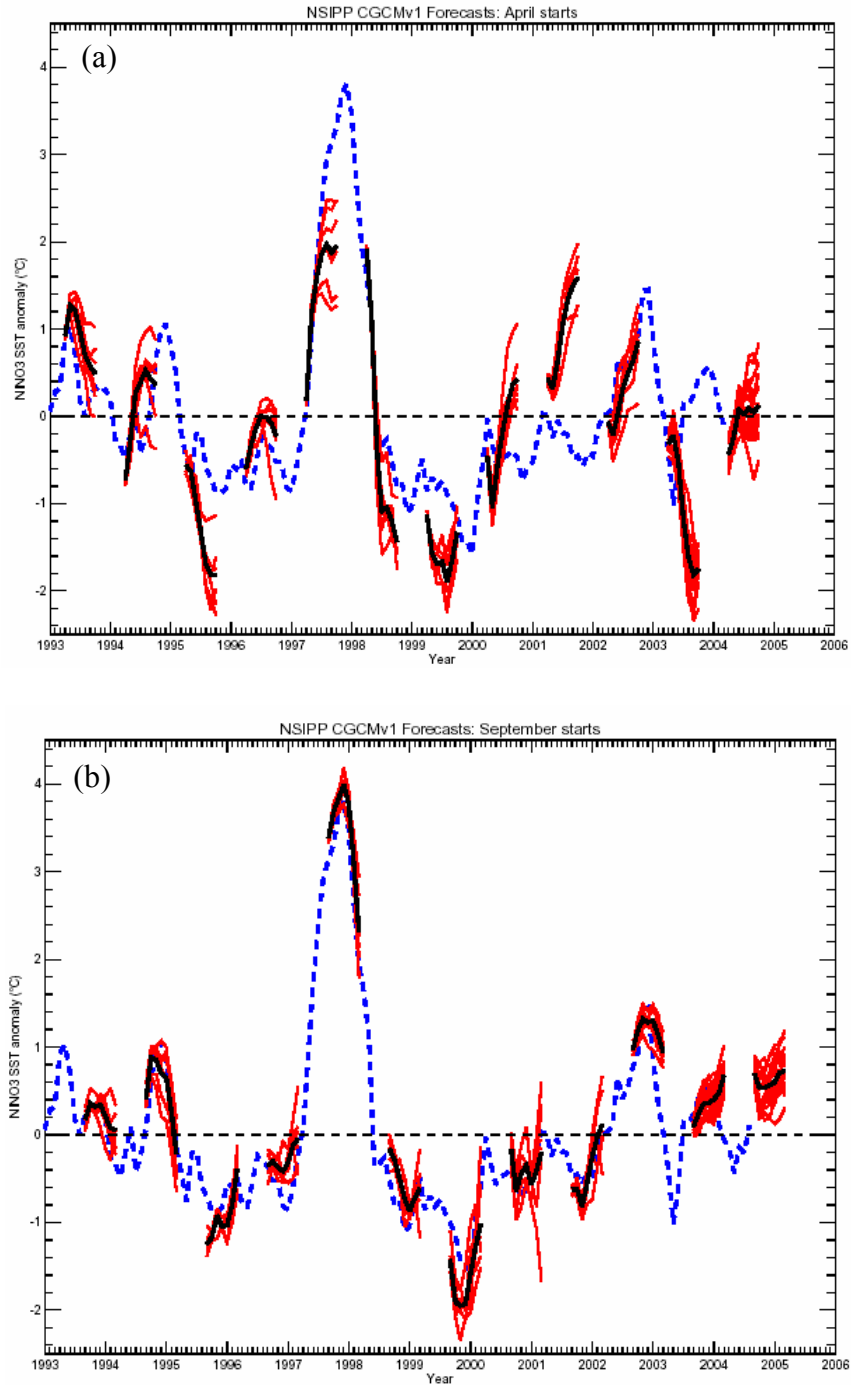
Figure 2.4.1 The forecast Niño3 index from the NSIPP ensemble hindcast experiments starting in (a) April and (b) September. The blue dashed line denotes the observed value, the red lines denote the ensemble member and the black solid line denotes the ensemble mean.

Chapter 3: Bred vectors in a perfect model experiment

In this Chapter we will discuss breeding experiments using the perfect model simulation introduced in Chapter 2. Starting from different random perturbations, two independent 10-year breeding experiments are performed from simulated year 2020 to year 2029. The unperturbed perfect model simulation will be referred to as the background state and bred vectors are defined as the difference between two nonlinear integrations (perturbed and unperturbed). Two parameters are used to control the breeding cycle. The rescaling period is chosen as one month, and the size of the perturbations chosen to rescale the magnitude of BV SST in the tropical Pacific domain (130°W-270°E, 15°N-15°S) to 0.085°C (about 10% of the natural variability). Two independently performed breeding runs have shown a very similar behavior. We then combined these two runs into a 20-year result in order to reduce the sampling error.

As a first step, characteristics of bred vectors are examined in order to understand their relationship to the background ENSO variability. Our results suggest that

- The breeding technique is able to detect the slowly-varying coupled ENSO instability if the rescaling period is chosen properly to be and physical meaningful.

- The bred vectors carry air-sea coupled features, which are characterized by their resemblance to the background ENSO-related variability.

- The characteristics of ENSO-related bred vectors are found to be robust by breeding in two independently developed coupled GCMs (NASA/NSIPP and NCEP/GFS). Comparisons show many similarities not only in the tropical Pacific but also in the PNA atmospheric teleconnected region.

- When the background state is under a La Niña event, the bred vectors also include the fast growing tropical instability waves near the northern part of the cold tongue region.

3.1 Growth rate of coupled bred vectors

Bred vectors represent, by construction, the instabilities that grow on the background flow. By choosing the rescaling period as one month, we focus on isolating the slow growing instability by saturating the unwanted fast growing instabilities, such as weather. The dominant slow growing instability is the ENSO-related coupled instability with an interannual time scale as we have discussed in Chapter 1.

An example of the instabilities captured by the coupled bred vectors is shown in figure 3.1.1, a snapshot of the bred vector SST field (contours) together with the corresponding background SST field (shading) on July 1 of the model year 2024. It shows that the bred vector field has a large amplitude along the sharp temperature

gradient in the equatorial cold tongue, coinciding with the background waves along the northern edge of the cold tongue. The bred perturbation suggests that the instability tends to make the waves break. Clearly, the formation of tropical instability waves is captured by the bred vectors. A main difference of the breeding method from the singular vector is that the breeding method allows the model to decide naturally which instability process will contribute to the growth of the instability of interest (i.e. slow-varying growing instability). By using the one-month rescaling period, bred vectors are able to pick up the growing signal associated with low frequency variations. The formation of the tropical instability waves strongly depends on the seasonal and interannual variability with a life time of 30-90 days. As a consequence, bred vectors naturally capture such instability and are able to identify the location of the occurrence of the tropical instability waves. The growing signal we measured in the Niño3 region contains coupled and uncoupled instabilities and we can view the net growth signal in this location as a result of the competition among these instabilities. In order to understand the components in the growing signals, we examine the relationship between the BV growth rate and the background ENSO variability, described by the Niño3 index.

The growth rate of the coupled bred vectors is calculated based on the chosen rescaling norm of the perturbation field within the tropical Pacific region:

$$G(t) = \frac{\sqrt{\sum_{NG} \mathbf{BV}_{\mathrm{SST}}^2(t)}}{\sqrt{\sum_{NG} \mathbf{BV}_{\mathrm{SST}}^2(t-1)}} \tag{3.1.1}$$

where *NG* is the total number of model grid points in the Niño3 region and *t* is the model time in months. In other words, we measure the growth rate of bred vectors by their amplification factor after one month. In the 10-year experiment, the typical value of the growth rate in one month of the NSIPP coupled model varies between 3 and 6, which is much larger than the coupled instability of about 1 to 3 per month found by Cai et al. (2003) for the ENSO mode in the ZC model. This is to be expected since the growth rate in the coupled GCM includes both coupled and uncoupled instabilities of any kind present in the system, which is much more complicated than the intermediate model. Figure 3.1.2 is an example showing how the background Niño3 index and the growth rate of the bred vector evolve with time. For comparison, we have removed the time mean of the growth rate (3.8 per month). The background Niño3 index has a seasonal dependence that peaks at the end of a year and the growth rate is large around the summer time as in Cai et al. (2003). In addition, we can see that the growth rate tends to have larger values when the background state is under a rapidly changing condition, for example, large growth rates occur before and after the warm event of year 2024. Therefore, we interpret the growth rate of 3-6 per month obtained from the CGCM as consisting of a noisy background growth rate of (mostly uncoupled) instabilities of about 4 per month (a component essentially absent in the ZC model), plus a coupled growth of about 1-2 per month, associated with the ENSO signal that we are seeking.

In order to test whether there is a component of the perturbation growth evolving upon the coupled ENSO background state (rather than growing randomly), we

calculate the lag/lead correlation between the growth rate and the absolute value of the background Niño3 index. We use the absolute value of the Niño3 index in order to account for the large amplitude of both positive and negative SST anomalies. It is evident in figure 3.1.3 that the growth rate of coupled bred vectors is sensitive to the phase of background ENSO and tends to be largest about 3-4 months prior to the time at which the background ENSO amplitude reaches its maximum stage (positive or negative). Our result is also qualitatively in good agreement with the results obtained with the ZC model in Cai et al. (2003), compared to figure 1.2.1.

Since the coupled GCM contains different types instability, the correlation level of 0.22, the maximum in figure 3.1.3, should be considered meaningful to the ENSO variability. In order to illustrate the point that the result is statistically significant enough to support our claim above, we construct the correlation between 1000 randomly generated time-series samples and the absolute value of background Niño3 index. The mean and variance of each sample are the same as the bred vector growth rates. The accumulated percentage of the values of correlations according to these 1000 samples is shown in figure 3.1.4. Among these randomly samples, the mean correlation value is 0.017 and the standard deviation is 0.07. Figure 3.1.4 indicates that the confidence level is close to 99% in stating that a 0.2 correlation value represents a significant correlation to the background Niño3 index. Therefore, the maximum appearing in figure 3.1.3 indeed is a result supporting that the bred vector growth rate leads the background Niño3 index.

The concept that the BV growth rate has a dependence on the phase of the background ENSO has been explained using a simple delayed oscillator model (Suarez and Schopf 1988). The delayed oscillator model for ENSO evolution can be written as

$$\frac{dT}{dt} = T - T^3 - \alpha T(t - \Delta) \qquad (3.1.2)$$

where, T represents the amplitude of the temperature anomalies, t is dimensionless time, Δ is the non-dimensional delay time, and α stands for the amplitude of the delayed signal relative to the linear and the nonlinear feedbacks. Taking T as the control background state, the tangent linear model corresponding to Eq. (3.1.2) is as

$$\frac{d\delta T}{dt} = (1 - 3T^2)\delta T - \alpha \delta T(t - \Delta) \qquad (3.1.3)$$

where, δT stands for the temperature perturbation. The growth of the temperature perturbation will have its maximum when the background state is at its neutral state (T=0). Therefore, our result indicates that the BV growth rate has reflected the unstable condition prior to the background ENSO event and thus proves our conjecture that breeding method can identify the slow, coupled instability related to ENSO variability. Such a characteristic has been pointed out by Evans et al (2004), showing that the BV growth rate can be used as a precursor in detecting the regime change of the background state (from warm condition to cold condition or vice versa).
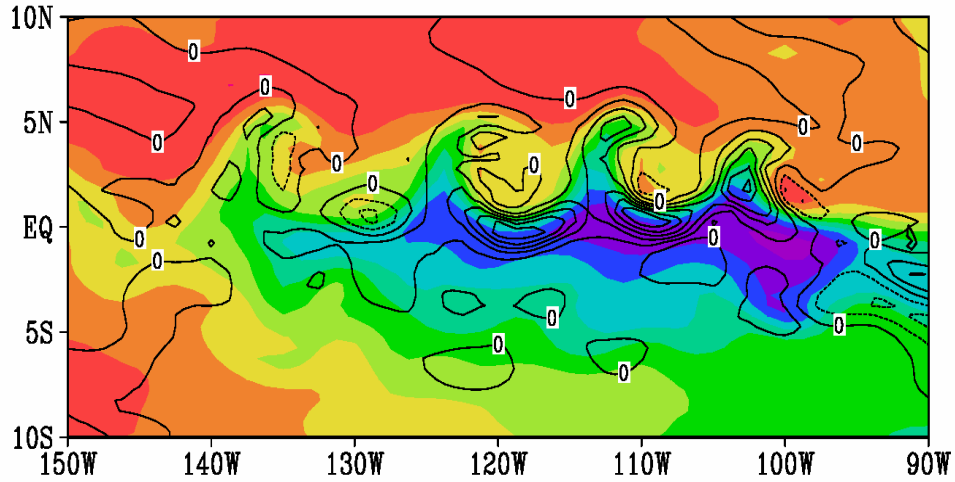
Figure 3.1.1 A snapshot of SST in Eastern Pacific showing the bred vectors perturbation (contour interval = 0.15 °C) evolving with the background flow (shadings with an interval of 1°C from 21°C to 30°C) on July 1 of the model year 2024. The dotted contours of the BV indicate negative values.
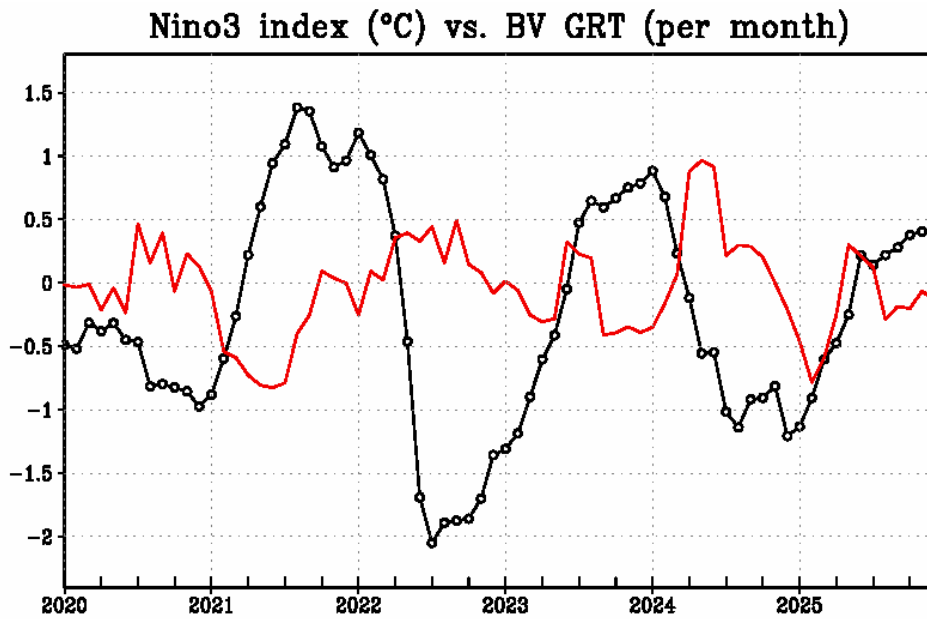
Figure 3.1.2 The background Niño3 index (unit: °C) vs. the growth rate of the bred vector (unit: per month). The time mean growth rate is 3.8 per month and is removed in this figure.

42
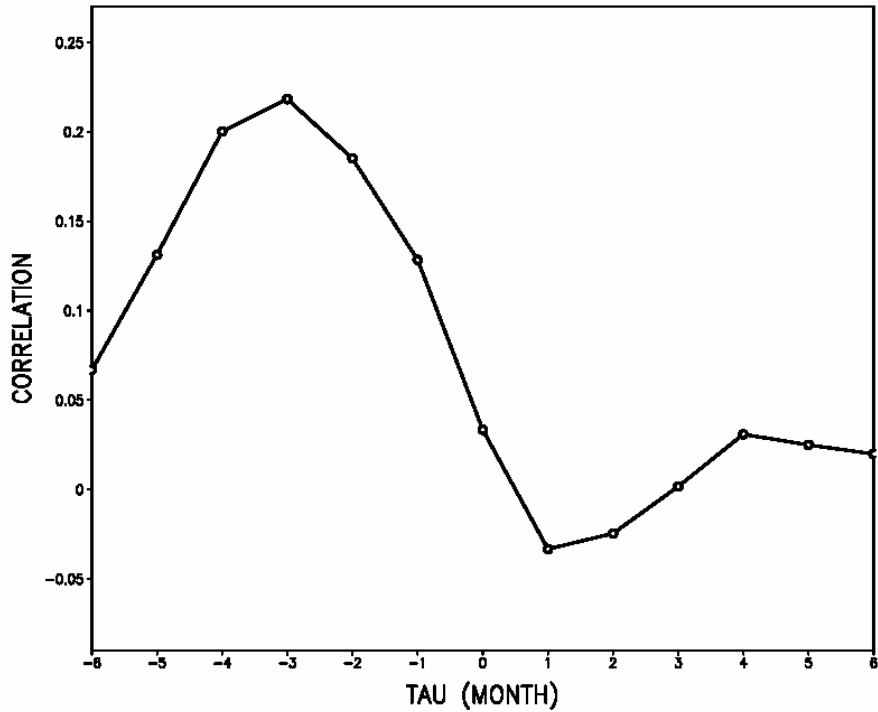
Figure 3.1.3 Lead/lag correlations between the BV growth rate and the absolute value of the background Niño3 index. Along the X axis, tau=0 stands for the mature stage of an ENSO event.
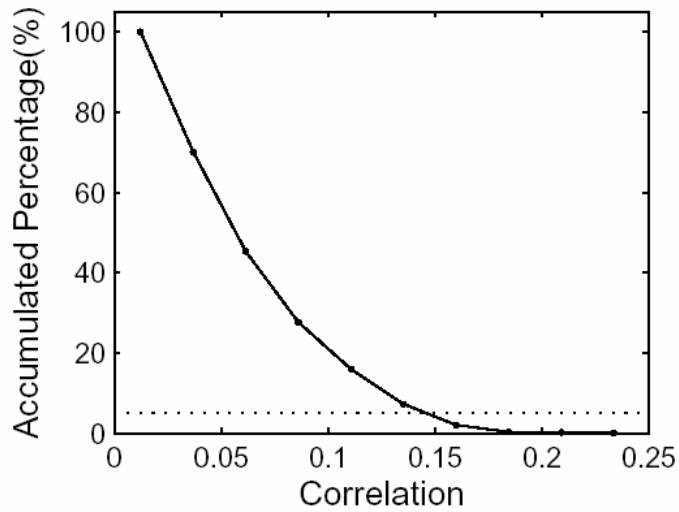
Figure 3.1.4 The accumulated percentage (%) constructed by 1000 randomly generated samples correlated with the absolute value of the background Niño3 index. Each sample has the mean and variance as the bred vector growth rate has. The dashed line stands for the 5% level.

3.2 The structure of the bred vectors

(a) Coupled BV modes

The spatial patterns of the coupled BV can be identified by constructing regression maps for both oceanic and atmospheric variables against the BV Niño3 index, defined as the spatial average of BV SST in the Niño3 domain. This regression will keep the low-frequency variabilities associated with ENSO instability and filter out perturbations unrelated with ENSO evolution. We will compare those maps with the background regression maps constructed with the same regression method but using the background Niño3 index in order to determine whether the coupled BV modes can capture the growing features associated with background ENSO variability.

The oceanic global regression maps for the background fields show the typical tropical variability corresponding to the ENSO mature stage (figure. 3.2.1(a)-(c)). These patterns include a large warming extending from the east to central equatorial Pacific, a deepening thermocline in the eastern equatorial Pacific, an accompanying shallowing feature off the equator in the western basin, and a basin-wide eastward current anomaly. The regression maps for the BV fields are shown in figure 3.2.1(d)-(f) representing the dominant coupled modes as the large growth rate occurs. The coupled BV mode exhibits a strong signal in the equatorial Pacific and fairly weak variability away from the tropics. The patterns of the coupled BV mode are reminiscent of those in the background state except that the BV mode is more confined to the east and to the equator. This feature is physically meaningful, since it

reflects a larger sensitivity to perturbations of the background flow in the shallower thermocline in the east along the equator. It is also consistent with the delayed oscillator theory, which considers that the growing perturbations are excited primarily over the eastern equatorial basin (Cane et al. 1990). It is known that in the mean the easterly wind stress is balanced by the zonal pressure gradient, which piles up warm water in the western basin, resulting in a thermocline that slopes down toward the west. The shoaling thermocline in the east implies that the thermodynamic feedback between SST and near surface ocean variables is much stronger in the east than in the west. As a result, oceanic perturbations either locally generated or coming from the delayed signals in the west and propagating into the eastern basin, will be easily amplified through positive feedbacks from air-sea interaction. This also suggests that the oceanic uncertainties are located mostly in the tropical eastern Pacific. Thus, if the initial perturbation of an ensemble member projects on coupled BV modes, we can expect that its forecasted SST anomalies should have large SST growth in the tropical eastern Pacific.

The background ENSO evolution has demonstrated a delayed oscillator behavior, associated with the equatorial wave propagations. The characteristics of the oceanic bred vector shown in figure 3.2.1 result from such tropical wave dynamics. Based on wave decomposition, we show that a large portion of variations of BV surface height and zonal currents is explained by the Kelvin mode and the first three Rossby modes [Appendix A]. This indicates that the coupled bred vectors obtained here have a strong component associated with the slow varying instability.

The atmospheric components of the ENSO mode derived from the control run and the BV field are displayed in figure 3.2.2. Since the regression maps are used to extract the linear response to lower boundary heating, it is rational that the regressed background wind anomalies in the tropics (figure 3.2.2(a)) exhibit a Gill-type model solution (Gill 1980). It shows that the patterns in the BV fields have several features in common with the patterns of the background state, such as the westerly wind perturbations located in the central equatorial Pacific, and the baroclinic structure in the height fields corresponding to the location of BV SST in figure 3.2.2(d). The BV outgoing radiation reflects an enhanced convective activity in the eastern basin. This atmospheric structure, implying that amplified perturbation in the eastern Pacific induces a westerly zonal wind perturbation, indicates that an unstable air-sea interaction in the eastern Pacific. These features reinforce our conjecture that the leading coupled BV mode is related to the coupled instability.

As we have mentioned in Chapter 2, the simulated SST variability from the NSIPP CGCM does not have enough variance near the coast in the eastern Pacific. An investigation suggests that such deficiencies may result from the atmospheric circulation in the tropical eastern Pacific not representing the Walker circulation well (Rienecker et al. 2001). Based on the high-low pressure patterns in figure 3.2.2(f) and (g), the coupled bred vector has shown their potential for ensemble forecasts since they are able to reflect the impact of coupled instability upon the background

atmospheric circulation (i.e. perturbing the longitudinal Walker circulation along the equator).

It is of interest to point out that the coupled BV also reflects the sensitivities in extratropical regions associated with background ENSO atmospheric teleconnections. Shown in figure 3.2.3 are the regression maps of surface pressure and geopotential at 200mb in Northern Hemisphere for the background state and for the BV field. The teleconnection patterns of the background state indicate a low-pressure anomaly over the North Pacific and a high-pressure anomaly over North America. This barotropic structure is very robust and extends to a high altitude. It is induced by wave-train patterns associated with the large scale heating in the tropics. For BV maps, strong responses can also be identified in those regions, especially where background regression maps show a strong gradient, for example in the mid Pacific at 30°N and east coast of North America. Wave-train patterns can also be found in BV regression maps. In the Southern extratropical region (not shown), atmospheric regression maps show a teleconnected pattern associated with background ENSO, and related BV dynamical sensitivities.
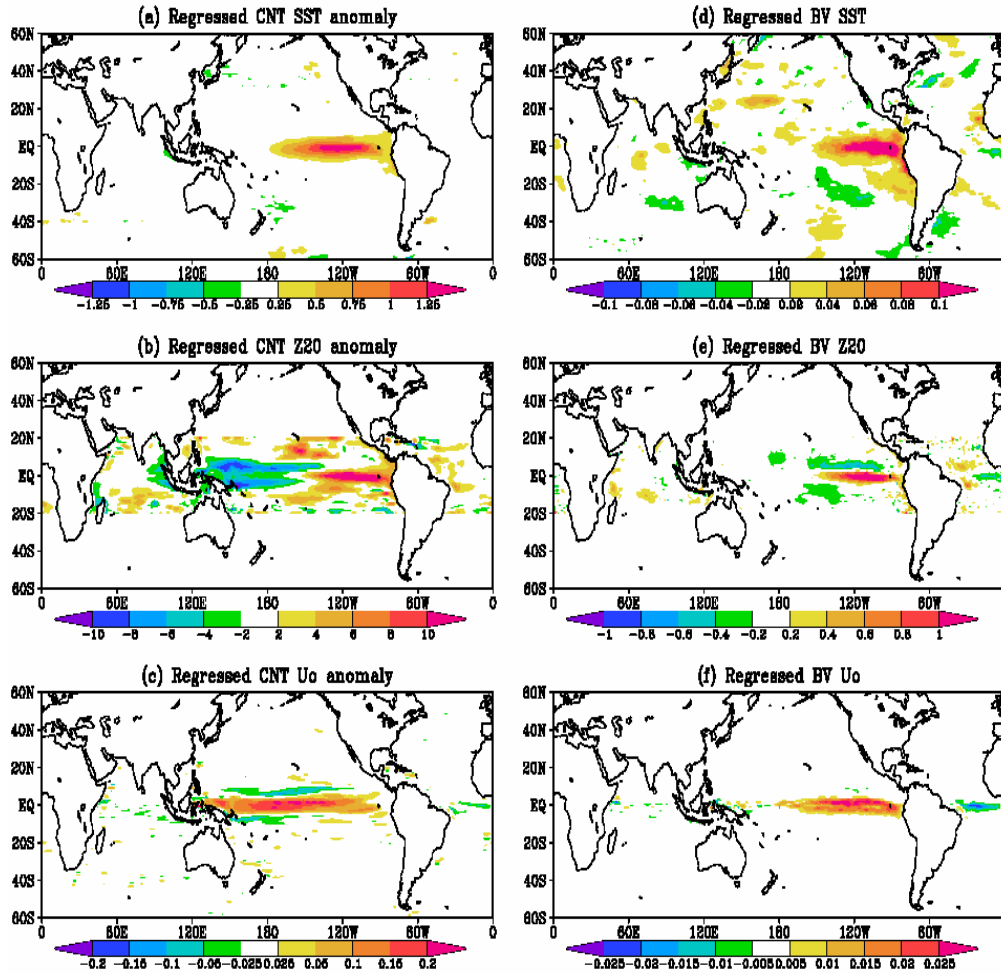
Figure 3.2.1 Oceanic regression maps in the global domain. Left panels are the background fields and right panels the BV fields. (a) SST anomaly (°C); (b) Z20 anomaly[1] (m); (c) surface zonal current anomaly (m/s). (d) BV SST (°C); (e) BV Z20[1] (m); and (f) BV surface zonal current (m/s). Background fields are regressed with the background Niño3 index and BV fields regressed with the BV Niño3 index[2]. The scales of BV fields are arbitrary but the ratio among BV variables (both oceanic and atmospheric variables) is retained as in the original.

---

[1] The background Z20 anomaly and BV Z20 are only plotted within 20°N/S. Because the definition for Z20 is the depth of 20 degree isotherm, it is not well defined beyond the tropics.

[2]The BV Niño3 index is defined as the area mean of BV SST in the Niño3 domain.
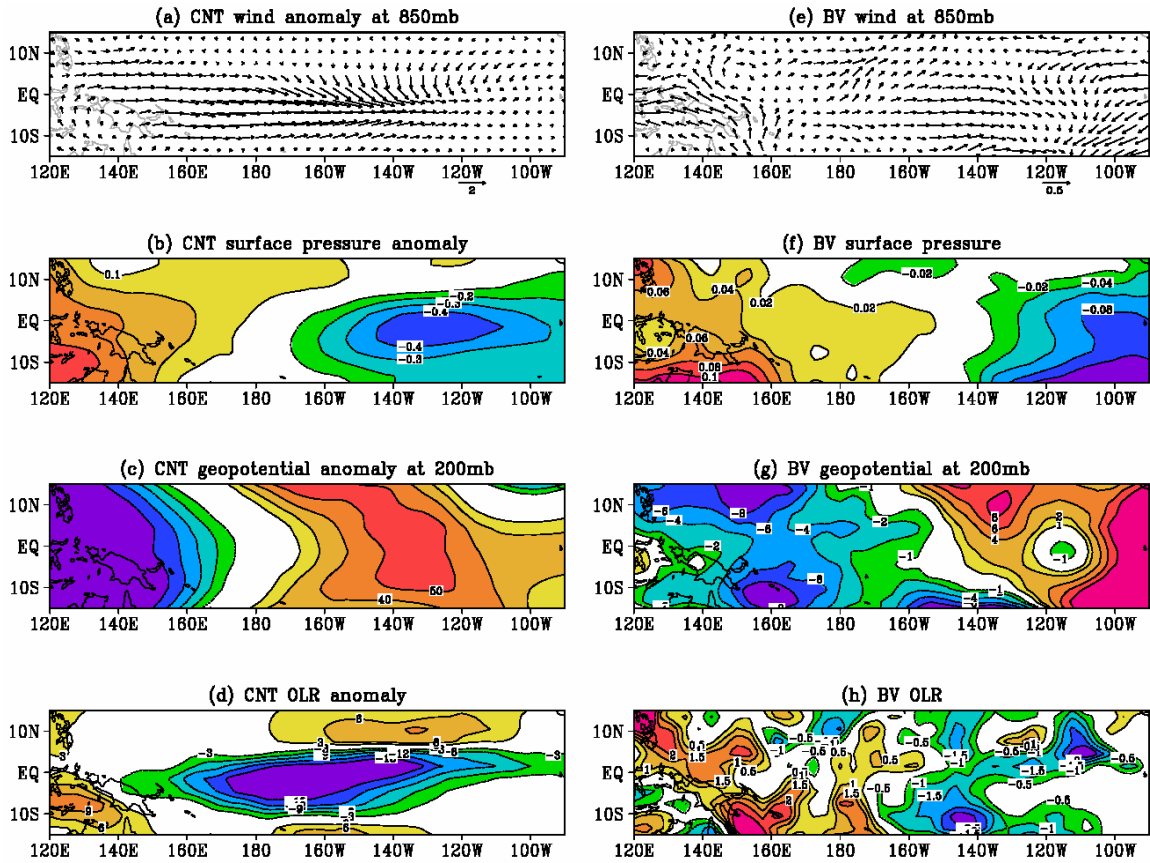
Figure 3.2.2 The same as Figure 3.2.1 except for the atmospheric regression maps over the equatorial Pacific basin. (a) wind field anomaly at 850mb (ms[-1]); (b) surface pressure anomaly (mb); (c) geopotential anomaly at 200mb (m[2]s[-2]); (d) outgoing long wave radiation (Wm[-2]); (e) BV wind field at 850mb (ms[-1]); (f) BV surface pressure (mb); (g) BV geopotential at 200mb (m[2]s[-2]); and (h) BV outgoing long wave radiation (Wm[-2]).
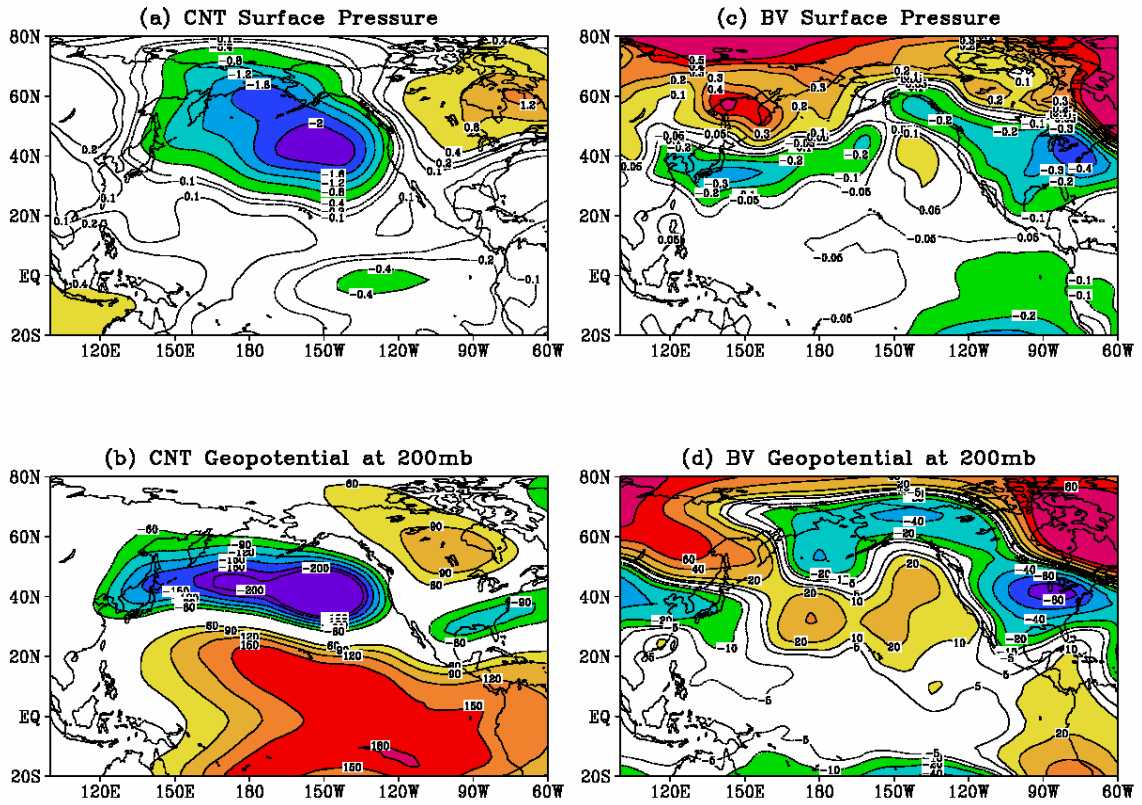
Figure 3.2.3 The same as Figure 3.2.1 except for atmospheric regression maps over the Pacific portion of the Northern Hemisphere. (a) background surface pressure anomaly (mb); (b) background geopotential anomaly at 200mb (m^2s^{-2}); (c) BV surface pressure (mb); and (d) BV geopotential at 200mb (m^2s^{-2}).

(b) The evolution of bred vector and its relationship to background ENSO

In the previous subsection, the regression maps provided the spatial distribution of coupled modes, including background and bred vector. The structures of bred vectors "inherit" the coupled variability from the background. Although the bred vectors in the tropical Pacific are characterized by strong structures in the equatorial eastern Pacific, such structures are not simply standing or oscillating alone at the surface in this location. Instead, the dynamic movement following a delayed oscillator mechanism appears also in the subsurface condition of bred vectors.

In the delayed oscillator theory (Schopf and Suarez, 1987; Kirtman 1997), the Rossby waves off the equator play an important role in determining the characteristics of an ENSO event, contributing to the time of formation and termination of an event. They are driven by the curl of the wind stress from the central Pacific. During the 10-year simulation, the meridional distribution of the curl of wind stress in the central Pacific (the Date line) is strongest at 5°S. Therefore, in the following, we examine the temporal evolution of the background anomaly and bred vector along this latitude.

Figure 3.2.4 (a) is the surface height anomaly (color shading) and bred vector surface height (contour) from December 2020 to December 2022 along 5°S. During this period, a strong El Niño event started on April 2021 and terminated on March 2022, as the consequence of the downwelling Kelvin wave in figure 3.2.4(c). A La Niña event then started on May 2022, which is clearly associated with the reflection

51

of the upwelling Rossby wave at December 2021. In figure 3.2.4(a), the bred vector also shows a westward propagation, which is a few months ahead of the background anomaly. The leading characteristic implies that the coupled instability triggering tropical waves does exist in the bred vector, and suggests its ability to estimate the occurrence/location of the instability. The forcing (background curl of the wind stress anomalies) that drove the Rossby wave is shown in figure 3.2.4 (b). This atmospheric forcing leads the surface height anomaly by about 2-3 months for both the warm and cold events. We found that bred vector and the background curl of wind stress exhibit a simultaneous behavior before these events (the dashed lines in figure 3.2.4(a), (b) denote the westward propagation in BV surface height). Such a temporal relationship suggests that the bred vector off the equator reflects the coupled instability originating from the atmosphere, even though the rescaling is simply done in the Niño3 region.

In order to show how the coupled BV mode evolves with the background ENSO evolution as in Cai et al. (2003) with the ZC model, we construct lead/lag regression maps against the time series of the amplitude of the background Niño3 index with a lead/lag time up to 6 months (figure 3.2.5). It is clear that the temporal evolution of the coupled BV mode is highly related to the background ENSO evolution. It shows that in the eastern basin the coupled BV mode leads the large amplitude of the background ENSO events by several months. The signal is clearly coming from the coupled dynamics because an increase of the ocean heat content and a warm SST anomaly in the eastern basin are accompanied by the presence of westerly wind anomalies. This coincides with the timing of the maximum growth of the coupled BV

mode, which also leads the background ENSO events by 3 months (figure 3.1.3). In figure 3.2.5 (c), the lead time in the BV zonal windstress is short and the large response is located in the central Pacific before and after the event. In addition, it is seen from figure 3.2.5(b) that west of 130°W, BV surface height and zonal wind stress exhibits a lagged response to the background ENSO by about 3 months. Figure 3.2.5 indicates that the dominant instability in the NSIPP GCM model is a coupled instability located in the eastern basin that precedes the development of background variability.

(a) Hsurf at 5°S

(b) Curl of Wind stress (x1.e7) at 5°S

(c) Hsurf at Equator
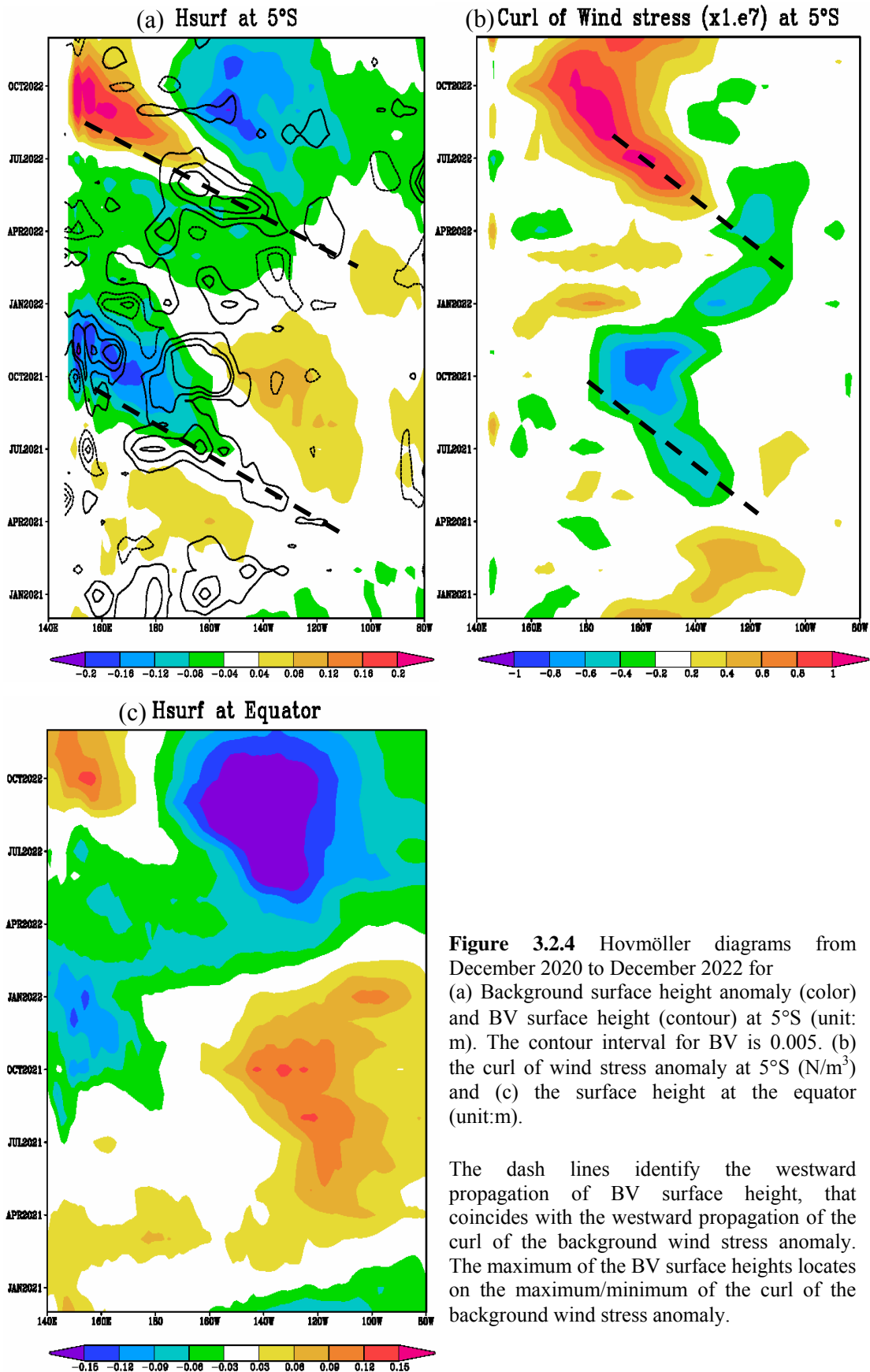
Figure 3.2.4 Hovmöller diagrams from December 2020 to December 2022 for
(a) Background surface height anomaly (color) and BV surface height (contour) at 5°S (unit: m). The contour interval for BV is 0.005. (b) the curl of wind stress anomaly at 5°S (N/m^3) and (c) the surface height at the equator (unit:m).

The dash lines identify the westward propagation of BV surface height, that coincides with the westward propagation of the curl of the background wind stress anomaly. The maximum of the BV surface heights locates on the maximum/minimum of the curl of the background wind stress anomaly.
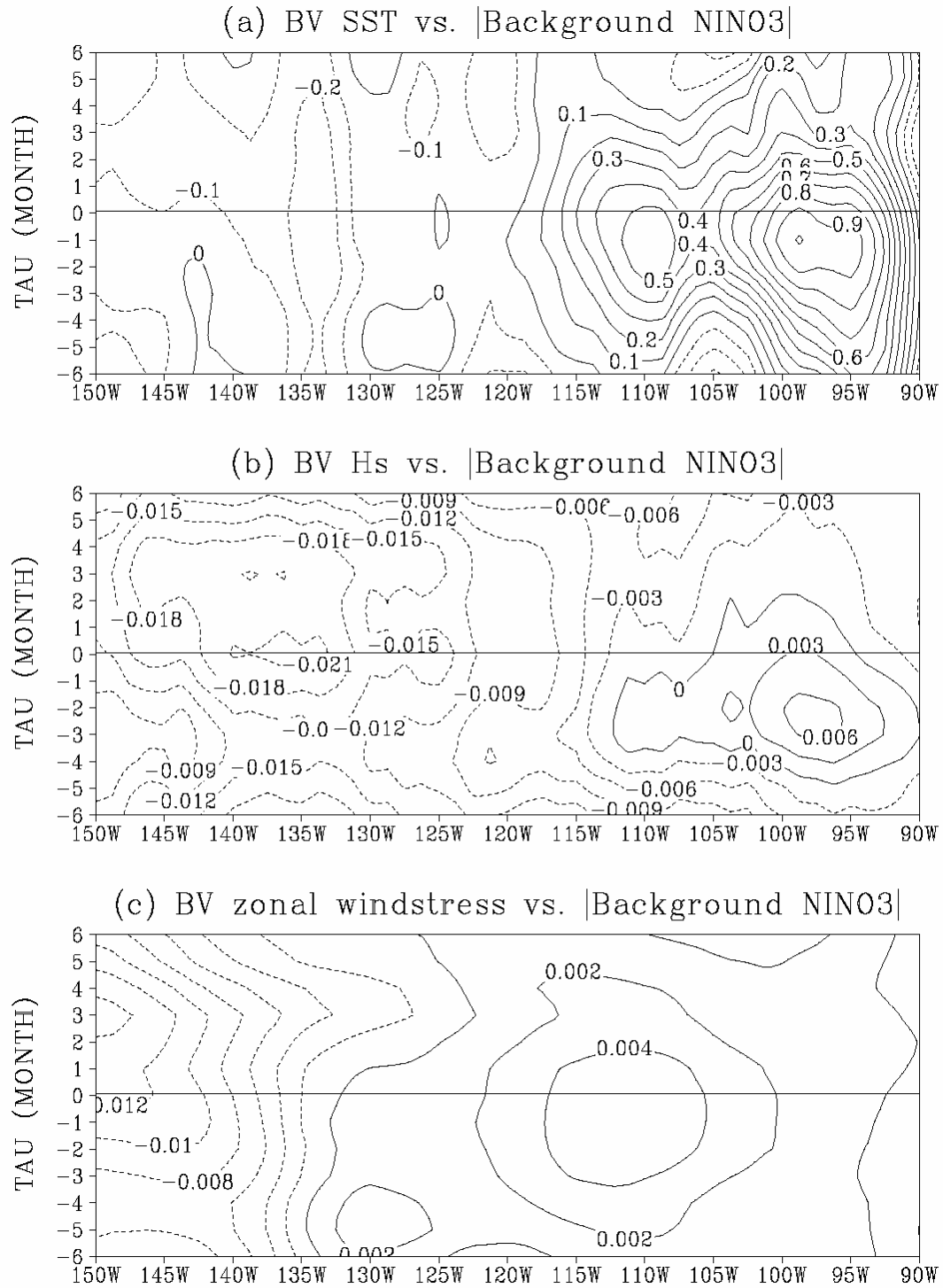
Figure 3.2.5 Lead/Lag regression maps along the equator for BV oceanic fields against the absolute value of the background Niño3 index. (a) SST (°C); (b) surface height (m); and (c) zonal wind stress (Nm^{-2}). The contours are arbitrary but the ratio among BV variables is retained as the original BV fields

3.3 Comparison of the NSIPP and the NCEP/CFS03 Bred Vectors

Similar breeding experiments were carried out with the coupled forecast system model (CFS03) developed in the National Centers for Environmental Prediction (NCEP). The atmospheric component uses the current version of medium range forecast (MRF) global model with a spectral truncation of 62 waves (T62) in the horizontal (equivalent to nearly 200 Km) and 64 vertical levels in sigma coordinate (Kanamitsu 1989; Kanamitsu et al. 1991; Caplan et al. 1997; Wu et al. 1997, Saha et al. 2004). The ocean component is the GFDL Modular Ocean Model V.3 (MOM3) with 40 layers in the vertical (Pacanowski and Griffies, 1998). The zonal resolution is 1° and the meridional resolution is 1/3° between 10°S and 10°N, gradually increasing through the tropics until it is fixed at 1° poleward of 30°S and 30°N.

Two independent breeding experiments were performed by choosing the last 4 years from a 23-year perfect model experiment as the background state. This 4-year period covers a warm event which matures at the model year 21, 2 years into the breeding run. The rescaling factor for perturbations is based on the SST norm in the whole tropical belt (10S-10N) and the perturbation size was chosen as 0.1°C. As in the breeding experiments performed with the NSIPP CGCM, we chose one-month as the rescaling period. Like the NSIPP coupled experiments, the two BV runs for the NCEP system were very similar despite having been started with different random perturbations so that their results are processed as a single 8-year time series.

Comparisons between the results from the NSIPP and the NCEP/CFS03 coupled systems are made for the purpose of exploring whether the bred vectors in two different coupled GCMs[3] share common characteristics.

Figure 3.3.1 (a)-(f) are background oceanic regression maps of two coupled GCMs. The oceanic components from the two GCMs successfully produce fundamental features of ENSO. Their differences also reflect differences of numerical schemes in the model dynamics or different choices of physical parameterizations. The meridional structure of warming and thickening surface height (figure 3.3.1(a), (b)) in the NCEP/CFS03 GCM in the Eastern Pacific is wider than that of the NSIPP GCM (figure 3.3.1(d), (e)). In addition, the regressed surface height of NCEP/CFS03 shows the southern branch of the shoaling patterns off the equator extends more southward instead of being meridionally limited as in the NSIPP case. This can also be seen in the SST and zonal current patterns. Despite those distinctions, the bred vectors from the two coupled systems have significant similarities linked with the background ENSO. To compare BV structures, we show the EOF modes of oceanic variables. Figure 3.3.1(a)-(c) are the first EOF mode of the BV SST and first two modes for the BV thermocline from the NSIPP CGCM and fig. 3.3.1(d)-(f) are the same modes using BV from NCEP/CFS03. Here, we use the first two EOFs for thermocline because in the background thermocline variability, it has two dominant modes along the equator controlling the phase of the ENSO event. There is one dominant mode in background SST variability. Despite the fact that

---

[*] Unfortunately, the experiments performed at NCEP were erased, so that we have only a limited number of diagnostic comparisons available.

these are two different CGCMs with very different background evolution, there is a strong resemblance between the BV EOF modes. Both the leading modes (EOF1) in NSIPP and NCEP/CFS03 bred vectors based on SST show an ENSO-associated warm feature in the tropical Eastern Pacific, farther east than in their respective background. Reflecting the different mean structures and background ENSO variabilities from a different coupled system, the NCEP/CFS03 BV SST EOF1 extends over a larger spatial scale, covering the whole Niño3 domain while the BV SST EOF 1 from NSIPP model is confined to east of 130°W and is meridionally limited. The NCEP EOF1 mode explain 11% and the NASA/NSIPP EOF1 modes 14% variance from the total growing SST perturbations. This suggests that the coupled growing perturbations associated with ENSO variability represent at least 10% of the total growing perturbations within the variety of instabilities that appear in a coupled GCM. *The fact that the leading EOF modes from BV fields in both coupled systems show an ENSO-like structure confirms that the breeding method can capture the coupled instability even in the presence of other types of instabilities in the fully coupled GCM model.* Moreover, this mode seems to be robust and dominant. Even if we enlarge the domain for the EOF analysis from the tropics to a global domain, the same mode appears, showing a fairly weak oceanic amplitude in the extra-tropics. This indicates that the breeding method can help to identify the largest growing error projecting on the ENSO variability in a global coupled model and can be applied to a model with full, complete physics such as a GCM, with the full resolution of model output.

Similar natural sensitivities in the eastern Pacific can also be found in the BV thermocline fields of both systems (Figure 3.3.2 (b), (c), (e) and (f)). Both leading EOF modes of the BV thermocline have a deepening feature along the equator and shoaling features off the equator, except that the EOF1 from NCEP/CFS03 extends across almost the whole basin. In addition, both EOF2 modes have a dipole pattern along the equator and establish a wave couplet off the equator in the western basin. All the information above reveals that oceanic perturbations will develop as Kelvin/Rossby wave packages, propagating the upwelling/downwelling signals in the tropical region, as we have discussed in the previous subsection. It should be noted that the EOF1 of the BV thermocline is close to the EOF1 of BV SST and there is a high correlation between their corresponding leading principal components (not shown), and also with its own Niño3 index. These two modes represent the dominant growing coupled instability, which has been retrieved by the BV oceanic regression maps (figure 3.2.1 (d)). The robustness of the results from two different coupled models supports our hypothesis that bred vectors are associated with the background ENSO variability. The differences between the BVs indicate that bred vectors are sensitive to model characteristics. For example, different vertical mixing schemes adopted in ocean models will have an impact on thermocline variations particularly in the shallow mixed layer region.

Based on the regression maps against with the BV Niño3 index (not shown) in the tropical region, we can also deduce that the coupling strengths are different in the two coupled GCMs. In both ocean models, a 1-meter variation (deepening) in BV

thermocline corresponds to 0.1 ºC warming in the eastern Pacific. The corresponding BV zonal wind stress shows a perturbation of 1.5 Nm^{-2} from the NSIPP AGCM, and it prevails in the central basin. By contrast, the corresponding stress perturbation is only 0.5 Nm^{-2} in NCEP/CFS03 case. In addition, the regressed BV surface pressure and geopotential height for the NCEP/CFS03 model are less organized in the tropics than for the NSIPP coupled model. This seems to suggest that perturbations are more strongly coupled in the NSIPP CGCM in the tropical domain.

There are also similarities between the two systems in the extratropical ENSO-associated teleconnection patterns. Figures 3.3.3 (a), (b) are the regression maps of BV geopotential height at 500mb from the two coupled models. Similar responses can be identified from the eastern basin of the North Pacific to the North Atlantic despite the different responses for other locations. The atmospheric seasonal to interannual variabilities in this ENSO-teleconnected region are constrained by the deep tropical heating. The resemblance for bred vectors in these locations indicates that the bred vector could be used as an "effective" ensemble member for the purpose of capturing the global seasonal-interannual related features. Additional experiments with NCEP/CFS03 CGCMS have confirmed that the relationship between the BV growth rate and the phase of ENSO events remains similar with a rescaling period of 15-day rather than the one month used so far (M. Peña, personal communication).
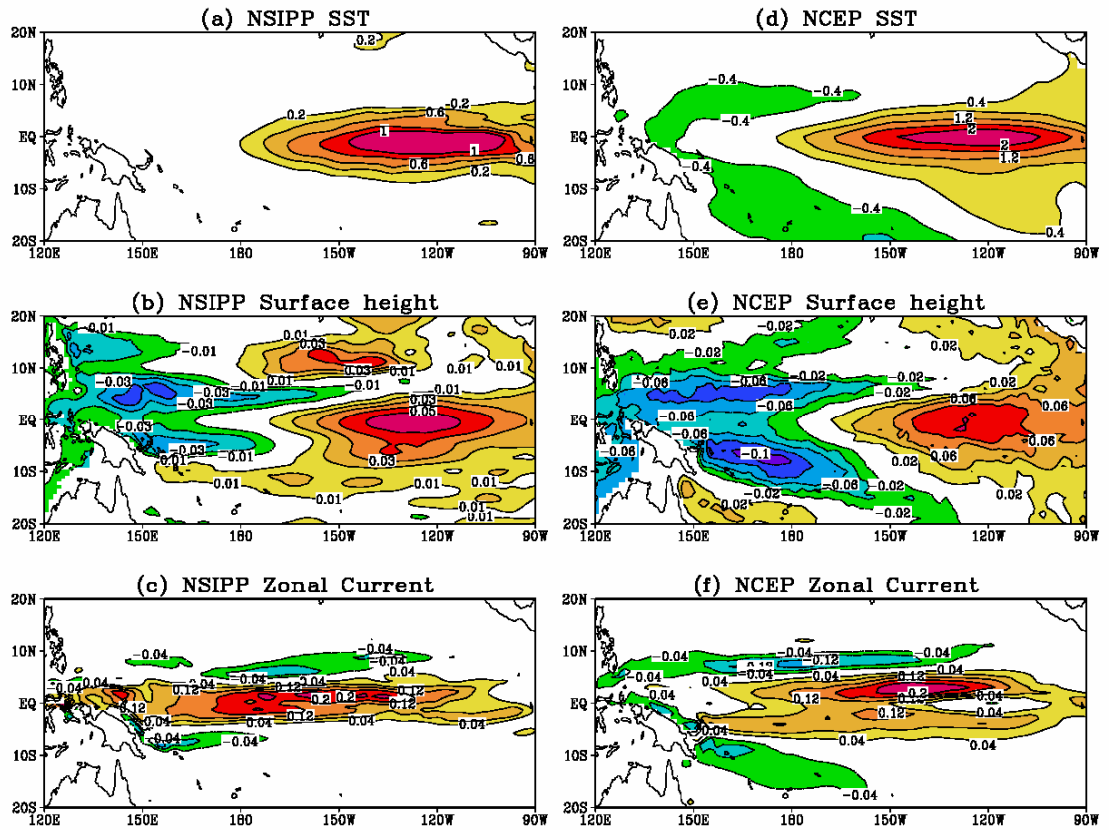
Figure 3.3.1 Background oceanic regression maps for two coupled GCMs in the tropical Pacific domain. Left panels are the NSIPP anomalies and the right panels the NCEP/CFS03 anomalies. (a) NSIPP SST (°C); (b) NSIPP surface height (m); (c) NSIPP surface zonal current (m/s); (d) NCEP SST (°C); (e) NCEP surface height (m); and (f) NCEP surface zonal current (m/s). The regression maps of NSIPP (NCEP) fields are computed using the NSIPP (NCEP) Niño3 index.
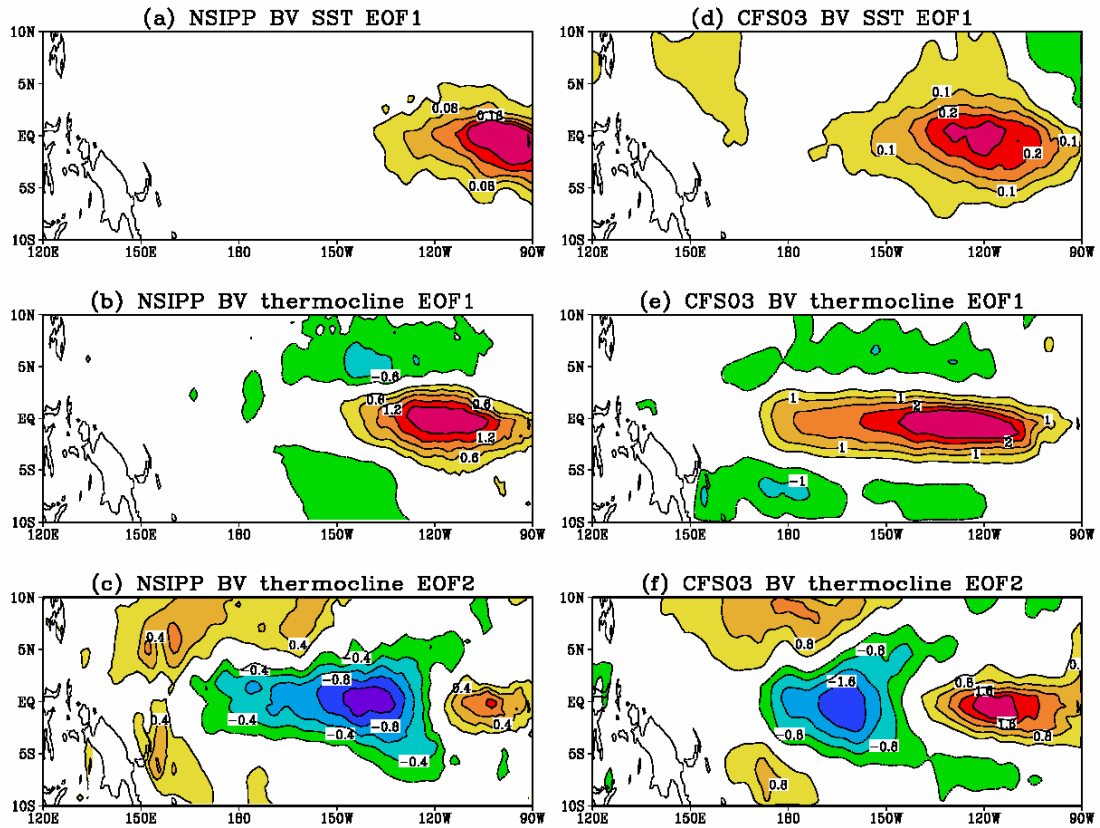
Figure 3.3.2 The leading EOFs of the BV SST and Z20 perturbations derived from the NSIPP and NCEP/CFS03 CGCMs. (a) EOF1 of NSIPP BV SST; (b) EOF1 of NSIPP BV Z20; (c) EOF2 of NSIPP BV Z20; (d) EOF1 of NCEP BV SST; (e) EOF1 of NCEP BV Z20; and (f) EOF2 of NCEP BV Z20. The scale is arbitrary.

(a) NSIPP regressed BV geopotential height at 500mb
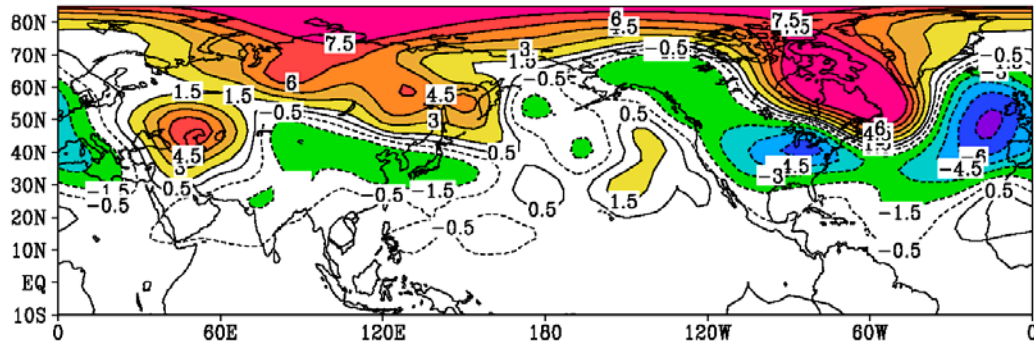
(b) CFS03 regressed BV geopotential height at 500mb
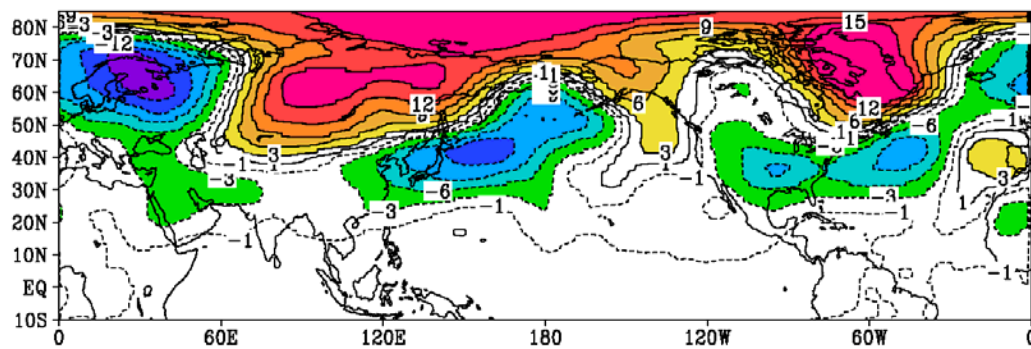
Figure 3.3.3 Atmospheric regression maps in the Northern Hemisphere of BV sea-level pressure (mb) (a) for NSIPP and (b) for NCEP/CFS03. Both fields are computed against their own BV Niño3 indices.

3.4 Bred vectors and tropical instability waves

In subsection 3.1, we have seen that bred vectors can capture the developing tropical instability wave at the edge of the cold tongue (figure 3.1.1). We notice that such instabilities are particularly active during the cold seasons and the La Niña events. In this section, we will show that breeding can be used as a dynamical tool to detect the tropical instability waves by capturing such instabilities in the BV fields. We also show their relationship with the background seasonal-to-interannual instabilities.

The presence of tropical instability waves is closely related to the evolution of background ENSO events and the seasonal cycle because these variations control the establishment of the cold tongue (Contreras 2002). The characteristic of tropical instability waves can also be represented by the CGCM and thus bred vectors naturally reflect the existence of such instabilities. Here, we illustrate their relationship by focusing on the domain located at the northern edge of the cold tongue away from the coast (140°W-110°W, 0.5°N-6°N). The average temperature gradient within this domain as shown in figure 3.4.1 (b), is particularly strong during the La Niña year (figure 3.4.1(a)) and also strong in late summer-fall season, but vanishes during the El Niño year. Also, after year 2025, the background SST remains as a weak but long-lasting El Niño event and the gradient of the SST in this domain is dominated by the seasonal variation. Those characteristics in figure 3.4.1 (b) are also

apparent in the BV SST represented by rms of BV SST in this domain (figure 3.4.1(c)). The amplitude of BV SST is particularly high during the La Niña year and small during the El Niño event, indicating that they are evolving upon the background flow. The Hovmöller diagram of bred vector and background SST in figure 3.4.2 demonstrates that both the instabilities develop along 4.5°N and propagate westward. In this figure, it shows that the background tropical instability waves commence during a La Niña event in June 2022 and the westward propagation is captured by the bred vector. The development of tropical instability waves in the bred vector can be seen as early as April 2022 (figure 3.4.3), indicating that the instability wave will grow upon the northern edge of the cold tongue. In addition, such structures of BV disappear when the background starts to evolve into a warm condition.

The results obtained with the perfect model scenarios have shown that bred vectors can capture the slow-growing instability, amplifying in the eastern Pacific and also detect the tropical instability wave wherever the background establishes a sharp SST gradient along the cold tongue. This encourages us to explore their potential applications in depicting the monthly error structures in more complex situations where real oceanic observations are assimilated. The result of the breeding method implemented in the current NSIPP operational forecast system will be presented in the next Chapter.
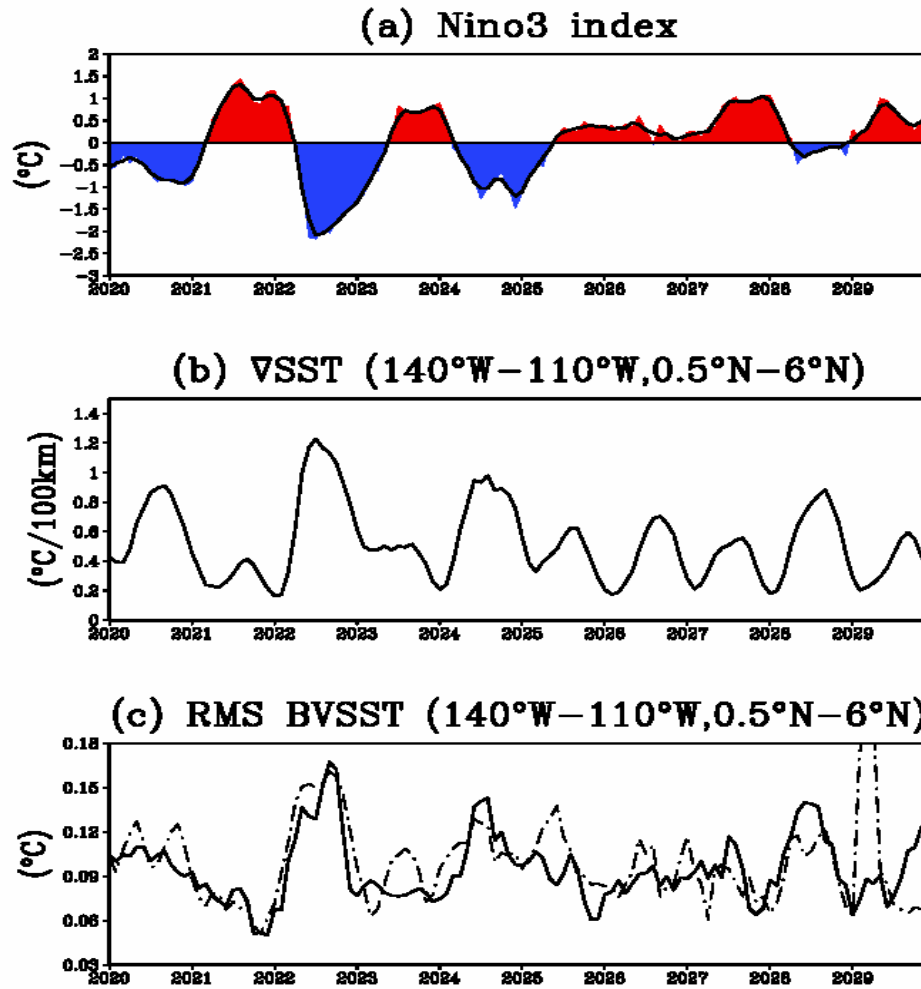
Figure 3.4.1 Time series plots for (a) Niño 3 index (°C) (b) average of SST gradient[*] within 140°W-110°W and 0.5 °N and 6°N (°C/100KM) and (c) RMS of BV SST (°C) from two independent breeding experiments

---

[*] $\left\| \nabla SST \right\| = \sqrt{\left[\dfrac{\partial SST}{\partial x} \right]^2 + \left[\dfrac{\partial SST}{\partial y} \right]^2}$
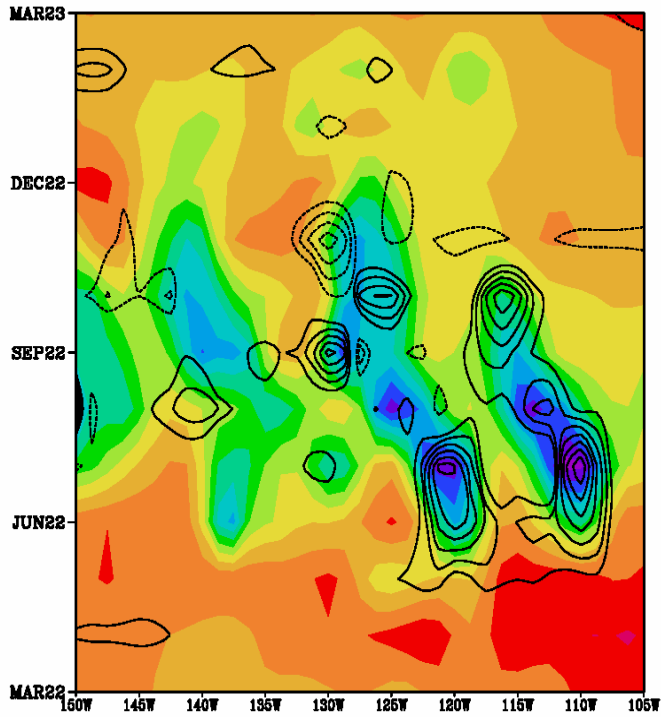
Figure 3.4.2 Hovmöller diagram for background SST (color) and BV SST (contour) at latitude 4.5°N from March 2022 to March 2023
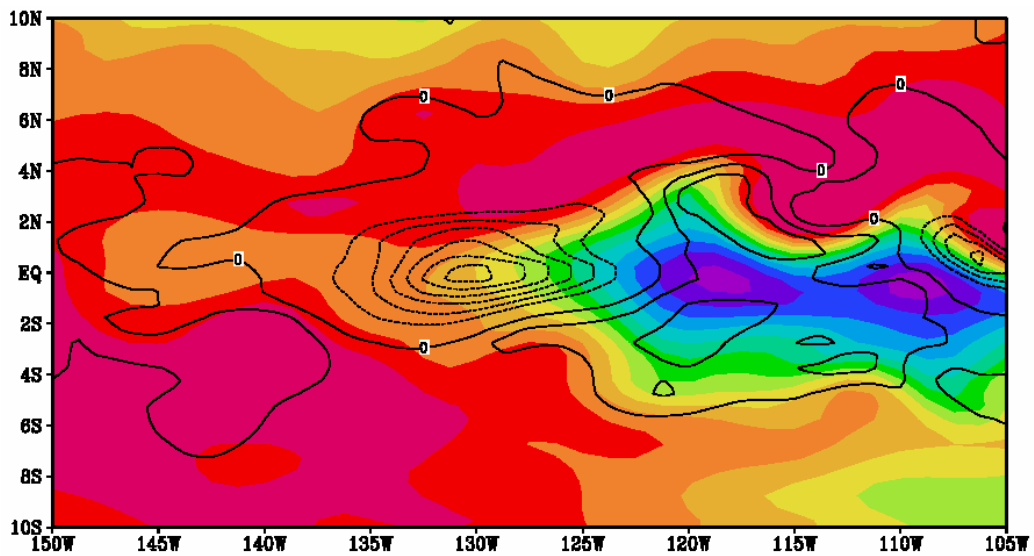
Figure 3.4.3 A snapshot of SST in the central Pacific showing the bred vectors perturbation (contour interval = 0.1 °C) evolving with the background flow (shadings with an interval of 0.5°C from 24°C to 30°C) on April 1 of the model year 2022. The dotted contours of the BV indicate negative values.

67

3.5 Comparisons between BV and SV in NSIPP CGCM

In this final subsection, we would like to compare our results with the work of Kleeman et al. (2003) using the same NSIPP CGCM. In their work, they constructed the linear propagator based on only one variable, SST, and with an optimization time of 6 months. The propagator was derived based on the subspace spanned by 5 dominant correlation EOF modes of the background SST anomaly.

Since the propagator used in the SV is only based on SST, the growth of the SST in the tropical Pacific is only determined by SST. They used the correlation EOFs to emphasize perturbations that have small amplitude but large-scales, particularly in the western Pacific. They also believe this is a way to include the effect of the air-sea interaction on the warm pool. However, their perturbations are mainly growing in the eastern Pacific and they can have a significant growing perturbation when the norm is in the eastern Pacific. Their initial SV (figure 3.5.1, from Kleeman et al. 2003) shows a very large-scale warming pattern from central to eastern Pacific and very strong in the central Pacific. In contrast, our results show the importance of the central Pacific lies in the subsurface thermocline (figure 3.3.2(c)). The final SV (not shown) is very similar to our ocean regression map (figure 3.2.1(d)) with a strong amplitude in the eastern Pacific representing the mature state of ENSO (figure 3.2.1(d)).

So far, there is no clear result showing that the initial SV is sensitive to the phase of the ENSO cycle. By contrast, the bred vector is sensitive to the phases of the

ENSO cycle and the structure also shows a delayed-oscillator behavior (Appendix A).

Also, the smallest growth rate of SVs was obtained when initialized from a simulated December, which Kleeman et al (2003) attribute to the fact that it corresponds to the onset of a La Niña event. In our case the BV growth rate is small during winter and early spring season (figure 3.1.2(b)). However, the result does not depend on whether it is the El Niño or the La Niña event. The BV growth rate is usually smallest at the mature phase of the ENSO event, since the uncertainty in the Niño3 region is saturated.
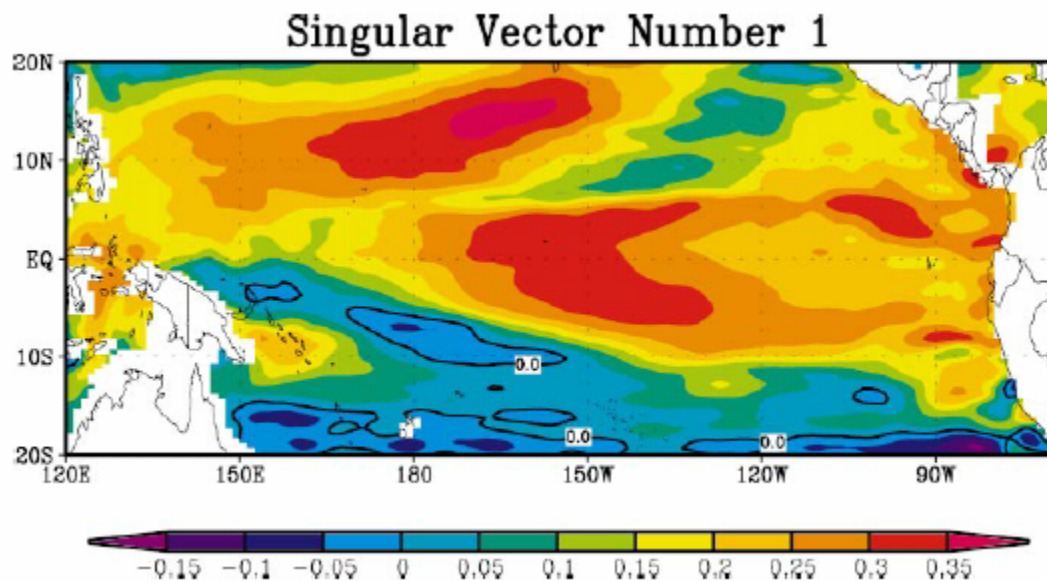
Figure 3.5.1 The climate-relevant singular vector for SST derived from NSIPP CGCM (from figure3 in Kleeman et al. 2003)

Chapter 4: Breeding experiments in NSIPP operational system

4.1 Introduction

In Chapter 3, we have shown the characteristic of bred vectors in a perfect model experiment. Now, we perform similar experiments with a much more challenging system: the operational CGCM with real observations involved. We introduce the discussion by summarizing the main results we obtained:

- Bred vectors show structure similar to the analysis increments (estimated one-month forecast errors), which are defined as the difference between the analysis and the one-month forecast. The agreement between the bred vector and the forecast errors is particularly good when the bred vector growth rate is large.

- Both the bred vector and the analysis increments are very sensitive to the phase of an ENSO event. Also, the subsurface structures of both the bred vector and the analysis increment project mostly on the variability associated with the ENSO variability.

- The bred vector growth rate has a strong annual cycle, indicating a strong relationship to the background seasonal variation and explaining the "spring barrier" in forecasting skill.

- Our results show that the BV structures are insensitive to the choice of rescaling norm. The local projection of the analysis increment on the subspace

spanned by three bred vectors is larger than the subspace spanned by three operational perturbations. This is valid for both tropics and extra-tropics.

We also performed preliminary experiments of ensemble forecasting to explore the impact from using coupled bred vectors in operational ENSO prediction. Bred vectors are used to represent the dynamic perturbations in the initial condition for the purpose of the effective growth of perturbations. As a result of these experiments, NASA has decided to incorporate these dynamic perturbations in their ensemble prediction system.

4.2 The relationship between bred vector and analysis increments

The procedure we used to perform the breeding cycle in the NSIPP operational system is the same as described in Chapter 1, with a rescaling period of one month. The rescaling factor is measured by the SST bred perturbations in the Niño3 region with an amplitude of 0.085ºC. The operational system has been described in Chapter 2 and includes coupling with an AMIP atmospheric state. In the coupled breeding cycle for an operational system, oceanic-bred perturbations are added to oceanic analysis fields and atmospheric bred perturbations are added to AMIP-style restart fields. Here, breeding experiments, including the control forecasts, are generated via the updated NSIPP operational CGCM from January 1993 to November 1998.

(a) The characteristics of bred vector and analysis increment during 1997-1998 El Niño evolution

In order to illustrate how bred vectors evolve upon the background ENSO variability, we examine the time evolution of the bred vector and the analysis increment with respect to the Niño 3 index. During ENSO evolution, variations in the subsurface have a strong influence on the SST anomalies, particularly in the eastern Pacific. The onset of an ENSO event has been traced back to perturbations in the subsurface of the western Pacific, traveling to the eastern Pacific through the motions of equatorial waves. We found that the analysis increment of temperature near the equator, defined as the difference between the analysis and forecast temperature, also carries this eastward propagating characteristic associated with Kelvin waves. This quantity (analysis increment) represents the optimized correction for the forecast state after combining observational information and as such is a proxy for *forecast error.* Here, we selected four months in order to illustrate the characteristic of the analysis increment corresponding to different stages of ENSO evolution (prior, developing, mature, and post stages). Figure 4.2.1(a) is the Niño3 index from September 1996 to September 1998, covering a strong El Niño event that peaked at the end of 1997. Figure 4.2.1(b)-(e) are snapshots of vertical cross sections along the equator of analysis increment (color shading) and bred vectors (contour) of temperature corresponding to prior, developing, mature and post stages. These snapshots are plotted in model coordinates (levels). We found that the analysis increment is mainly located in the subsurface of the western Pacific and close to the eastern coast before the warming commences as shown in figure 4.2.1(b). The analysis increment shifted eastward and extended vertically as the warm anomaly started building up in the

eastern Pacific (figure 4.2.1(c)). At the mature stage of the event, the analysis increment/forecast errors were smallest near the eastern coast and upper levels (figure 4.2.1(d)). When the warm anomaly diminished and the background state resumed to normal condition, the analysis increment accumulated mostly in the eastern Pacific (figure 4.2.1(e)). It is remarkable that the same longitude-vertical variations also appear in the bred vectors. As shown in figure 4.2.1(b)-(e), bred vectors (contour) are characterized by eastward propagation synchronized with the analysis increment movement and their shapes tend to capture the large-scale analysis increment (i.e., the forecast error). Our results support our claims that the bred vectors obtained from such a complicated CGCM with data assimilation do include information related to realistic ENSO development and, further, indicate where the forecast errors will be located.

Despite not knowing the observations that lead to the analysis corrections, the bred vector contains the dynamic errors that appear in the correction term from data assimilation. Good agreement between bred vector and analysis increment in the model coordinate suggests their potential application in the data assimilation process. The shapes of the dynamic errors can be obtained from the bred vectors in model coordinates. The characteristic error shapes can be used to augment the background error covariance used in data assimilation, as done by Corazza et al. (2002).

The similarity between bred vectors and forecast errors remains after we convert bred vector and analysis increment from levels to depths by combining the variations

of model thickness and temperature. Figure 4.2.2 is an example to show the error distribution in both model levels and physical depths. A large analysis increment was located near the thermocline, where sharp temperature gradients can easily introduce instabilities. As could be expected, both the analysis increment and bred vector maximum occur near the thermocline if we convert them to depths. The structure of the bred vector suggests it will be effective when used as an initial ensemble perturbation for ensemble forecast since it projects strongly on background instabilities.
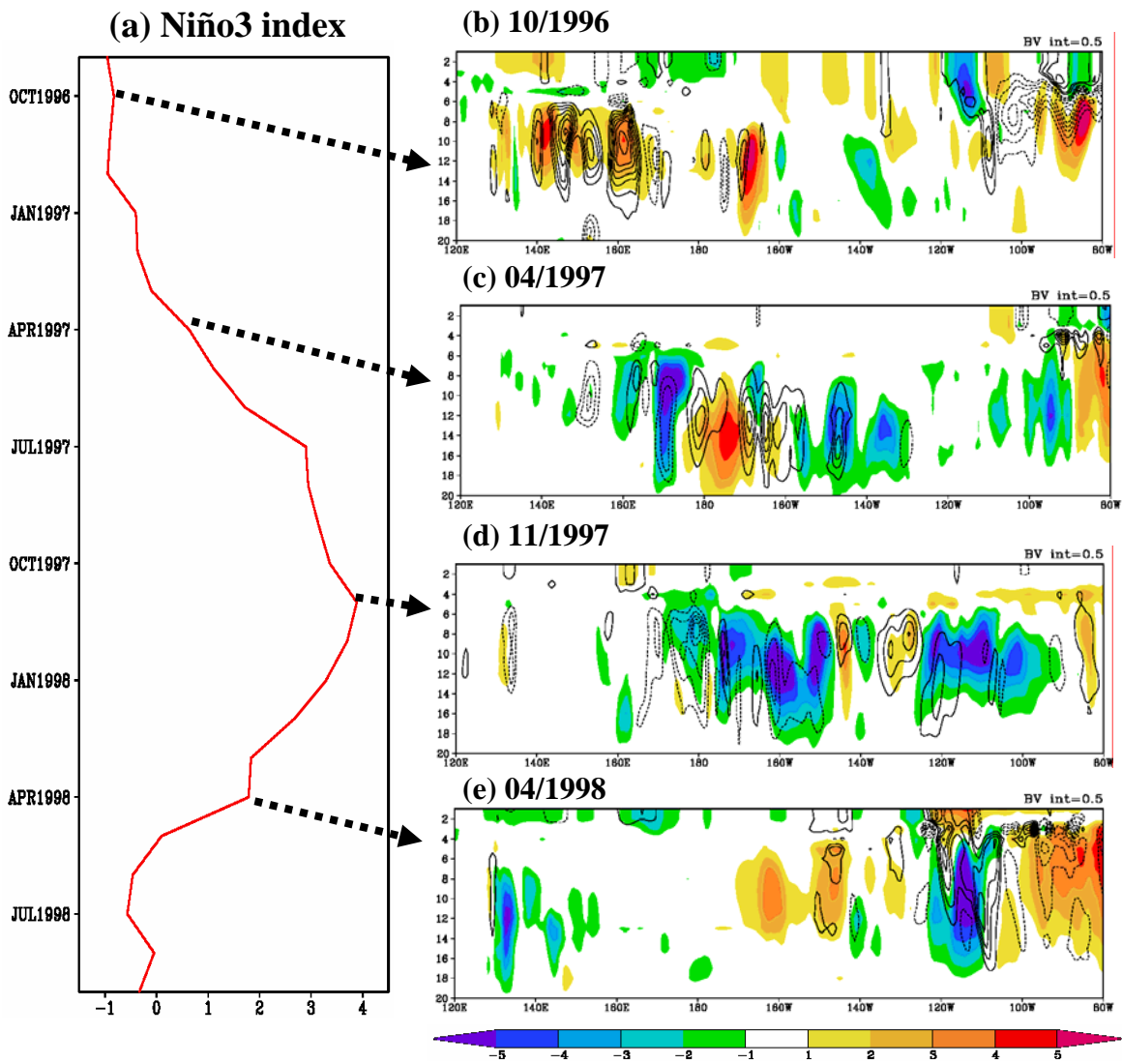
Figure 4.2.1 (a) Background Niño3 index (°C) and vertical cross-section of temperature analysis increment (°C, color) and BV temperature (°C , contour) corresponding to (a) October 1996, before warming developed (b) April 1997, warming started (c) November 1997, warming is strongest, and (d) April 1998, warming diminished. The contour interval is 0.5°C and the zero contour is not plotted.
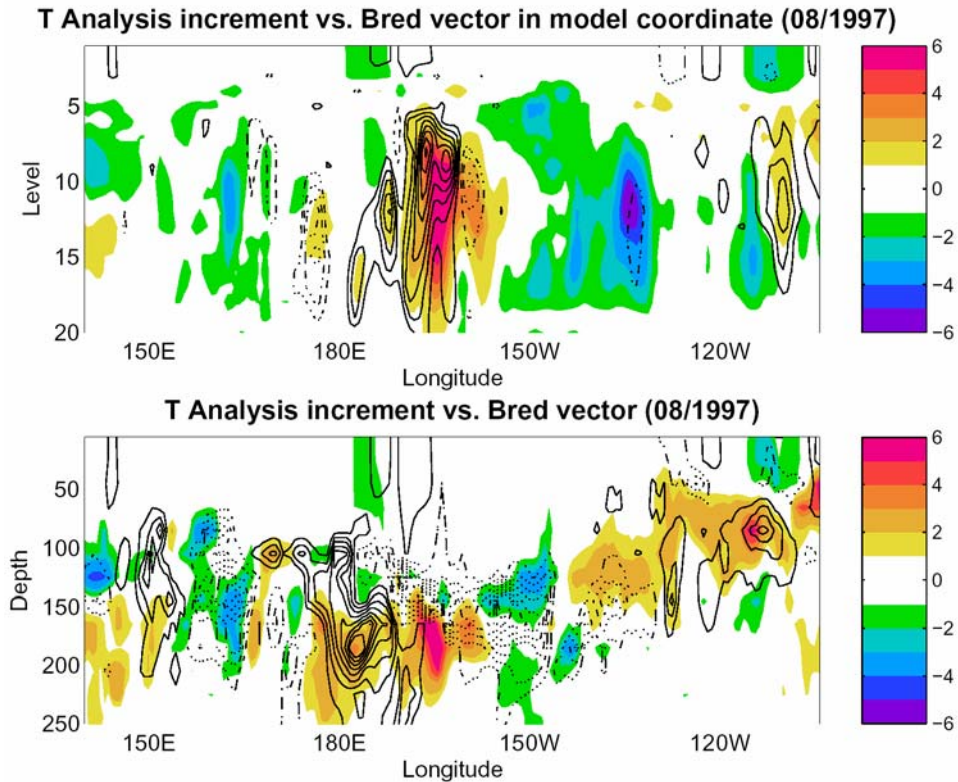
Figure 4.2.2 Vertical cross-section at equator of analysis increment temperature (color) and BV temperature (contour) (°C): (a) in levels (model coordinate) and (b) in depths.

(b) The temporal and spatial relationship between bred vector, analysis increment and background variability

One main conclusion from Chapter 3 was the time-dependent relationship between ENSO variability and the bred vector growth rate. This relationship remains for the growth rate obtained from a system with data assimilation in a perfect model simulation (figure 4.2.3(b)), indicating the growth rates are large before and after ENSO and smallest at the mature state of the ENSO event, described by the Niño3 index in figure 4.2.3(a)). The background Niño3 index is defined using oceanic analysis SST, which is very similar to the observations ("Reynolds SST"). We find

that the growth rate is also strongly correlated to the magnitude of the SST analysis increment (figure 4.2.3(c)), measured by rms of the SST analysis increment in the Niño3 region. It shows that the amplitude of the analysis increment is also very sensitive to the phase of ENSO and related to background variability. The relationship between these two time series suggests that the bred vector has similar error structures as the analysis increment field and that bred vector growth rates can provide information about large forecast errors, particularly when the background SST is at neutral state before and after an ENSO event.

In order to measure the relationship between bred vectors, analysis increment and background state, we used the pattern correlation between the bred vector and the analysis increment in the Niño3 region to determine how well the breeding method could capture the dynamical error. We grouped monthly pattern correlations into bins based on the BV growth rate (bins are categorized by growth rate < 2.5, 2.5–3.5, 3.5–4.5 … and > 8.5 per month). We then calculated the mean of the absolute value of pattern correlation in each group. For comparison, we also divided the Niño3 index into eight groups using the same BV growth rate classification and calculated the mean value of the absolute Niño3 index within each group. As shown in figure 4.2.4, the Niño3 index is high when the growth rate is small and close to zero when the growth rate is large, as discussed above. The mean pattern correlation between BVs and analysis increments in the last two groups with large growth rates are particularly high with respect to other groups. An example is in May 1994, chosen from the last group in figure 4.2.4 and shown in figure 4.2.5. In this figure, the BV growth rate is

10 and the pattern correlation between analysis increment and bred vector is 0.52. As shown in this figure, BV SST (contour) captured most of the dynamically evolving error with large amplitude as shown by SST analysis increment, mainly located in the eastern Pacific.

Combining the results from figure 4.2.3 and figure 4.2.4, we can conclude that bred vector is naturally influenced by the background instability. The shapes of analysis increments (forecast errors) are also influenced by these instabilities. Therefore the bred vector detects the rapidly changing stage of the background occurring in the eastern Pacific and represents the shape of forecast errors. This time-dependent feature suggests the potential application of bred vectors in data assimilation. The information obtained from bred vectors can be used to augment the background error covariance and to avoid underestimating it with "errors of the month" in areas of large error.
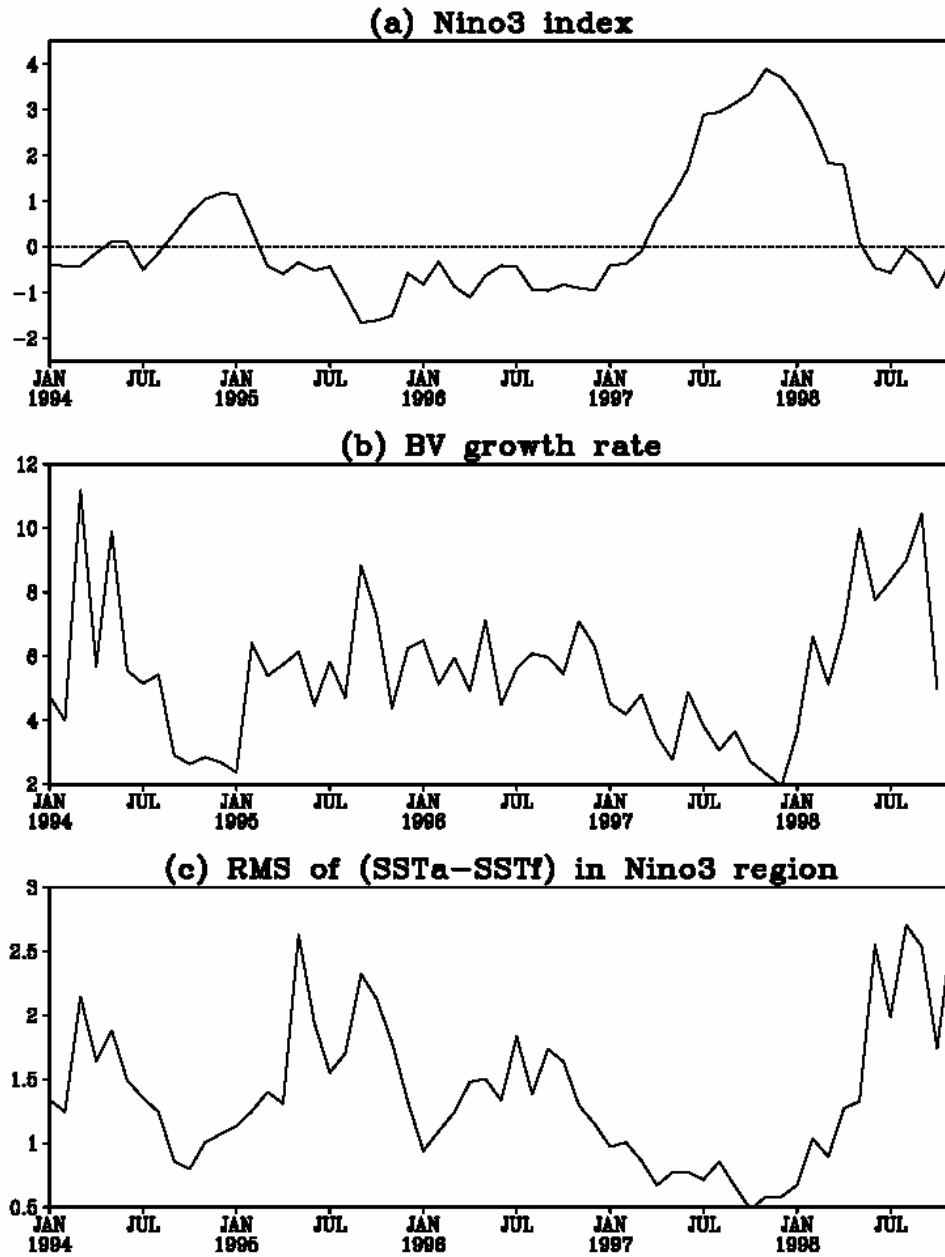
Figure 4.2.3 (a) Background Niño3 index (°C), (b) bred vector growth rate (per month) and (c) root mean square of the difference between SST analysis and one-month forecast in the Niño3 region (°C).
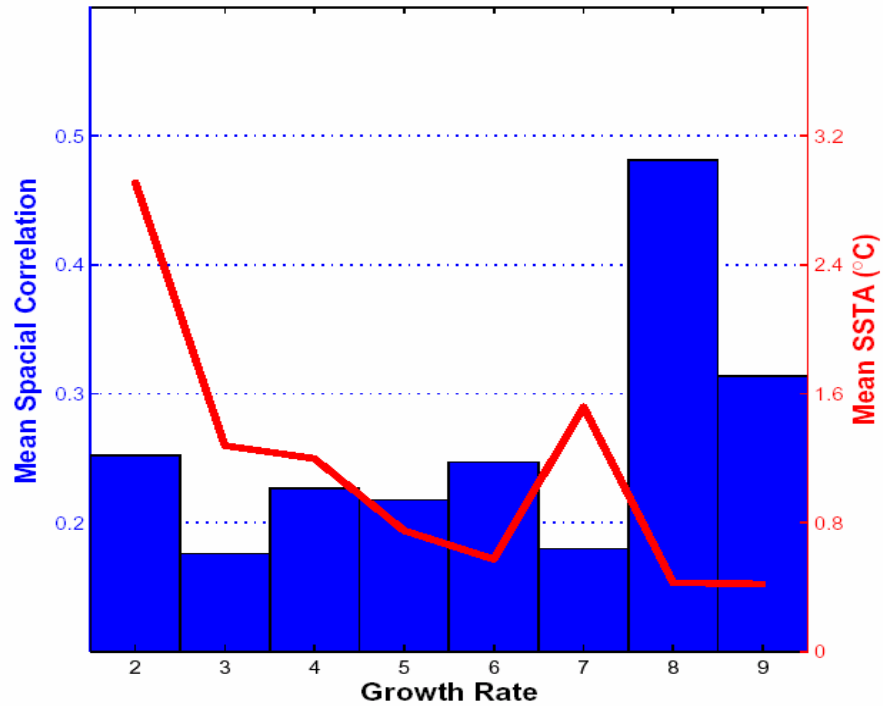
Figure 4.2.4 Mean value of the pattern correlation (bar) and the Niño3 index (red line) in catalogs defined by the BV growth rate. Pattern correlation and the Niño 3 index are grouped based on their corresponding growth rate and bins are the growth rate < 2.5, 2.5-3.5, 3.5-4.5,…. and > 8.5. Pattern correlation is defined as the spatial correlation between the bred vector and analysis increment in the Niño3 region and the absolute value is used for both the pattern correlation and the Niño3 index.
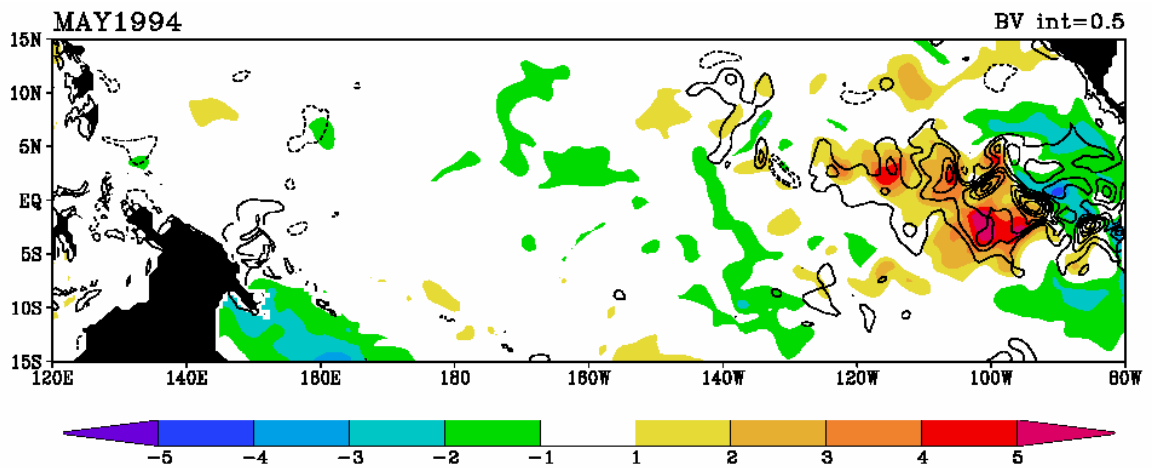
Figure 4.2.5 SST analysis increment (color) and BV SST (contour) (°C) in tropical Pacific in May 1994. The contour interval is 0.5°C, and the zero contour is not plotted.

80

(c) Climate variability and dynamic error structure

In the previous subsections, we have shown that both BV SST and analysis increment are dominated by the dynamic error in the eastern Pacific. The growths of SST perturbations in the western Pacific are limited by its low natural variability. This difference can be understood by the mean structure of the background thermocline, since the depth of the well-mixed surface warm water is much deeper in the western Pacific than the eastern Pacific. Therefore, the growths of any temperature perturbations are less detectable in the western Pacific. Despite this, a small perturbation in the subsurface, such as a small displacement generated by a wind perturbation, will propagate into the east Pacific and influence the growth of SST perturbations in the shallow-thermocline region.

It has been recognized that maximum subsurface temperature variability is found in the mean thermocline. We performed an EOF analysis for the equatorial temperatures showing there are two dominant modes associated with the ENSO evolution, (figure 4.2.6 (a) and (b) where the mean depth of the thermocline is represented by the dashed black line). The first mode has large variability located in the eastern Pacific and explains 44% of the total variability. The principal component (time series) corresponding to this mode is in phase with the SST variation in the eastern Pacific. The second mode exhibits large variability that starts in the western Pacific and peaks in the central Pacific. The principal component shows that it leads the first mode by about 7 months and explains 19% of the total variability. During

ENSO development, the subsurface anomaly first appears near the thermocline in the western Pacific (as the 2nd EOF mode) and propagates eastward along the path of the thermocline (as 1st EOF mode).

It is clear that for ENSO prediction, the ability to describe thermal variations in subsurface conditions will determine the how good the SST forecast can be extended. It has been documented by Rienecker et al. (2001) that simulated subsurface temperature variations from the NSIPP CGCM have only half the magnitude of observations. This will crucially influence the efficiency of forecasting SST anomalies in the equatorial eastern Pacific. Such a deficiency is a common feature of many CGCMs. This implies that there is a need for ensemble perturbations to project on these large-scale variabilities, when undertaking seasonal-to-interannual prediction.

The same EOF analysis is applied to the analysis increment and bred vector of equatorial temperature. Figure 4.2.7(a)-(c) are the first three EOF modes of analysis increment and their explained variances are 19%, 11% and 5%. The first two modes of the analysis increment have similar structures to the dominant ENSO-related modes, showing large variability along the thermocline. The first mode has large variabilities in both the eastern and western Pacific. Instead of a one-sign dominated pattern as indicated in the background temperature anomaly (figure 4.2.6(a),(b)), the leading mode of the analysis increment has the variance amplitude in the west more comparable to the one in the east. The second mode is mainly located in the central

Pacific. Combining these patterns explains that how temperature perturbations grow rapidly when the thermocline slope becomes sharp. It also demonstrates that the analysis increment in the subsurface is dominated by dynamical evolving error. Most importantly, their patterns suggest that the forecast error projects strongly to the large-scale features associated with ENSO variability.

We then applied the same EOF analysis to the bred vector equatorial temperature. Here, we use the unrescaled bred vectors, i.e. the bred vectors are weighted with their growth rates, in order to find the patterns that dominated the growing part. Figure 4.2.8(a)-(c) are the first three EOF modes of the analysis increment and their explained variances are 7%, 8% and 9%. The EOF modes are strikingly similar to the EOF modes of the analysis increment and their roles are relatively the same since their explained variance is very close. This supports our conjecture that bred vectors can capture the shape of the related dynamic error that dominates the analysis increment. Currently, the background error covariance in the univariate Optimal Interpolation analysis uses an isotropic shape in the horizontal and in vertical (Eq. (2.3.5)). By contrast, the patterns of EOF modes suggest that the background error in the subsurface should have considered large-scale shape expanding near the thermocline in order to ensure the corrections from the data assimilation scheme project on the ENSO-related low dimensional space.

We also notice that in the BV EOF modes, there is a feature locally trapped at the very end of the eastern equatorial Pacific and near the surface, which is absent in

the EOF modes of the analysis increment. This may be contributed from the rapid dynamical adjustment of Ekman pumping in the shallow region off the coast when introducing the bred perturbation to the unperturbed control background, i.e. the analysis field.
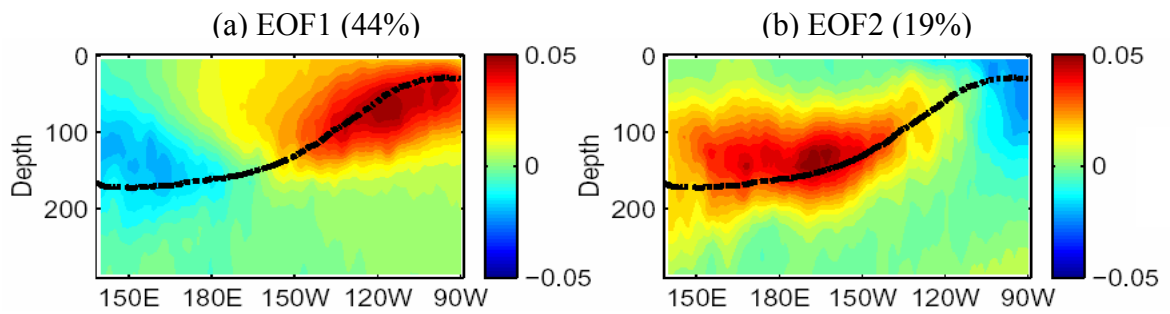
Figure 4.2.6 (a) The first EOF mode of the equatorial temperature anomaly and (b) the second EOF mode. The thick dashed line is the depth of the mean thermocline. EOF modes are normalized.

Figure 4.2.7 (a) The first EOF mode of equatorial temperature of analysis increment, and (b) the second EOF mode and (c) the third EOF mode.

Figure 4.2.8 (a) The third EOF mode of equatorial temperature of the unrescaled bred vector, and (b) the second EOF mode and (c) the first EOF mode.

(d) The relationship between the bred vector growth rate and background seasonality

The fact that eastern Pacific SST anomalies peak at a specific season indicates the interaction between the interannual and seasonal variabilities and the phase-locking of ENSO with the seasonal cycle. Also, the amplitude of SST anomalies in the eastern Pacific (Niño3 region) is smallest during spring and is large during the fall and winter seasons as described by the standard deviation of SST anomalies in figure 4.2.9(b). The eastern Pacific is dominated by a seasonal cycle, indicating a seasonal warming in boreal spring and cooling in fall (figure 4.2.9(a)). Zebiak and Cane (1987) pointed out that summertime is the most favorable period for anomalies to have rapid growth. The background seasonal cycle plays a crucial role in ENSO intensification and in its irregular periodicity due to the nonlinear interactions between these time scales. In fact, the variations of the upper ocean heat content associated with ENSO will also modulate the amplitude of seasonal cycle. Thus, the deepening in the thermocline in the warm phase of ENSO will reduce the seasonal cooling in SST and lead to a phase-locking of ENSO to the seasonal cycle (Xie 1995 and Gu and Philander 1995).

The bred vector growth rate also contains a seasonal cycle component as indicated by the black line in figure 4.2.9(b). Seasonal variation in the growth rate remains large from early summer to fall and is lowest during winter time. In addition, it shows a 90° phase shift (one season) to the background seasonal cycle of SST (figure 4.2.9(a)). This demonstrates that when the background seasonal cycle is at its

inflection point (about to change), it provides the dynamical condition for the perturbations to grow vigorously. This agrees with the results of Zebiak and Cane (1987). Also, in figure 4.2.9(b), the growth rate annual cycle peaks one season earlier than the SST anomalies' seasonal variations.   This BV growth rate, peaking in late spring and summer, is what explains the skill spring barrier (Cai et al. 2003).

The seasonal variation explains 20% of the total variance of the growth rate. We notice that when the background SST anomalies are in a cooling stage, the growth rate is generally large (figure 4.2.3 (a) and (b)). This may be a result of the presence of tropical instabilities which are vigorous due to the sharp temperature gradient when the cold tongue is well established (figure 4.2.11). After we remove the seasonal variations and the mean value from the growth rate, the large growth rate clearly stands out before the 1997 warm event. We then calculate the lead and lag correlation between the growth rate and the background Niño3 index. As shown in figure 4.2.10, the result shows that the growth rate "anomaly" leads the background Niño3 index four months and lags five months. Our result from a complex CGCM confirms the relationship between BV growth rate and ENSO variability in Cai et al. (2003) obtained from a comparatively simple and single mode dominated model.
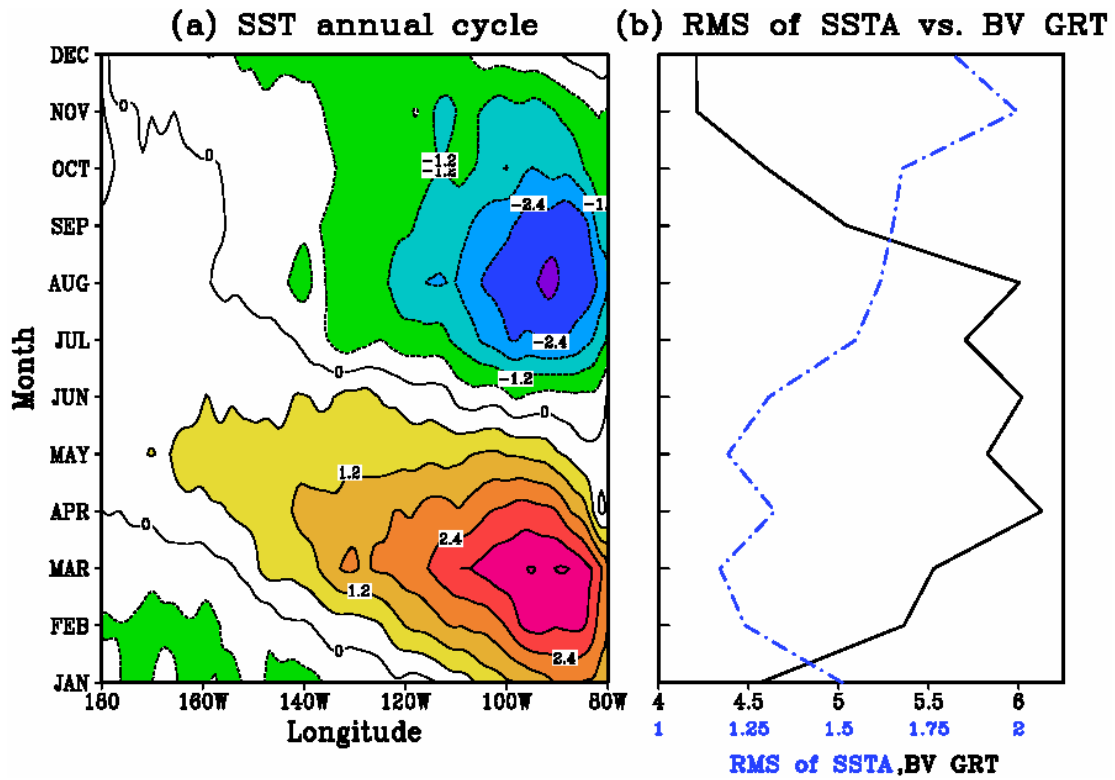
Figure 4.2.9 (a) The annual cycle of SST in the eastern equatorial region (4°S to 0) and (b) the annual cycle of the BV growth rate (the black line) and the rms of background SST anomaly within the Niño3 region.
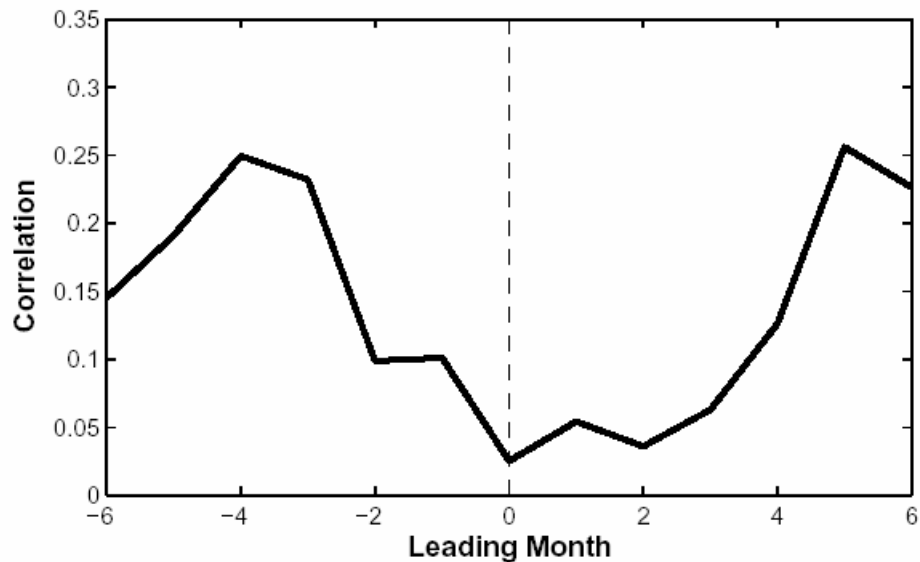
Figure 4.2.10 Lead/lag correlation between the BV growth rate anomaly and the background Niño3 index. An ENSO event peaks when the leading month equals zero, as represented in the dashed line.

4.3 Bred vectors with different rescaling norms

Cai et al. (2003) demonstrated that bred vectors in a coupled system are not sensitive to the choice of either an energy or a L2[*] norm. In order to test the robustness of bred vectors in a more complicated system, we performed different breeding experiments by choosing a different rescaling norm. Hereafter, we will refer to the bred vector obtained from rescaling BV SST in Niño3 region as BVa. Two different norms we chose for comparison are: (1) the tropical thermocline norm (10°N-10°S, 130°E-80°W) with a size of 2 m; and (2) the Niño3 SST norm with a size of 0.15°C but breeding is only done in the tropics (perturbations are damped beyond 30°N/S). They will be referred to as BVb and BVc respectively.

During the breeding experiments of BVa, BVb and BVc, large amplitudes of BV SSTs are often found in the eastern Pacific and sometimes extend along the equator when the background cold tongue is rapidly established. Figure 4.3.1 (a)-(c) are the snapshots for BVa, BVb and BVc SSTs compared with the analysis increment in August 1993. Clearly, in this example, the structures of bred vectors are insensitive to the choice of rescaling norm. Bred vectors from three breeding experiments all tend to locate at those regions characterized by a large amplitude of analysis increment. This also suggests that the background state has the location of low dimension, where one bred vector is able to capture the dominant dynamic instability

---

[*] L2 norm is constructed based on the perturbation vector containing six model variables: surface current in zonal and meridional direction, thermocline, sea surface temperature and windstress in zonal and meridional direction.

in the local region near 120°W, 0°N. Because the bred vector is constructed with finite time and finite amplitude, it is not necessary that it has the same sign in the local region. Therefore, opposite signs are shown in the equatorial eastern Pacific. In addition, there are still some discrepancies in the local structures between bred vectors since this dynamic system contains a variety of nonlinear processes and instabilities.

We also examine the subsurface temperature structure between bred vectors. Comparing figure 4.3.2 (a),(b) with figure 4.2.8(a)-(c), the first three EOF modes in BVa are captured by the first two BVb EOF modes. In addition, BVc EOF modes (figure 4.3.2(c)-(e)) also show high resemblance compared with modes obtained for BVa and BVb. Their EOF structures confirm the robustness of bred vectors, which emphasizes similar variability of analysis increment (figure 4.2.7 (a) and (b)).

As shown in those modes, some large variability is divided into two modes. For example, in figure 4.2.8, the variability located near 30°W appears in both EOF1 and EOF2 modes. Therefore, it will not be appropriate to define a mode-to-mode correlation. For the purpose of quantifying their relationship, we define a total correlation to express how the BV modes correlate to the analysis increment as in Eq. (4.3.1): the total correlation for the ith EOF mode of analysis increment is the square root of the summation of the square of the correlation between this mode and the BV j^{th} EOF mode.
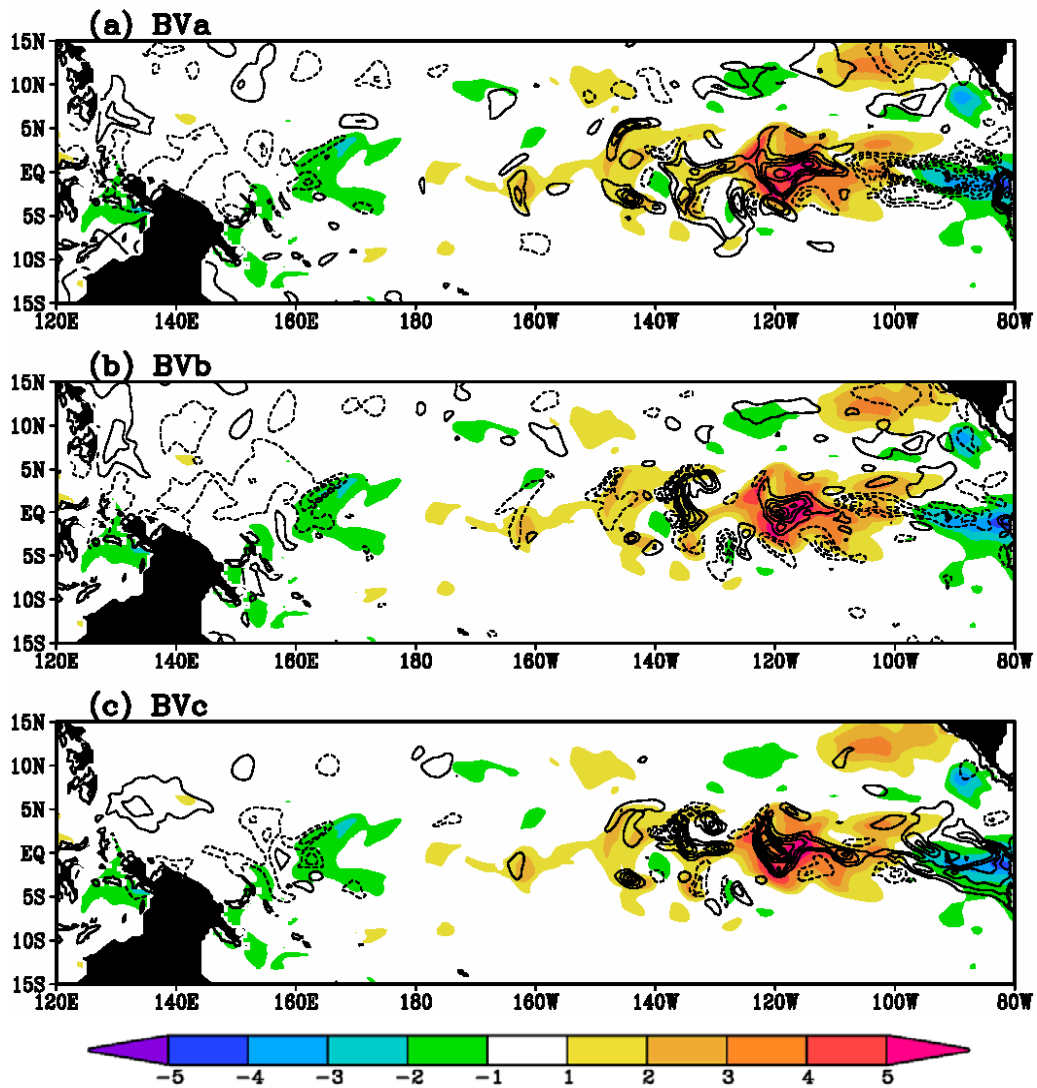
Figure 4.3.1 SST analysis increment (color) and BV SST (contour) in the tropical Pacific in August 1993 (a) BVa using a Niño3 SST norm (b) BVb using a tropical Z20 norm and (c) BVc, the same as (a) but breeding is done in tropics only.

$$\hat{\rho}_i = \sqrt{\sum_{i=1}^{i=3} \rho_{i,j}{}^2} \qquad\qquad (4.3.1)$$

Table 4.3.1 Total correlation between analysis increment (An.Inc) EOF modes and bred vectors (BV)

	An.Inc EOF1	An.Inc EOF2	An.Inc EOF3
BVa	0.80	0.84	0.62
BVb	0.84	0.75	0.49
BVc	0.80	0.64	0.50

In table 4.3.1, we summarize the total correlation between analysis increment and bred vectors. The first two EOF modes of analysis increment are largely related to the bred vectors. This indicates that in the subsurface, the growing part of the bred vector structure will dominate the shape of analysis increment and such growing parts are robust and insensitive to the perturbation norm. This also reflects that nature decides where are the growing instabilities by itself, instead of isolating the instability from choosing a norm (as is the case with initial singular vectors).
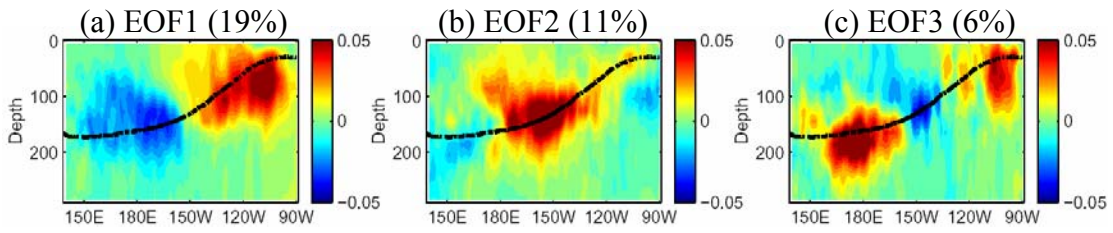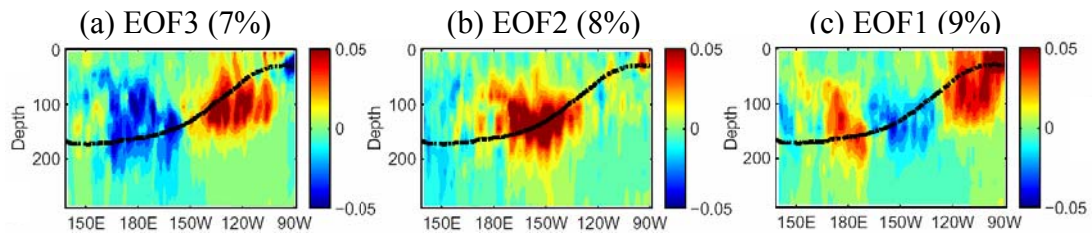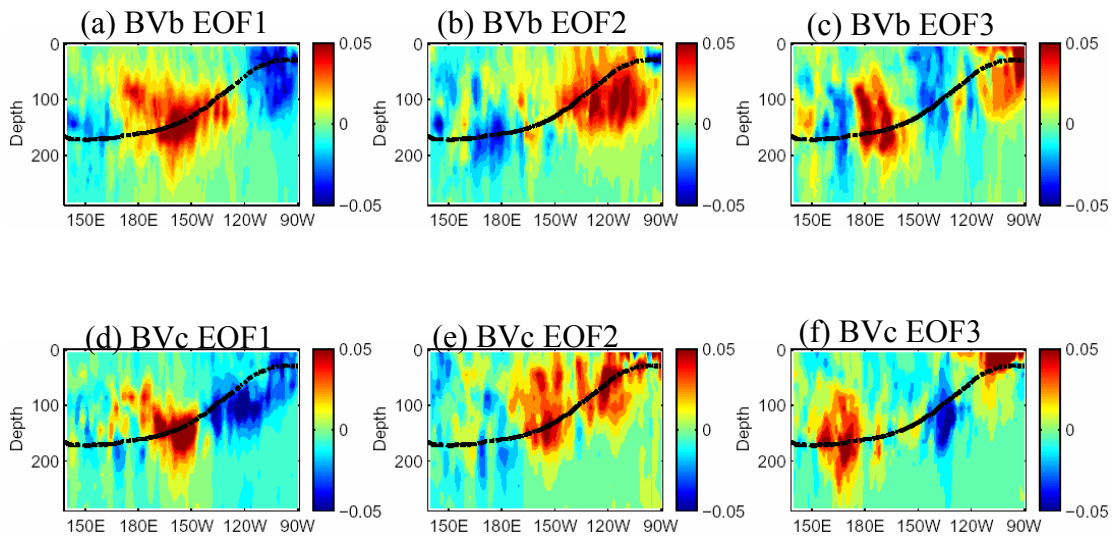
Figure 4.3.2 (a) The first EOF mode of equatorial temperature of the unrescaled BVb, and (b) the second EOF mode and (c) the third EOF mode. (d),(e) and (f) are the same as (a),(b) and (c), except for the unrescaled BVc.

(a) Comparison of bred vectors with operational ensemble perturbations

As we have illustrated in previous sections, the analysis increment in the tropics contains dynamic errors associated with seasonal to interannual instabilities, whose structures can be described by bred vectors. The characteristics of bred vectors obtained with different rescaling norms are robust and have high pattern correlations with each other. However, those bred vectors will not collapse onto one single mode and will allow themselves to project onto other sub-growing directions. We assume that such features should benefit the NSIPP oceanic data assimilation scheme by supplying growing information for the background error covariance. In order to illustrate this point, we will project the analysis increment onto the subspace spanned

by three bred vectors and then compare to the projection on the subspace spanned by three operational ensembles.

For this purpose, we define the local projection of the SST fields of analysis increment on bred vectors and on operational ensemble perturbations. We looped a local patch centered at each grid point with a size of 5×9 grid points (about 3° in both longitude and latitude) and calculate how much of the analysis increment can the bred vector or operational perturbations explain locally. The detailed process to obtain the local projection follows a method proposed by Patil et al. (2001).

First, the analysis increment and the bred vectors are defined on a local patch centered at a grid point (m,n). We then reshape this local patch into a vector with 45 elements. As described in Eq (4.3.2) and Eq. (4.3.3), $\mathbf{y}$ is the local vector corresponding to the analysis increment, $\mathbf{x}_k$ is the local vector corresponding to k^{th} bred vector. K =3 is the total number of bred vectors. To define the subspace spanned by these three bred vectors, we put $\mathbf{x}_k$ into a matrix and transform this into an orthogonal matrix, whose k^{th} column vector is defined as $\hat{\mathbf{x}}_k$. We then remove the projection on these orthogonal column vectors from $\mathbf{y}$ sequentially (Eq. (4.2.4)). The remaining part, $\hat{\mathbf{y}}_{\mathbf{K}}$, stands for that portion of the analysis increment unexplained by the bred vectors.

$$\mathbf{y} = T_a(i,j) - T_f(i,j) \qquad (4.3.2)$$

$$\mathbf{x}_k = bv_k(i,j) \qquad (4.3.3)$$

where $i = m-2, m-1, m, m+1, m+2,$ $j = n-4, n-3,\ldots,n,\ldots,n+3, n+4$ and $k=1,\ldots,K$

$$\hat{\mathbf{y}}_k = \hat{\mathbf{y}}_{k-1} - \langle \hat{\mathbf{y}}_{k-1}, \hat{\mathbf{x}}_k \rangle \hat{\mathbf{x}}_k \qquad (4.3.4)$$

where k goes from 1 to K and $\hat{\mathbf{y}}_0 = \mathbf{y}$

If α represents the angle between bred vector subspace and analysis increment, the explained variance is then defined as $\cos^2 \alpha$, which can be calculated based on $\hat{\mathbf{y}}_K$. The explained variance can also be interpreted as the percentage of the analysis increment that projects on bred vectors:

$$\cos^2 \alpha = 1 - \sin^2 \alpha = 1 - \left[\frac{\|\hat{\mathbf{y}}_K\|}{\|\mathbf{y}\|} \right]^2 \qquad (4.3.5)$$

We made the same calculation for the operational ensemble perturbations simply by replacing bred vectors in the above equations with the three operational SST perturbations, which as described before, are the 3-day differences between ocean analyses.

Figures 4.3.3(a) and (b) show the amount of analysis increment projection on bred vector and operational perturbations for October 1996. In this example, the local analysis increment can be largely explained by the space spanned by three bred vectors. Comparing figure 4.3.3(a) and 4.3.3(b), the main difference is the projection in the tropics is generally lower for the bred vectors than for the operational perturbations. In addition, both bred vectors and operational perturbations can explain

a large amount of the analysis increment outside the tropics. This is because SST variations outside the tropics are slower and smaller. By contrast, the strong coupling between the tropical ocean and the atmosphere and horizontal temperature gradient makes the tropical region contain more complex instabilities as discussed in Chapter 2. Therefore, the dynamic-based bred vector can be expected to capture more instabilities in the tropics than the operational perturbations, even though the latter are based on analysis tendencies.

Based on horizontal projection maps as in figure 4.3.3, we calculate the longitudinal mean of such projection for all the months. The results are shown in figure 4.3.4 (a) and (b). As we have seen in the previous example, the result shows that the projection is higher outside the deep tropics and with bred vectors the projection amount is even higher. Overall, the projection on bred vector is considerably higher than the projection on the operational perturbations. The mean projection of the 15°N-15°S tropical band for two types of perturbation is shown in figure 4.3.5. Our result suggests that the mean projection on bred vectors is quite robust and remains on a level of 75%. Although, occasionally, the operational perturbation has the comparable projection amount, it is more random in time. Since the operational perturbations are generated by randomly choosing three-day differences between two analysis states, it will have a higher chance to capture the dynamical error when the background varies slowly under a stable condition.

We noticed that, when the growth rate is large, the projection on bred vectors tends to be higher than the projection on operational perturbations. There exists a particular low projection band with operational perturbation before the onset of the 1997-1998 El Niño event (background state changes rapidly). This suggests we can expect bred vector, as an effective ensemble member, to excel the operational ensemble perturbations, particularly when a large growth rate occurs. The example in figure 4.3.4 is such a case of large growth rate that the tropical-averaged projection is 74% for bred vector and only 58% for operational perturbations.

In summary, the impact of using bred vectors rather than operational perturbations can be identified when the bred vector growth rate and analysis increment are both large. This also implies their important role in detecting background variations. The bred vectors can also provide a solution to compensate for the deficiency from the background error covariance ($\mathbf{B}_{OI}$) being time-independent in the operational univariate OI. It is possible to augment the background error covariance at a very low cost by adding a weighted outer product of bred vectors (Eq. (4.3.6)).

$$\mathbf{B}_{augm} = \mathbf{B}_{OI} + \alpha bb^T \qquad (4.3.6)$$
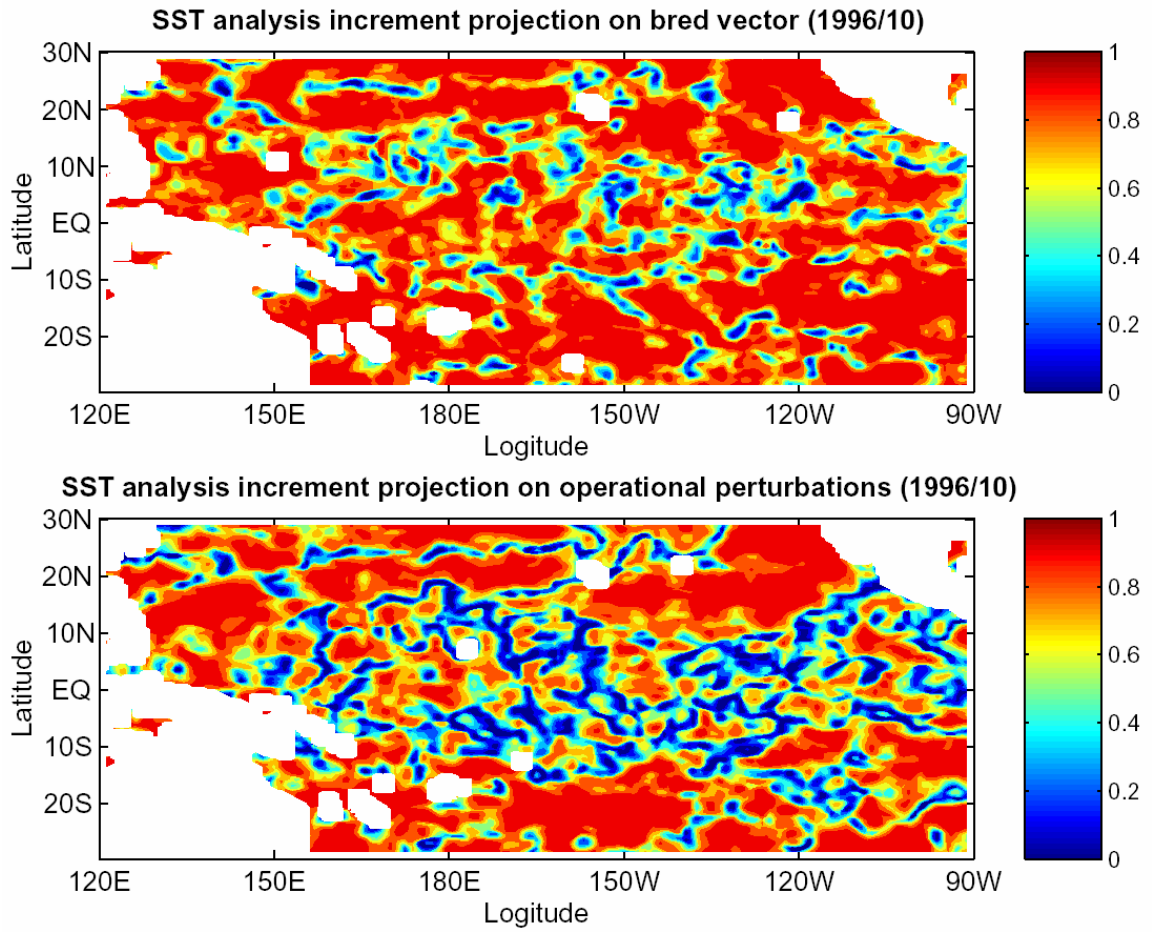
Figure 4.3.3 SST analysis increment local projection on (a) three bred vectors and (b) three operational perturbations
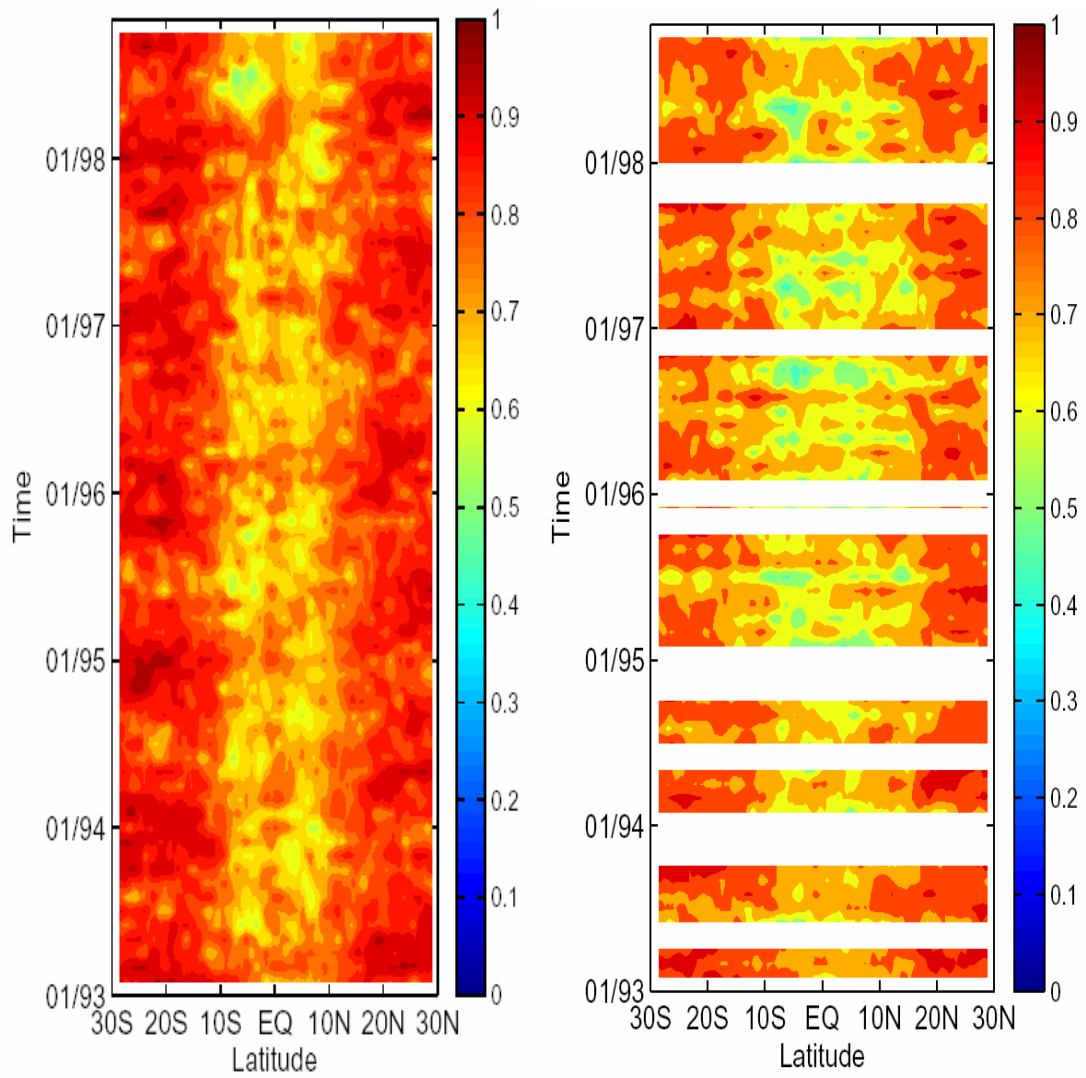
Figure 4.3.4 The latitude-time plot for the longitudal-mean local projection of SST analysis increment on (a) three bred vectors and (b) three operational perturbations
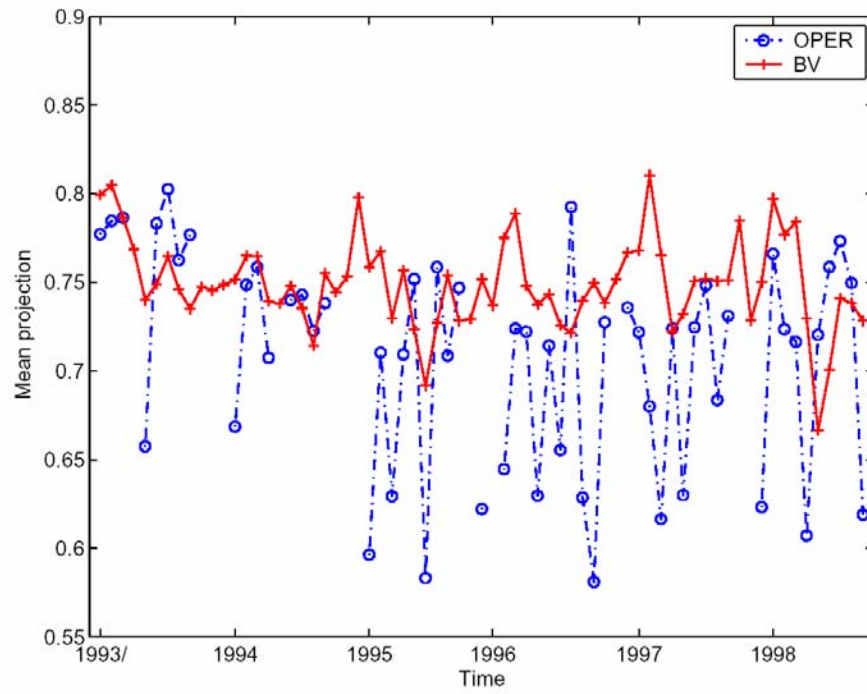
Figure 4.3.5 The mean local projection within the tropical region (15°N-15°S) for analysis increment projections on bred vectors (red) and operational perturbations (blue)

4.4 Breeding experiments for ensemble forecasting

In this section, we describe the application of a two-sided breeding cycle to create dynamical perturbations for initial perturbations. The two-sided breeding cycle was proposed by Toth and Kalnay (1997) for the application in an ensemble forecast system. The two-sided breeding makes use of ensemble forecasts and may result in a nonlinear filtering of forecast error. The schematic plot in figure 4.2.1 explains how to implement this method in an operational CGCM. The two-sided breeding cycle starts by adding and subtracting a set of random perturbations to the corresponding initial fields of the control forecast, which are oceanic analysis and AMIP atmospheric initial fields. In the next step, we integrate both positively and negatively perturbed runs for one month, take the difference between their one-month forecasts, and divide by two. We then rescale this difference to the initial size of the perturbation and add/subtract the bred perturbation to the next initial coupled fields. In this scheme, the pair of bred vectors (positive and negative) is self-propagated without the need for a control forecast. We use the rescaled positive and negative bred perturbations as dynamic perturbations for ensemble forecasts. As discussed in the previous section, bred vectors share many temporal and spatial features with the analysis increment, providing the shapes of dynamic errors but without knowing their signs. Therefore, either the positive or the negative bred vector used for the ensemble perturbation is expected to reduce errors in the initial condition and make it closer to the true state.

The growth rate obtained from the two-sided breeding cycle has a behavior similar to the one obtained from a one-sided breeding cycle as described in the previous section, except its magnitude is smaller (Figure 4.4.2). This reduction could be due to the reduction of spinup by averaging the two atmospheric runs, rather than replacing the atmosphere with a random AMIP state. The spatial structure of the bred vector is similar in both methods. Here, we compare our results with ensemble forecasts carried out by the NSIPP operational center. The generation of ensemble perturbations has been provided in section 2.3. In this section, we will explore the potential improvements we can gain for control forecasts from using dynamic perturbations (bred vector) and identify the benefits of using the coupled bred vectors as the initial perturbations.

Starting from September 1996, we performed 12 month ensemble forecasts every three months. Figure 4.4.3 shows the forecasted Niño3 index from ensemble forecasts using one pair of bred vectors (+/–BV, as shown in the green lines) and operational ensemble perturbations (as shown in the blue lines). The model configuration was updated in 2004 with respect to the 1999 version that was used to generate the operational ensemble forecast. Since in our experiments we use an updated model for comparison, for our ensemble experiments we need to generate new control forecasts with the updated CGCM (shown as the solid red line in figure 4.4.3). The control forecast corresponding to the earlier CGCM is marked as the dashed red line in figure 4.4.3. New control forecasts do not always improve the old control forecast since the modifications are designed to improve forecasts on the average. For example, the new

102

2004 control forecast starting from March 1997 generally improves the warming strength of the forecasted SST anomaly, but not for the case starting from December 1996.

Since the BV and the operational ensemble forecasts are generated with a different version of CGCM, it is difficult for us to make a direct comparison. Despite this, we can still gain some valuable insights based on the results from these 8 cases. Ensemble forecasts with dynamic perturbations (one pair of bred vectors) seem to have better forecast skill when forecast initialization is prior to the beginning of the event and when the control forecast skill is low, which also tend to be cases with large bred vector growth rate. The skill of the ensemble of bred vectors is comparable with that of the operational ensemble perturbations when the starting month is near the mature stage of the event, i.e., when the initial background field has had a strong signal about anomalous variation in the tropical Pacific and the uncertainty is "saturated". Recall from section 2.3, that the operational ensemble perturbations contain the observed memory since they are constructed in a form which reflects the tendency of the analysis. When the forecasts start at these months (June 1997, September and December 1998), the ensemble spread is limited. In fact, the BV growth rate has indicated such limitation on the ensemble spread. In figure 4.4.2, the BV growth rate is small at the mature time of an ENSO event; SST perturbations are smallest in the eastern Pacific and thus will not grow much.

By contrast, the ensemble spread is larger before and after the event and is also large for the cases starting in March which are the cases with large growth rates. Such large growth rate also suggests that the ensemble forecasts have higher chances to include the background uncertainties and improve the control run.
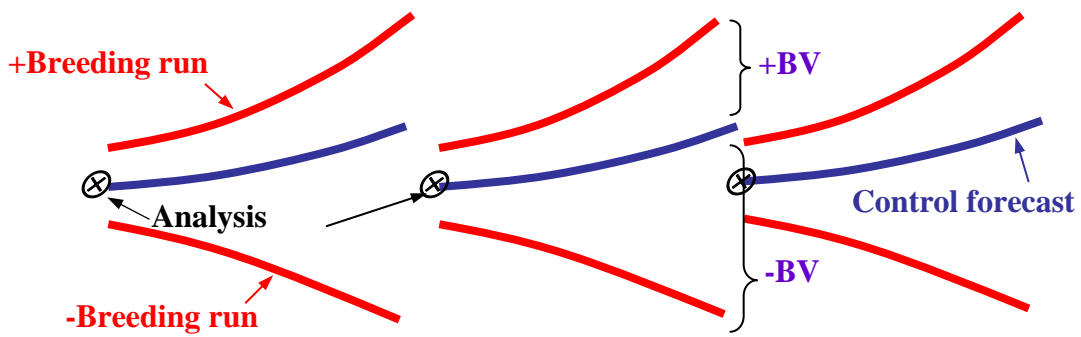
Figure 4.4.1 Schematic plot for performing the two-sided breeding cycle
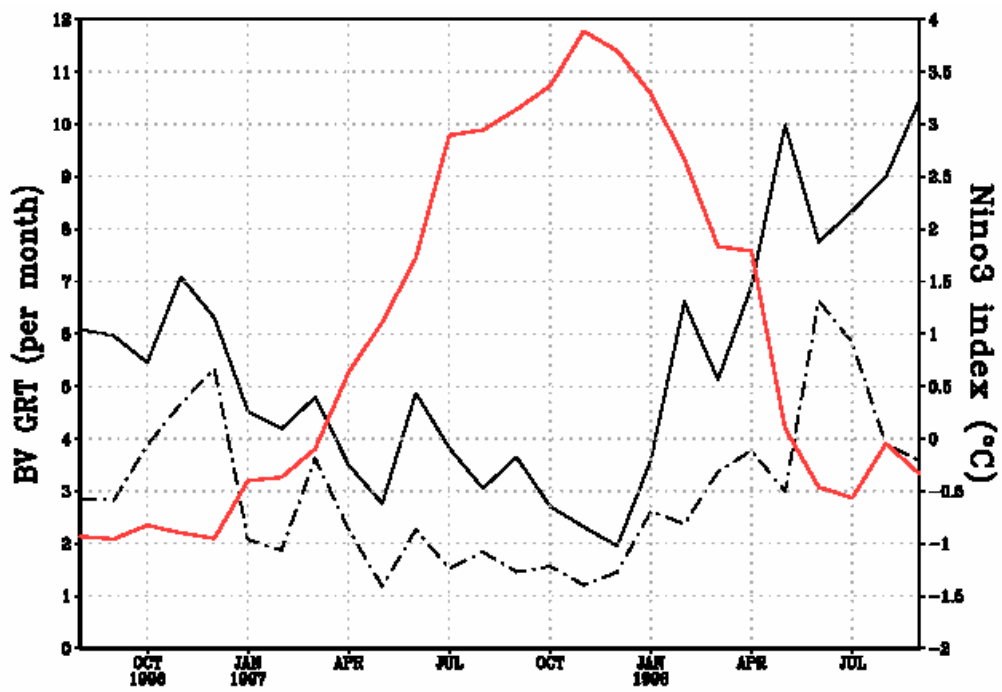
Figure 4.4.2 Growth obtained with one and two-sided breeding experiments (solid: one-sided bred vector; dash-dot: two-sided bred vector). The solid red line is the background Niño3 index and their value is indicated in the right y axis.
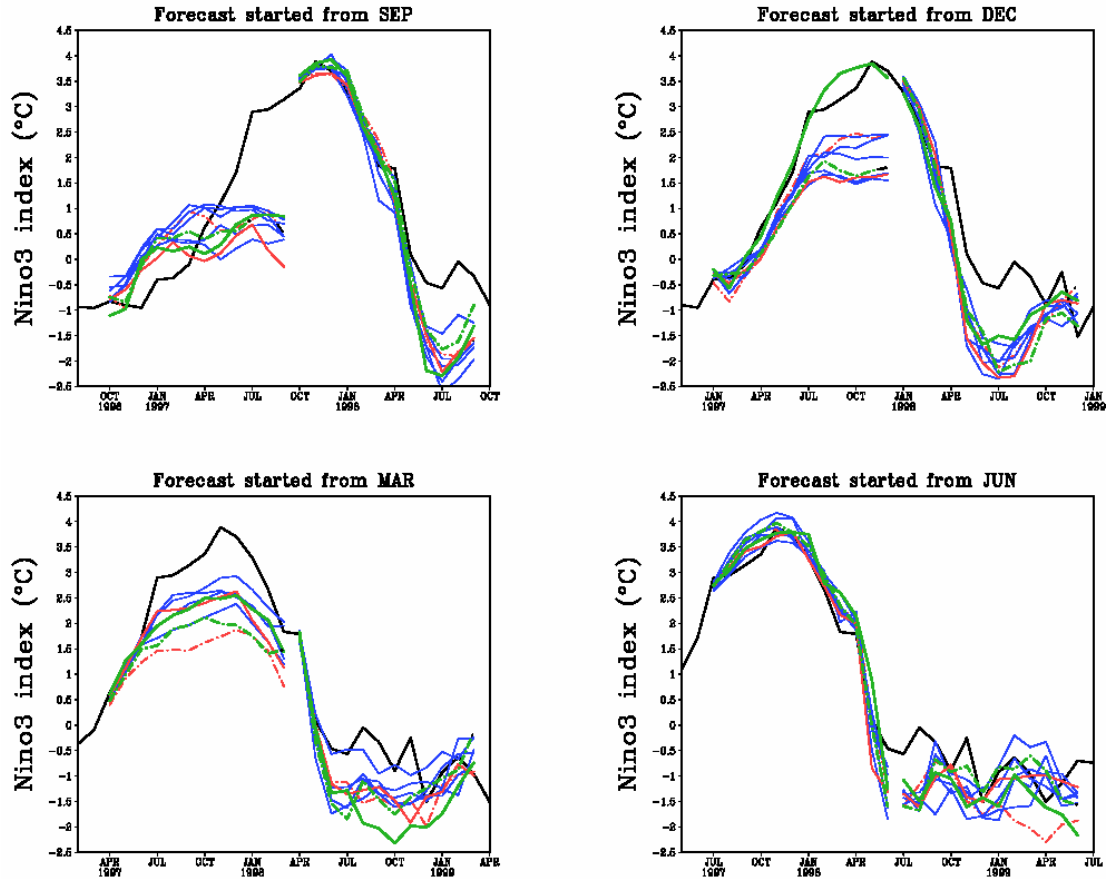
Figure 4.4.3 Forecasted Niño3 index, for forecasts started (a) in September, (b) in December (c) in March and (d) in June. The Niño3 index has been corrected with the climatology in regard of each starting month. The solid/dashed green lines are the ensemble forecasts with initial perturbations of +BV/-BV. The blue lines denote the operational ensemble forecasts (5 ensemble members). The BV solutions are generated by the updated model, whose control forecast is represented by the solid red line. The operational ensemble forecasts are generated by early version of the model, whose control forecast is represented by the red dash line.

106

(a) The role of coupled bred vector in predicting '97-'98 El Niño event

The large improvement of the ensemble initiated on December 1996 and in particular the positive bred vector forecast, allows us to explore the role of the coupled bred vector in predicting the warming strength of 1997-1998 episode. The forecasted Niño3 indices for this case with the +/–BV and operational ensemble perturbations are shown in figure 4.4.3(b). None of the operational ensemble members (the blue lines) and the control forecasts (the red lines) captured the warming strength after July 1997. The coupled +BV solution not only has the forecasted SST anomaly intensifying correctly but also includes a realistic decaying of the anomaly after November 1997. The successful result of this +BV solution is attributed to the good representation of the coupling process between the atmosphere and the ocean.

The Hovmöller diagram of thermocline anomaly in the analysis along the equator shows a clear delayed oscillator behavior (figure 4.4.4(a)): the deepening thermocline anomaly is initiated in the western Pacific and propagates eastward with amplifying amplitude. This deepening of the thermocline in the east Pacific together with the shoaling in the west Pacific reduced the longitudinal slope of the thermocline, and is a consequence of positive feedback of air-sea interaction (Schopf and Suarez 1987). Figure 4.4.4 show that all the ensemble members, including the control forecasts, are able to initiate the deepening eastward-propagating feature forecast. However, comparing with figure 4.4.4(a), only the thermocline anomaly from using the coupled +BV forecast (figure 4.4.4(d)) amplifies strongly to allow the deepening

of the thermocline to reach the eastern Pacific and continue amplifying in the eastern Pacific. The thermocline anomalies from the control, operational ensemble mean, and coupled –BV are much weaker and lack the deepening feature near the eastern Pacific. In addition, only the coupled +BV solution produces a distinctive shoaling in the western Pacific, which is crucial to maintaining the intensity of an El Niño event. The rest of the ensemble members are unable to predict the shoaling feature correctly, indicating the zonal wind stress did not generate enough vertical displacement. Such results reveal that the atmosphere and ocean in those ensemble members do not couple well enough to lock the feedback process in the tropical Pacific.

As we have discussed in Chapter 2, the operational ensemble perturbations for the ocean and for the atmosphere are generated separately and thus, the perturbations are not coupled in the initial state. To address this concern, we separate the coupled +BV into the oceanic perturbation and atmospheric perturbation in order to test whether the high prediction skill can be achieved by simply applying each of them individually. Figure 4.4.5 is the Hovmöller plot for forecasted zonal wind stress (total value field). We took the 15-day average zonal wind stress between 2°N and 2°S for all the ensembles. As shown in figure 4.4.5(b), the forecasted zonal wind stress from the +BV has a very different pattern from all the other ensemble members. Its relaxation is strongest starting at 135°W in April 1997, showing that the +BV solution has the stronger feedback from the coupling process. Such a successful simulation cannot be achieved if we use only the oceanic or atmospheric bred vector from the coupled +BV solution. Additionally, the strength and timing of wind stress

relaxation cannot be reproduced with either the oceanic or atmospheric component alone (figure 4.4.5(d) and (e)). This suggests that the bred vector indeed includes the information related to the coupled instability. Thus, using only the perturbation from either model component will suppress and even destroy such information. We also show the forecasted wind stress field from the operational ensemble and control forecasts for comparison. Figure 4.4.5 suggests that the relaxation of zonal wind stress in April 1997, which is most evident in the +BV solution, is the key to determine the strength of the warming in this case.
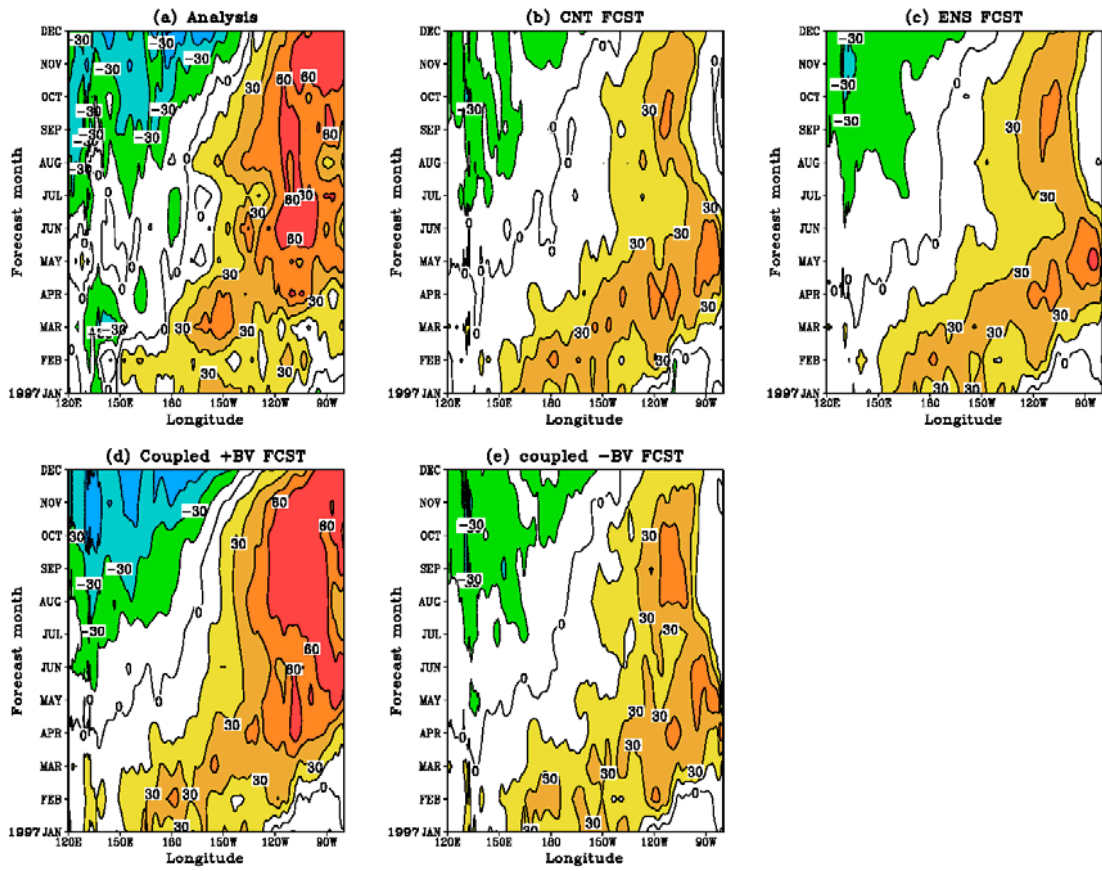
Figure 4.4.4 Thermocline anomaly at the equator starts in January 1997: (a) analysis (b) control forecast (c) ensemble mean of five members (d) coupled +BV forecast and (e) coupled –BV forecast (climatology and drift of OGCM were removed from forecast thermocline anomaly).
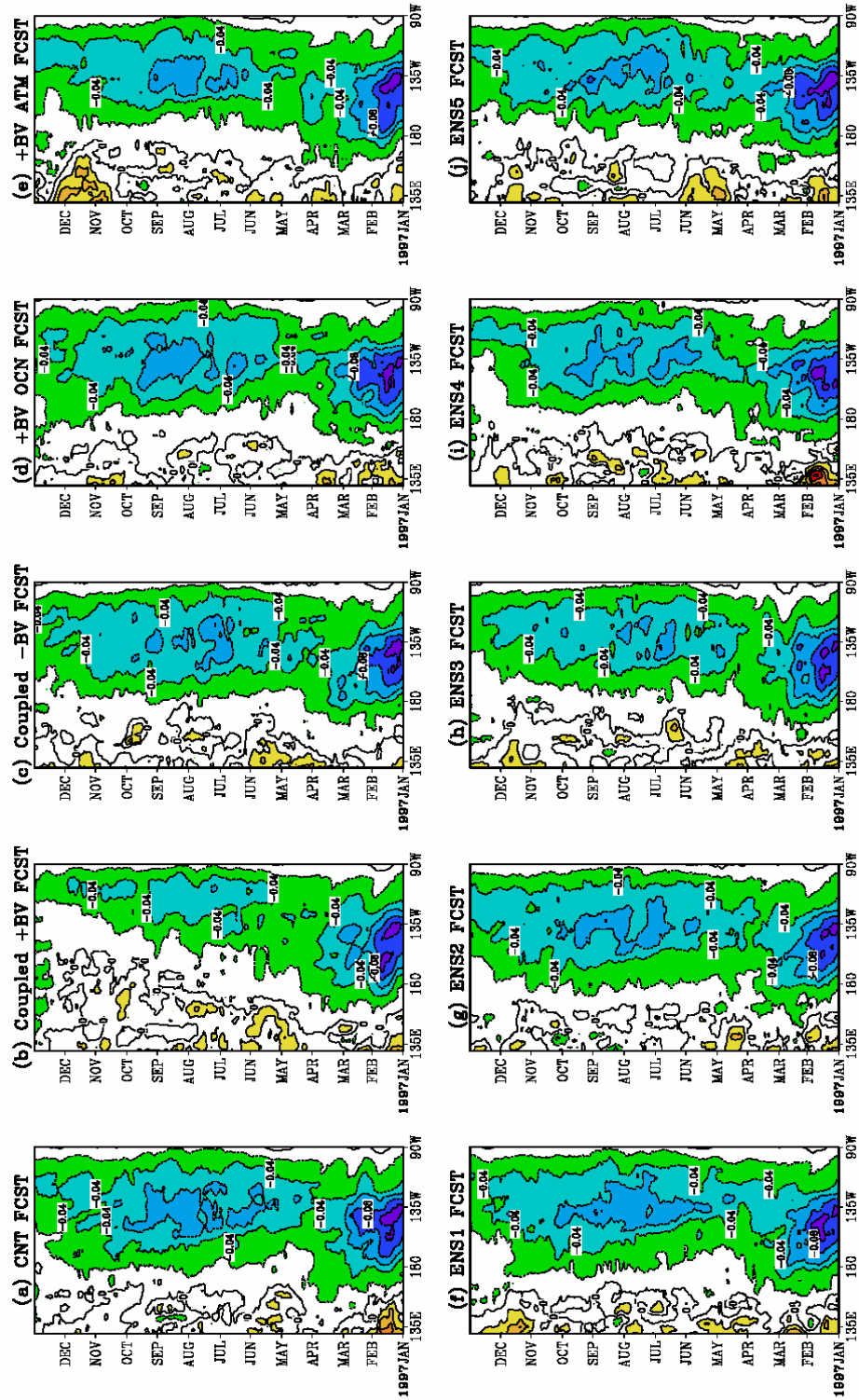
Figure 4.4.5 Forecasted wind stress (total value) started in December 1997 (a) from control forecast and from using initial condition (b) using coupled +BV (c) coupled –BV (d) oceanic component of coupled +BV (e) atmospheric component of coupled +BV and from five operation initial ensemble as shown in (f),(g),(h),(i) and (j)

111

4.5 Two-sided breeding experiments from a modified breeding cycle

There are still some concerns on the coupled bred vectors obtained for NSIPP operational system described in the previous subsection.

- In every breeding cycle, the atmospheric perturbation is added to the AMIP-type initial condition. However, the AMIP initial conditions are generated separately, which forces the NSIPP AGCM to start from a new state instead of from one balanced with the oceanic state. This may make the bred vector evolve upon a background flow which is not well coupled at each new forecast initialization time.

- The AMIP states are the atmospheric response to the observational SST. We know that variations of the wind stress field in the central Pacific particular,ly play an important role in generating equatorial waves, which may trigger or terminate an ENSO event. However, such features cannot be considered in the AMIP initial condition.

- We notice that the magnitudes of the operational ensemble perturbation size within Niño3 region are much larger than the BV perturbation size before and after the 1997-1998 warm event, which are also the cases with large BV growth rates. This suggests that the magnitude of perturbation applied to the ensemble initial condition should depend on the growth rate: use a larger perturbation to encompass the larger dynamic errors.

Therefore, we modified the breeding method and performed the ensemble forecast experiments to address these concerns individually.

(a) The atmospheric initial condition

In order to examine the role of the atmospheric initial condition in the breeding cycle, we replaced the AMIP atmosphere initial condition with two different types of atmospheric fields: one is the NCEP reanalysis data and the other is the mean of the atmospheric fields from the pair of BV one-month forecasts. The former was chosen as a more realistic atmospheric initial condition despite not being well balanced with the initial SST condition. The latter is used as an ensemble mean state "trained" by its own AGCM for the purpose of reducing the initial shock from using uncoupled oceanic-atmospheric initial conditions.

The impact of the atmospheric initial condition is first examined by the difference in the control forecasts (unperturbed ones) by replacing the AMIP atmospheric initial condition with these two different types of atmospheric initial conditions. Their 12-month forecasted Niño3 indices are shown in figure 4.5.1 (the green lines are initialized with AMIP data, the red lines are initialized with NCEP reanalysis and the blue lines are initialized with the atmospheric mean state from the +BV/−BV forecasts). In general, the forecasted Niño3 index retains the similar tendencies for predicting warm or cold anomalies with no clear winner except for the forecasts starting from September 1996. This is because of the dominant role of initial SST and upper ocean thermal variations in determining the phase of an ENSO cycle.

113

Among these 8 cases in figure 4.5.1, there are two cases showing a better prediction skill after we change the atmospheric initial condition. First, the forecast starting from September 1996 with NCEP reanalysis exhibits a more realistic SST anomaly variation than the one starting with either the AMIP or BV mean state. It not only lessened the unrealistic warming at the early forecasted months but also captured the warm anomaly correctly after April 1997 while forecasts with two other atmospheric initial conditions show a rather neutral condition (figure 4.5.1(a)). Such results indicate that providing a more realistic atmospheric condition does help to capture the onset of the 1997 El Niño event. Second, the forecast starting from December 1996 (figure 4.5.2(b)) with the mean of the BV atmospheric fields has a large improvement after six forecasted months compared to the forecast initialized with the AMIP run. This suggests that the mean of the BV forecasted atmospheric fields contains the information underlying the air-sea interaction. If we recall from figure 4.4.3(b), such a benefit from BV fields was apparently recognized in the +BV forecasts when initializing with the AMIP field since its initial perturbations are obtained from the previous one-month positively and negatively perturbed forecasts.

However, we also found that the prediction skills of the forecasts using the NCEP reanalysis data quickly drop when initializing in March. As shown by the red lines in figure 4.5.1(c), the forecast Niño3 index initializing with the NCEP reanalysis are much cooler than the other two forecasted Niño3 indices. We know that the longitudinal SST gradient is smallest during this month since the background SST is warmest in the eastern Pacific (shown in the seasonal cycle in figure 4.2.9(a)). The

oceanic state is sensitive to the atmospheric initial condition most during this season. Therefore, the NCEP reanalysis data may create a stronger initial shock if it is less well balanced to the oceanic state in that month. We also found that the differences between the geopotential height fields of AMIP and NCEP atmospheric fields are strongest at this month.
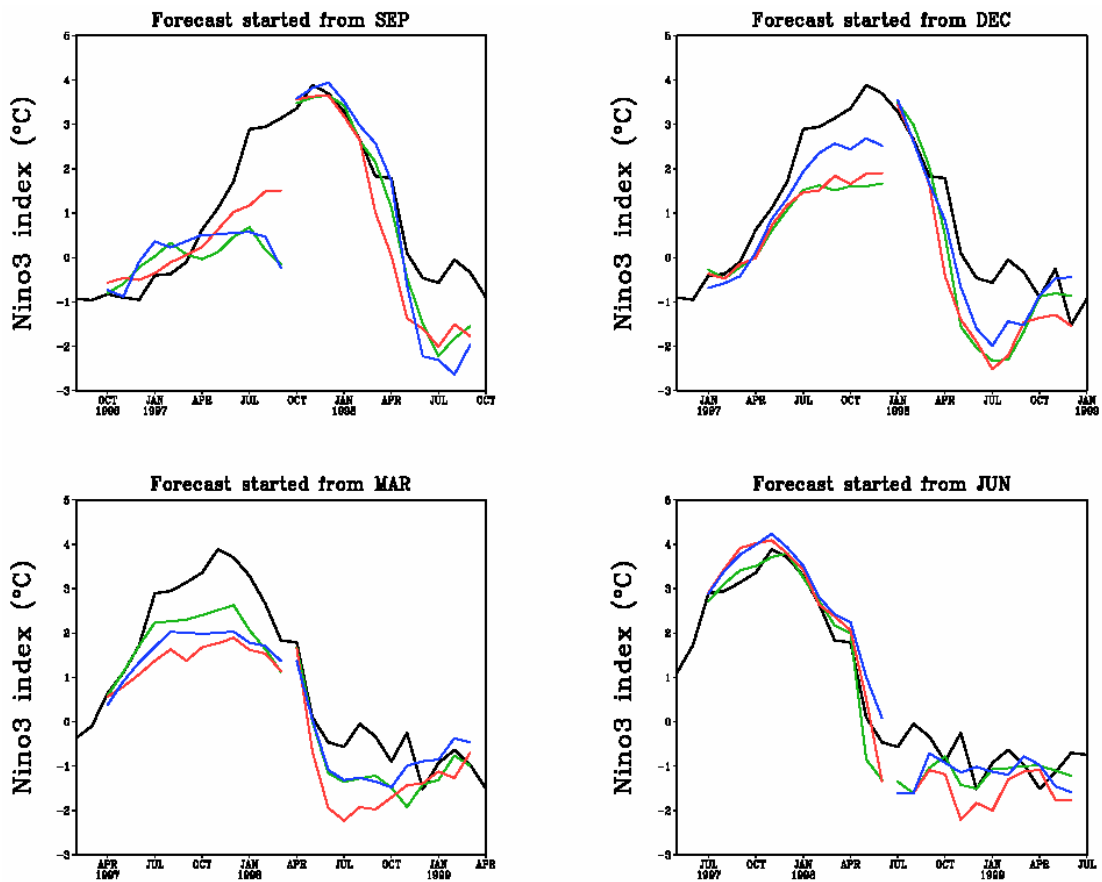
Figure 4.5.1 the same as figure 4.4.3, except they are control forecasts without perturbations and their initial ocean states are all from ocean analysis. The green line is using the AMIP data, the red line using the NCEP reanalysis and the blue line using the mean of +BV and –BV one-month forecasted atmospheric fields as the initial atmospheric field.

Based on the control forecasts (unperturbed oceanic and atmospheric fields) in figure 4.5.1, we then performed two-sided breeding experiments and the results are shown in pairs in figure 4.5.2 (initializing with NCEP data) and figure 4.5.4 (initializing with the mean of BV atmospheric fields), both comparing with the BV ensemble initializing with AMIP data. Mainly, the differences between those pairs of BV solutions occur at the early stage of the 1997 El Niño event. From figure 4.5.2(a), when the forecasts start from September 1996, the –BV solution of the NCEP reanalysis is benefited from the improved control forecast and shows better prediction skill in the later months. Also, the results seem to suggest difference betweens +/–BV solutions tend to be large when predicting the cool anomalies. For example, the cases starting from September 1997 exhibit a large difference when the +BV forecasted a warm anomaly but –BV forecasted a cold anomaly in April 1998. Figure 4.5.3 shows the rms error of forecasted SST when compared with the analysis data. It is clear that the +BV solution reduces a large amount of error from spring to early summer, improving the control run for using unperturbed AMIP or unperturbed NCEP data.

From figure 4.5.4, we do not find significant impact from the two-sided breeding experiment from using the mean of the BV atmospheric fields, expect that the –BV solution shows some improvement for capturing a more realistic warming trend after April 1997. Also, the span between +BV and –BV forecasts seems to be less than the ones using the AMIP initial conditions.

We also note that the forecasted rapid cooling in both control and ensemble

forecasts during April-June 1998 cannot be improved by replacing the atmospheric initial condition. Such feature is related to the model climatology, which shows a much weaker annual cycle, especially during the spring season (March-May). Thus, in order to correctly predict ENSO events, it is important for the model to have a more realistic annual cycle within the background state.
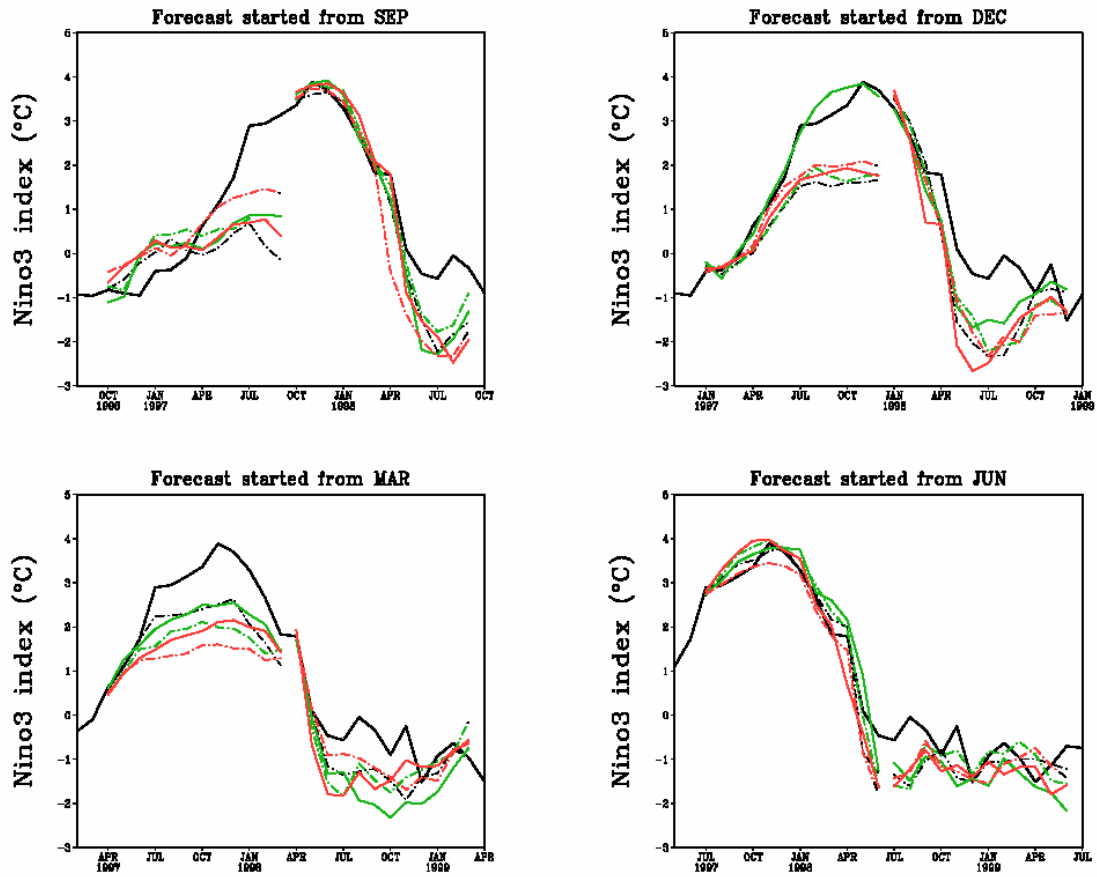
Figure 4.5.2 the same as 4.4.3, the black dashed line is the control forecast, the green lines are the BV ensembles initialized with AMIP data and oceanic analysis, the red lines are the BV ensembles initialized with NCEP (initial oceanic fields remain the same).
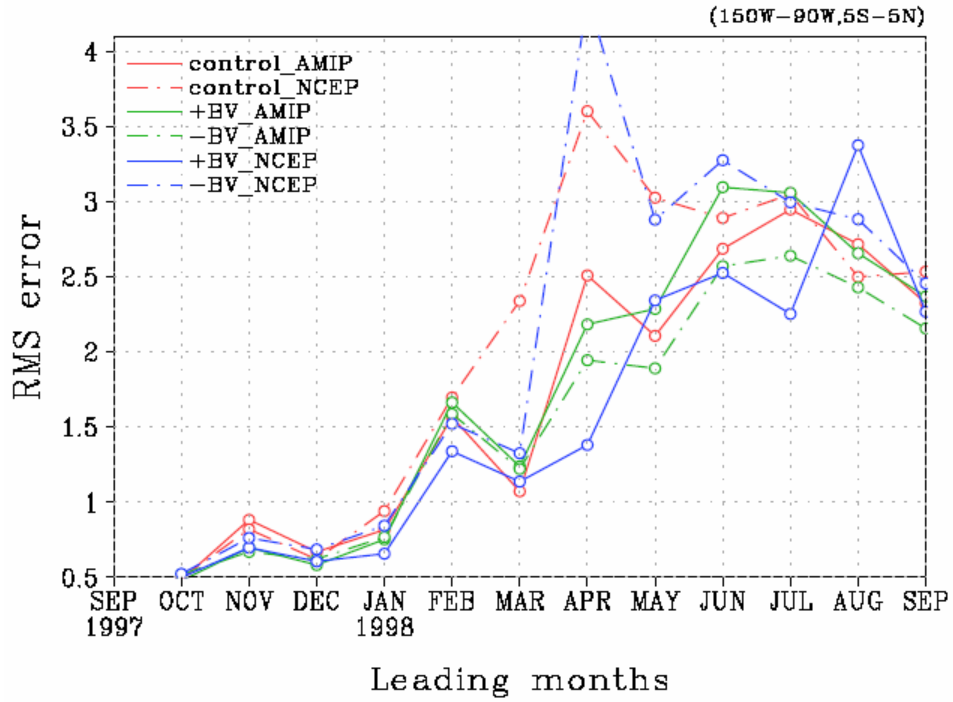
Figure 4.5.3 RMS error between the 12-month forecasted SST and the analysis field. The forecast started from September 1997, the rms error of the control forecast is denoted as the red line, solid: AMIP and dash-dot: NCEP. The green lines are the forecasts initialized with AMIP data and the blue lines are the forecasts initialized with NCEP data and both are using the same ocean analysis (the solid lines: +BV and the dash-dot lines: –BV).
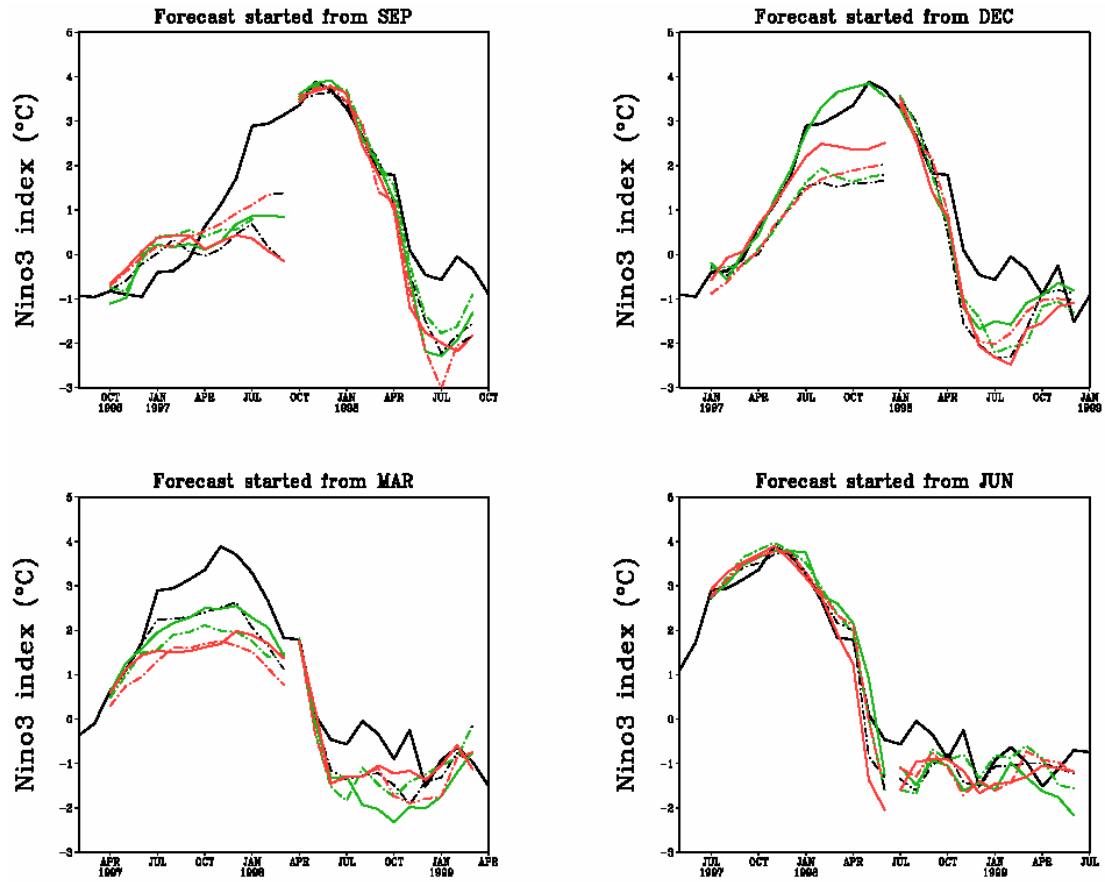
Figure 4.5.4 the same as 4.4.3, the black dashed line is the control forecast, the green lines are the BV ensembles initialized with AMIP data and oceanic analysis, the red lines are the BV ensembles initialized with the mean of the +BV and –BV one-month forecasted atmospheric field (initial oceanic fields remain the same).

(b) Case study for using different sizes of initial perturbations

We found that the size of the operational ensemble perturbation has a tendency that was large before and after the ENSO events, particularly for the cases during 1997-1998. This is similar to the behavior of the BV growth rate. But the bred ensemble perturbations have been used with a constant size in the Niño3 region, which is much smaller than the size of operational perturbations (in Niño3 region). Therefore, we propose to let the initial perturbations be proportional to the BV growth rate for the purpose of making better use of the ENSO-related growing instability obtained from the breeding method. Toth and Kalnay (1997) had also indicated that enlarging the size of the perturbations will improve the forecast in the later forecast months.

Here, we choose the case starting from September 1996 in order to test this idea. This case has a large growth rate (figure 4.4.2) and the control forecast shows a rather neutral condition after 6 forecast months instead of a significant warming occurring in observations (figure 4.5.1(a)). We multiply the original +/–BV perturbation by a factor of 5 and add/subtract it to the same initial conditions. The factor is used to emphasize the structure of the bred perturbations we obtained. Figure 4.5.5 is the rms error between the BV ensembles and the analysis field (shown as the blue lines), compared with the original size of the initial perturbations (shown as the green lines). With the 5-times large +BV perturbation, we are able to reduce the error in the later month forecasts (after 6 months). As a result of such improvement, the forecasted

warming trend is more realistic after April 1997.

A similar experiment is also tested for the case starting from March 1998, in which the BV growth rate is also large and the control forecast has the enormously cool anomalies in the early summer of 1998. We found that when using the large perturbation, the –BV forecast is able to lessen such a rapid cooling trend, which cannot be achieved when we replaced the atmospheric initial conditions.
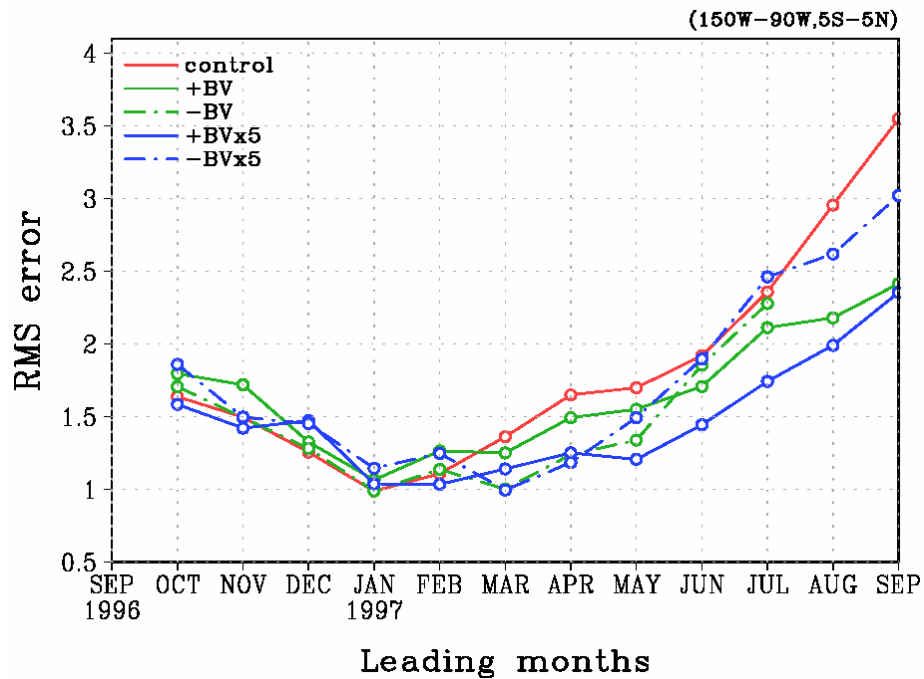
Figure 4.5.5 RMS error between the 12-month forecasted SST and the analysis field. The forecast started from September 1996, the rms error of the control forecast is denoted as the red line. The green lines are the forecasts with original size of the BV perturbations (‖BVSST‖=0.085°C) and the blue lines are the forecasts with BV perturbations 5 times larger than the original ones (the solid lines: +BV and the dash-dot lines: –BV).

Chapter 5: Summary and discussion

In this work we have studied the possibility of creating coupled initial ensemble perturbations for ENSO prediction in order to solve the deficiency of current strategies where the ensemble perturbations are not coupled. We used the breeding method to attempt to construct the initial perturbations associated with the slowly varying coupled ENSO instability by choosing the rescaling amplitude and time interval based on physical scales. We have examined the characteristics of the *coupled* bred vectors obtained from the NASA/NSIPP coupled General Circulation Model (CGCM) and also explored their potential applications for the oceanic data assimilation.

Breeding experiments were first performed under a perfect model scenario using a 10-year run of the NASA CGCM, and performing breeding with one month as the rescaling interval. This interval was chosen in order to allow for the saturation of "weather noise" while still retaining the slower ENSO instabilities. Our results suggest that the breeding method is indeed capable to capture the coupled instability associated with ENSO variability when dealing with a complex system such as the NASA CGCM that includes various instabilities of different time scales. In addition, our results agree in several aspects with those of Cai et al. (2003), who performed breeding experiments on the Zebiak-Cane model designed to represent the ENSO phenomenon. This is also a coupled model, but the atmosphere is simply a diagnostic

"slave" of the tropical ocean SSTs. First, Cai et al. (2003) demonstrated that the BV growth rates from this intermediate model is large between the ENSO events and is smallest at the mature stage of the events. The BV growth rate from the full NASA CGCM also exhibits such dependence on the phase of the ENSO cycle: the BV growth rate peaks 3 months earlier than ENSO events. Second, the structure of the bred vectors in Cai et al. (2003) shows that SST perturbations with large magnitudes are located in the tropical eastern Pacific that resemble the background ENSO variability. Similar structures are also found in the coupled BV mode derived from the NASA CGCM. Our results show that oceanic SST perturbations mainly grow in the eastern Pacific and that their structure is confined close to the equator. The tropical atmospheric bred vector exhibits a high-low pressure pattern in longitude in response of the boundary heating from the growing perturbations. As a consequence, the coupled bred vector is able to effectively perturb the Walker circulation. The geographic location of the bred vector suggests that the longitudinal structure of the mean thermocline (warm water distribution) determines the most dynamically sensitive regions that allow the amplification of the perturbations. The bred vector in the tropical upper ocean also exhibits a structure, comprising the propagation of the off-equator Rossby waves and of the equatorial Kelvin waves.

We have demonstrated the robustness of coupled bred vectors in this work by comparing the bred vectors obtained from two independent coupled GCMs (the NASA/NSIPP and NCEP/CFS CGCMs). There are many similarities between the main structures of these two bred vectors. In tropical Pacific, both bred vectors

confirm the role of the eastern Pacific in generating the fast growing perturbations associated with ENSO variability, including a couplet of off-equator upwelling Rossby waves that are accompanied by a downwelling Kelvin wave. The similarity between these two coupled systems can be also identified in an ENSO teleconnected region: North Pacific-North America. Both systems show that bred perturbations in this extratropical teleconnected region can be influenced by the growing signals in the deep tropical Pacific. Based on these results, we can confirm that the bred vector constructed in a coupled framework carries the information associated with the seasonal-to-interannual variability in time and in space. Therefore, the coupled bred vector seems an appropriate approach to be used for creating initial coupled ensemble perturbations that project on the ENSO-related instabilities.

We then applied this method to the current operational NASA Data Assimilation/CGCM used for predicting ENSO. This is a much more complex system than the perfect model coupled system since it assimilates oceanic observations daily and is coupled once a month with an AMIP atmosphere to launch one year forecasts. Our experiments should establish whether the present approach can be applied in practice within the operational ensemble forecast system. We also tried to explore whether breeding can have a potential impact in the data assimilation system for better use of oceanic observations in current data assimilation. For this purpose we focused on the one month forecast errors, measured by the difference between a one-month forecast and the analysis based on the observations. This "analysis increment" or forecast error, represents the corrections of the forecast field after the data

assimilation procedure. Since the analysis does not completely correct forecast errors, the analysis increment is also a proxy for the shape of the *analysis* errors.

The analysis increment has structures sensitive to the phase of the ENSO. Large analysis increments were found in the subsurface (at the thermocline) of the western Pacific at the onset of the 1997 El Niño event, and they propagated eastward as the event evolved. Their magnitude is smallest at the surface during the mature phase of the event and is large in the subsurface of the eastern Pacific. The bred vector, computed independently from the observations, exhibits a very similar structure and evolution to that of the analysis increment, especially in the eastward propagation along the equatorial thermocline. By examining the EOF modes of the subsurface temperature in the upper ocean, we found that both the analysis increment and bred vector project strongly on the background anomalous variations associated with ENSO events. Since the large portion (> 60%) of the subsurface temperature variance is associated the ENSO variability, it suggests a low dimensionality in the upper ocean heat content. The strong projection and the resemblance between the analysis increment and the bred vector imply that the evolution of dynamical errors is limited to a low dimensional space, which is constrained by the ENSO variability.

These results strongly suggest that the bred vector is able to capture the dynamic error structure as it appears in the forecast errors and analysis increments. We also found that such agreement between the analysis increment and the bred vector is particularly high when the BV growth rate is large and the background SST is close to

the neutral state. In other words, the dynamic error will grow rapidly and dominate the forecast error and the bred vector when the background is near neutral phase, as obtained by Cai et al (2003) with a much simpler model. The phases of the ENSO cycle are known to lock with the background annual cycle and such seasonal dependence is also identified in the BV growth rate. By examining the annual variation of the background SST, SST anomalies and the BV growth rate, we can link the interplay between the seasonal and interannual variability through the BV growth rate. This emphasizes the role of the eastern Pacific (the seasonal and interannual variabilities are much smaller in the western Pacific, but the signals appear first in the western thermocline). The background seasonal variations provide a mean dynamical structure that allows perturbations to grow rapidly in the early summer, since the background SST tendency is strongest around this time. Therefore, the BV growth rate reveals the chances that the background anomalies will develop in this region and reach their maximum in winter season.

Such time-dependent and background state dependent characteristics imply that there is a potential impact of using bred vectors in data assimilation. The current NASA data assimilation method is based on an ocean univariate Optimal Interpolation (OI) scheme, where the background (forecast) error covariance is independent of time. Although it has been documented that the ENSO prediction skill is strongly improved after the temperature data (TAO array) is assimilated in the oceanic initial condition (Rienecker 2000), the observations may not be optimally utilized since the constant background error covariance does not consider the

dynamic error evolution projecting on seasonal to interannual time scale.

In our results, the structure of bred vector is clearly insensitive to the choice of rescaling norm but there is a need to rescale the perturbations in the slowly varying component (ocean). Our result suggests that similar structures dominate the independently computed bred vector, but choosing different rescaling norms allows us to capture in some areas more than one growing direction. By locally comparing three initial bred vectors with three operational oceanic ensemble perturbations, we found that the bred vectors generally can explain larger amount of the oceanic analysis increment in the tropical domain, and that the total explained amount is more stable in time, compared with the projection obtained by using the operational ensemble perturbations. This implies an advantage of generating initial perturbation with a more dynamical framework. Our results also suggest that a few bred vectors constructed at a low computational cost can effectively span the space of the forecast error. This advantage can be applied in the ensemble Kalman filter (EnKF) data assimilation scheme (Keppenne and Rienecker, 2002), currently developing in NSIPP operational system. Coupled bred vectors can provide monthly dynamical evolving error and augment the ensemble-based background error covariance by supplying the effective correction pattern. In future experiments, the atmospheric slowly varying part may be also considered in the rescaling norm by combining the monthly-average wind field with the oceanic bred perturbation.

We then performed the ensemble forecasts with the two-sided breeding cycle for

a preliminary study of the impact of dynamically generated coupled initial perturbations. We chose 8 cases covering the 1997-1998 El Niño and the following La Niña event. The forecasts are initiated from March, June, September and December over two years. Since the ensemble forecasts with bred perturbations and operational ensemble perturbations were carried out by different versions of the NSIPP CGCM, it is difficult to compare them directly. Some of the preliminary results are: First, the ensemble members with positive bred perturbations tend to have better prediction skill than the ones with negative bred perturbations, suggesting that perturbations will carry their memory of the impact over many months. This would not be possible with operational ensemble perturbations generated separately and independently in time. Second, ensemble forecasts using one pair of bred vectors seem to have better forecast skill when forecast initialization is prior to the beginning of the event and when the control forecast skill is low, which tend to be also cases with large bred vector growth rate. Finally, the influence from using the dynamic perturbations is comparable with the ensemble forecasts using operational ensemble perturbations when the starting month is near the mature stage of the event, i.e., the initial background field has had strong signal about anomalous variation in the tropical Pacific and the uncertainty is "saturated".

In addition, the impact of using coupled bred vector was explored on the case of the +BV forecast of the case starting from December 1996, which captured the amplitude of warming strength, while the rest of the ensemble members show much weaker warming strength. The success of this +BV forecast can be identified as due

to a better representation of the wind stress in central Pacific and, thus, the forecasted thermocline is able to have a realistic displacement in both the eastern and the western Pacific. However, we do find it difficult to reduce the unrealistic rapid cooling happened in the early summer in 1998 (no matter when the starting forecast month is). This deficiency implies the importance for a CGCM to describe a more realistic annual cycle in order to correctly predict an ENSO event. In the NSIPP CGCM, it has been documented that the timing of the development of the cold tongue is not well represented.

Finally, we explored the impact of different atmospheric initial fields for the unperturbed forecast and for the two-sided breeding experiments. In these experiments, we replaced the AMIP initial fields with either the NCEP reanalysis data or the mean atmospheric fields from one-month forecasts of the pair of BV ensemble. Our results suggest that such replacement can have a strong influence on the prediction of the onset of the 1997 El Niño event.

We suggest that more experiments starting with different ENSO phases or different ENSO cases (strong, weak and long lasting El Niño/La Niña events) need to be performed in order to further understand the sensitivity of the bred vector in those situations. Also, the amplitude of the ensemble perturbations needs to be carefully investigated in order to best utilize the fast growing coupled perturbations. In our original experiments, we used perturbations after they were rescaled, i.e., with constant initial amplitude, without using the information of the bred vector growth

rate. One experiment showed that the prediction skill for the forecast in later month will increase when we amplified the perturbation size for a case with large BV growth rate. Therefore, we propose that ensemble perturbation should have an initial size proportional to the BV growth rate. This can be done in a simple way: use the bred vectors before they are rescaled times a constant chosen to insure that the varying perturbations are within reasonable ranges.

I plan to work with the NASA scientists to test and implement these ideas on the NASA system.

Appendices

A. The application of equatorial wave decomposition on oceanic bred vectors

Following a decomposition method proposed by Boulanger and Menkes (1995), we examine the wave structure in the oceanic bred vector and the background anomaly with the long wave approximation. This wave decomposition method projects the sea level and zonal current onto the meridional wave structures in order to investigate the wave propagation in the Pacific basin.

Defining a vector (**u**) including the nondimensionalized zonal current (u), and sea level height (h), we can expand it into Kelvin wave and Rossby waves.

$$\mathbf{u}(x,y,t) = \begin{pmatrix} u(x,y,t) \\ h(x,y,t) \end{pmatrix} = \sum_{n=0}^{\infty} r_n(x,t)\mathbf{R}_n(y) = \sum_{n=0}^{\infty} r_n(x,t)\begin{bmatrix} R_n^u(y) \\ R_n^h(y) \end{bmatrix} \qquad (A.1)$$

In Eq (A.1), $\mathbf{R}_0(y)$ is the Kelvin wave meridional structure and $\mathbf{R}_n(y)\big|_{n>0}$ ($n{>}0$) is the Rossby wave meridional. Those meridional structures of Kelvin and Rossby waves can be written as combinations of the Hermite functions. With the normalized Hermite functions (ψ_n),

$$\mathbf{R}_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} \psi_0 \\ \psi_0 \end{pmatrix} \text{ and } \mathbf{R}_0 = \sqrt{\frac{n(n+1)}{2(2n+1)}\begin{pmatrix} \dfrac{\psi_{n+1}}{\sqrt{n+1}} - \dfrac{\psi_{n-1}}{\sqrt{n}} \\ \dfrac{\psi_{n+1}}{\sqrt{n+1}} + \dfrac{\psi_{n-1}}{\sqrt{n}} \end{pmatrix}} \qquad (A.2)$$

The wave coefficients are calculated as

$$r_n(x,t) = \mathbf{u} \bullet \mathbf{R}_n = \int \left[u(x,y,t)R_n^u(y) + h(x,y,t)R_n^h(y) \right] dy \qquad (A.3)$$

Figure (A.1) is the meridional wave structure for the Kelvin wave and first four

Rossby waves (calculated for a 3 m/s Kelvin phase speed).

We then reconstruct the zonal current and sea level height using the Kelvin and

first to third Rossby waves according to Eq(A.1). This procedure is applied to the

background anomalies and the bred vector. For the background anomalies, these 4

modes had explained most of the variance in the tropical Pacific as shown in figure

A.2. Large amount of explained variance in the BV sea level height is particularly

located in the eastern Pacific and 5-7.5 degree off the equator (figure A.3(a)) and the

explained variance for BV zonal current is more limited to the equator (figure

A.4(b)). The spatial explained variance pattern in the bred vector suggests that the

evolution of the oceanic bred vector also has those equatorial waves structure. In

addition, the patterns in figure A.3(a) are very similar to the results obtained from

EOF analysis in section 3.3 (figure3.3.2(b),(c)), which has indicated the importance

of the Kelvin wave and the first Rossby wave.

We reconstruct the background and the bred vector surface height according to

the wave coefficient individually. Figure A.4 show the Kelvin and first two Rossby

waves from December 2020 to December 2022. The sea surface height for both

background and the bred vector during the same period are plotted in figure 3.2.4. By

this wave decomposition, we can see that the bred vector has the information

associated with the equatorial long waves, which control the ENSO cycle. The Kelvin

wave component in the bred vector exhibits eastward propagation particularly during

these warm and cold events and also locates locally in the eastern Pacific at the

mature phase of those events. Comparing to figure 3.2.4, similar feature in Rossby waves are also found in figure A.4(b) and (c).
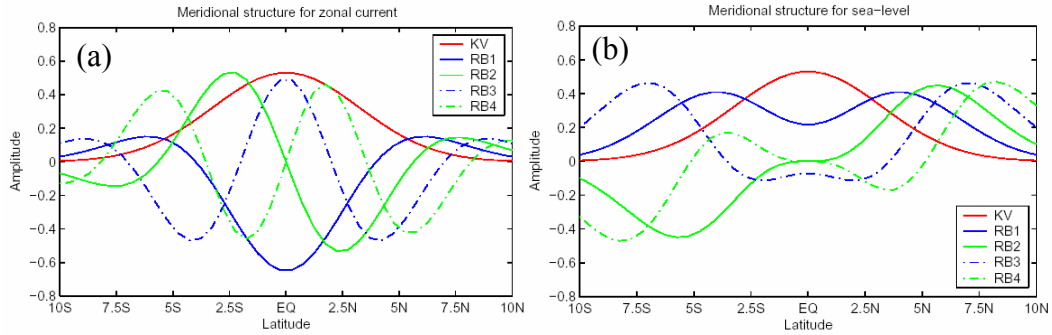
Figure A.1 Meridional structures of (a) zonal current (m/s) and (b) sea level height (cm) for Kevin and first to third Rossby waves. Modes are calculated for a 3 m/s Kelvin wave phase speed.
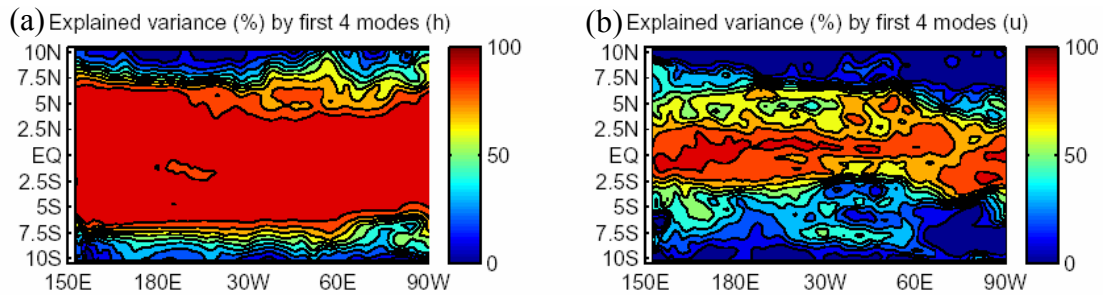
Figure A.2 Percentage of explained variance by using the Kelvin wave and the first to third Rossby waves for (a) background sea level height anomaly and (b) background zonal current anomaly.
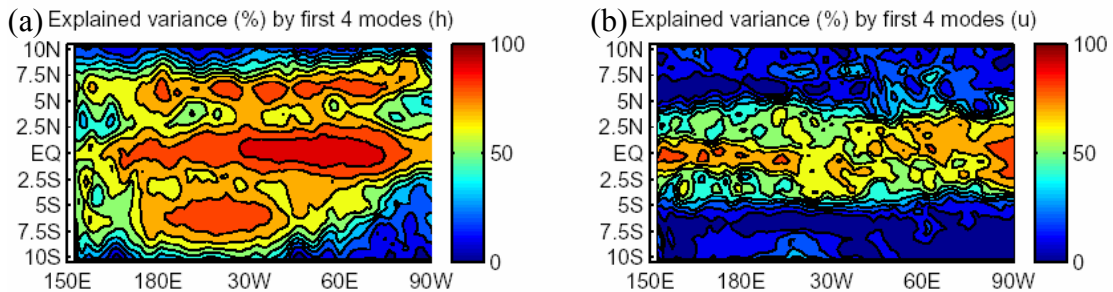
Figure A.3 Percentage of explained variance by using the Kelvin wave and the first to third Rossby waves for (a) BV sea level height and (b) BV zonal current.
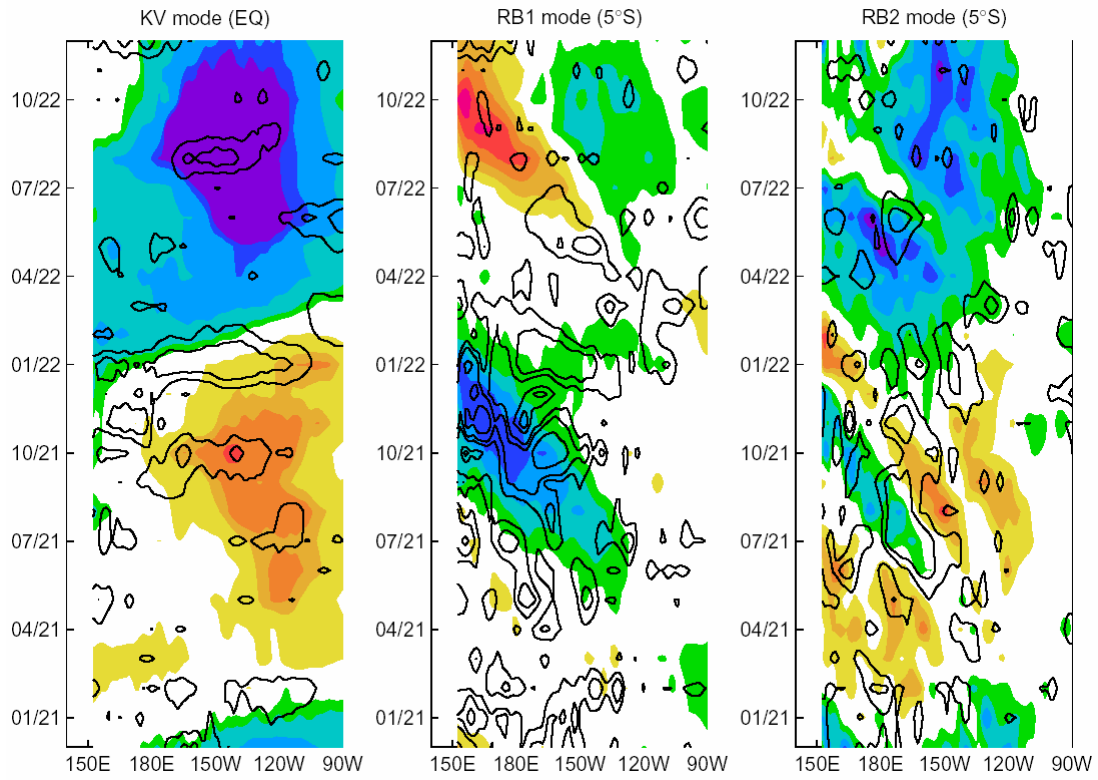
Figure A.4 The background sea surface height anomaly (color) and the bred vector (contour) are reconstructed for (a) Kelvin wave, (b) first Rossby wave and (c) second Rossby wave.

B. Errors of the day, bred vectors and singular vectors in a QG atmospheric model: implications for ensemble forecasting and data assimilation

B.1 Introduction

Numerical forecasts are sensitive to errors in the initial condition because of the chaotic characteristic of the atmospheric flow (Lorenz, 1965). Ensemble forecast are used to contain the uncertainties of initial conditions. Bred Vectors (BVs, used at NCEP) and singular vectors (SVs, used at ECMWF) are the current main methods for generating effective ensemble perturbations.

Day-to-day variability in the atmospheric state shows "errors of the day". Therefore, in data assimilation systems, the background error covariance needs to evolve with time in order to capture the evolution of the errors of the day. Current data assimilation schemes with this time-varying property are ensemble-based Kalman filter (EKF and LEKF) and 4DVAR, related respectively to BVs and SVs. We investigate the degree to which background (forecast) errors are related to BVs and SVs.

B.2 Experiment setup

The experiments in this section are done by a quai-geostrophic (QG) model developed by Rotunno and Bao (1996). It is a mid-latitude, beta plane, finite difference, channel model that is periodic in x and has impermeable walls at the north and south boundaries, and rigid lids at the top and bottom. Pseudo-potential vorticity is conserved except for Ekman pumping at the surface, ∇^4 horizontal diffusion and forcing by relaxation to a zonal mean state. The model is written in nondimensional

form and has 64 grid points in the zonal direction, 32 grid points in meridional direction and 7 levels in vertical direction. The model variables are potential vorticity defined at the interior levels (from levels 1 to 5) and potential temperature defined on the bottom and top levels (levels 0 and 6).

As in Morss (1998) and Hamill et al. (2000) and others, we use a single model integration as the true or "nature" run. 64 "Rawinsonde observations" are generated every 12 hours by randomly perturbing the true state at fixed observation locations, which were randomly chosen at the time of initialization. The simulated data assimilation is performed with a 3D-Var data assimilation scheme, constructed by Morss (1998). In our experiments, the same model is used to generate the truth and forecasts, assuming a perfect model scenario.

- Bred vector

The implementation of a breeding cycle and definition of bred vector has been described in Section 1.2 and Chapter 2. The rescaling period is 12 hour for the breeding experiments in this QG model.

- Singular vector

With a chosen norm and a chosen optimization time, the singular vector (SV) is a set of perturbation that will maximize the growth of the perturbation (also discuss in section 1.2). The initial SVs are the eigenvectors of the matrix, $\mathbf{L}^T\mathbf{L}$, where $\mathbf{L}$ is the tangent linear model of the original nonlinear model (i.e. the QG model) and $\mathbf{L}^T$ is the adjoint of the tangent linear model. The final singular vector is obtained by linearly evolving the initial SV for the chosen optimization time. In our results, the tangent

linear and the adjoint models are derived from the "Tangent linear adjoint model compiler" (TAMC, Giering, 1996). This compiler provides automatic differentiation in forward and adjoint (reverse) mode for programs written in Fortran. In our experiments, SVs are defined with the potential enstrophy norm and the optimization time is chosen as 24 hour. Also, we used the Lanczos algorithm to calculate the first 10 SVs.

B.3 The relationship between error of the day (background error), BV and SV

There exist several main differences in the spatial patterns obtained from these two methods. For example, the initial SV patterns are very sensitive to the choice of norm while the BV is not. Figure B.1 and B.2 are examples demonstrating such characteristic. In order to compare with the structure of the initial and final SVs, we present the two BVs, initializing with different initial random perturbation and show the BVs after 24 hours. Figure B.1 shows that BVs have the structure locally similar to the background error (contour). In addition, the difference between these two BVs suggests that they will not collapse into one single mode. This characteristic is also apparent in the BVs after 24 hours. By contrast, the first two initial SVs (figure B.2(a), (b)) are dominated by large scale features and did not project much on the background error even though their final SVs demonstrate a structure very similar to the BVs and project largely on the background error. In order to quantify the relationship between BV, SV and background error, we use two different quantities to represent their characteristics: (i) the local explained variance of forecast error by the BV and SV ensemble, representing how well the BV/SV describe the structure of

137

the background error and (ii) the E-dimension represents the local effective dimension of BV or SV ensemble.

- Comparison of the representation of forecast errors

The explained variance is calculated following the description in Section 4.3 as projecting 10 BVs /SVs to the background error and the local domain is chosen as 5×5 grid points. The same calculation is done for 5 BVs/SVs in order to understand the impact of ensemble size. In addition, for comparison, we construct a surrogate (random) BVs by choosing the BVs randomly from different time steps, in other words, we destroyed the time-dependency in the BV. The result is shown in figure B.3 and summarized as follows

- The BVs and final SVs have similar result and explain similar amount of the variance of the forecast error. The initial SVs always explain least of the forecast error variance no matter the size of the ensemble.

- When randomly choosing the BVs, they did not include the information of "error of the day" therefore explained variance is low when considering the same ensemble size.

- When we increase the BV and SV to 10 vectors, the results explain more than 90% of the forecast error. But changing the size of the SVs did not change the initial structure of the subspace spanned by SVs and the main difference only shows in their final structure.

- The E-dimension

The method of computing the E-dimension is following the work by Patil et al. (2001). This quantity estimates the local "effective" dimension of an ensemble, in other words, it describes the similarities among the ensemble members.

$$\Psi(\sigma_1, \sigma_2, \cdots, \sigma_k) = \frac{(\sum_{i=1}^{k} \sigma_i)^2}{\sum_{i=1}^{k} \sigma_i^2} \qquad (B.1)$$

Illustring in Eq. (B.1), we applied singular value decomposition to the matrix composed by 10 vectors (k=10) and the ith singular value is denoted as σ_i. Details and more discussions of the E-dimension can be found in Patil et al. (2001) and Corraza et al. (2002).

- As shown in figure B.4, our results show that the BV E-dimension converges to a level of 2.7 after 10 days and final SV consistently has a lower E-dimension. Also, the initial SV has a lowest E-dimension (mean E-dimension is less than 2), suggesting the initial SVs comprises a rather flat shape. This indicates that even though the initial and final SVs are globally orthogonal, locally they have very similar shapes.

- We also notice that when adding small amount of random perturbations during the breeding cycle, it can help to accelerate the convergence of the E-dimension. This may indicate the potential application in the ensemble-based data assimilation to generate effective perturbations. Refreshing the bred vector by small amount of random perturbations may be able to help to expand the subspaces spanned by the ensemble vectors. Therefore, the probability of projecting the perturbations on the favorable growing direction is increased by the additional random perturbations.

– Our results using surrogate BVs (randomly chosen) exhibit a much high E-dimension than either BV or SV, i.e. the ensemble vectors are quite different to each other. Combining with figure B.3, it suggests that the perturbations are unable to effectively project on forecast error since the surrogate perturbations have included many ungrowing directions. However, the high and consistent E-dimension also implies that the probability for such perturbations to project on growing direction is stable although is low. This can explain the stable performance of the 3DVAR scheme since the background error covariance is constant in time in the 3DVAR.

B.4 Implications

Our results suggest that beyond the optimization period SVs and BVs should be similar. But, earlier forecast and initial time, the BVs are better representing the "error of the day", which implies the influence of having the background error evolving with time. The relationship between SVs and 4D-Var suggests that the final 4D-Var state is accurate, but the initial state may be not. In addition, "refreshing" BVs with random perturbations implies that increasing the probability of projecting the perturbations on the favorable growing direction may be a better alternative to variance inflation. The impact of the small amount of perturbations (some articles refer to this as a stochastic seeding) has been found contribute to large amount of improvement in data assimilation system (Corazza et al, 2002, Szunyogh et al, 2004 and Yang et al., submitted).

(a) Initial BV1

(c) Final BV1 (after 24 hours)

(b) Initial BV2

(d) Final BV2 (after 24 hours)
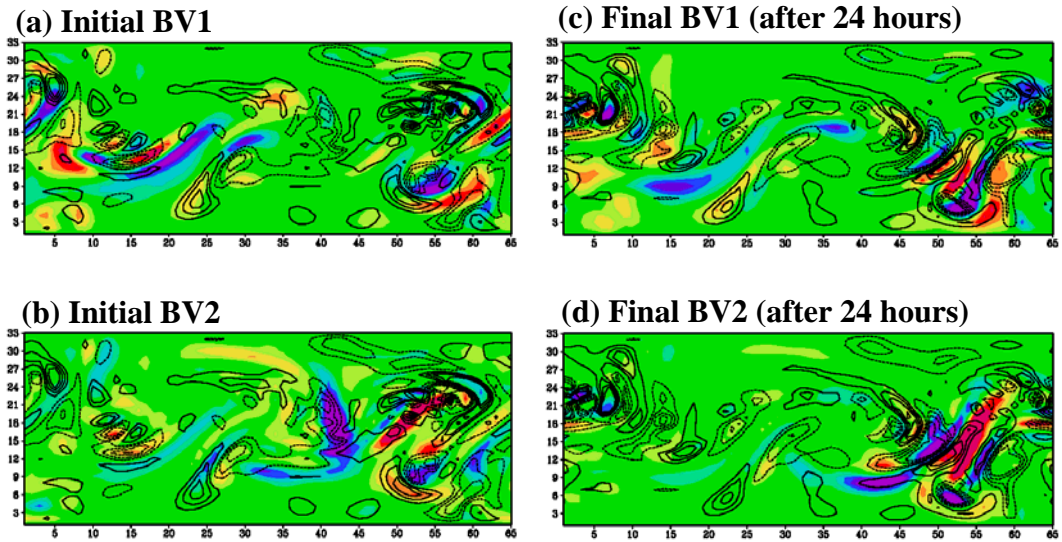
Figure B.1 (a) the initial bred vector (color), selected as the bred vector after 3-day integration. The contour lines are the background error at the initial time. (b) same as (a) but initialized with different random perturbation. (c) the final bred vector obtained by integration (a) for 24 hour and the contour lines are the background error after 24 hours, and (d) the same as (c) but with (b) as the initial condition.

(a) Initial SV1 sigma^2=2815

(c) Final SV1 (after 24 hours)

(b) Initial SV2 sigma^2=1035

(d) Final SV2 (after 24 hours)
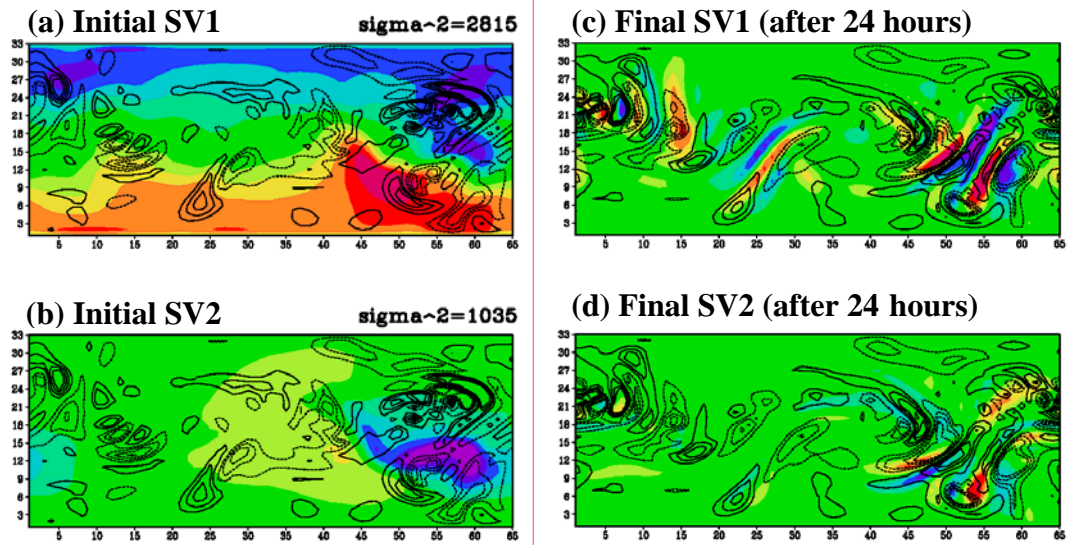
Figure B.2 (a) the first initial singular vector (color), the reference trajectory is using the background state in figure B.1. The contour lines are the background error at the initial time. (b) the second initial singular vector, (c) the final singular vector corresponds to initial SV in (a). The contour lines are the background error after 24 hours, and (d) the final singular vector corresponds to initial SV in (b). The sigma square values are the corresponding singular values and the contour lines have the same definition in figure B.1.
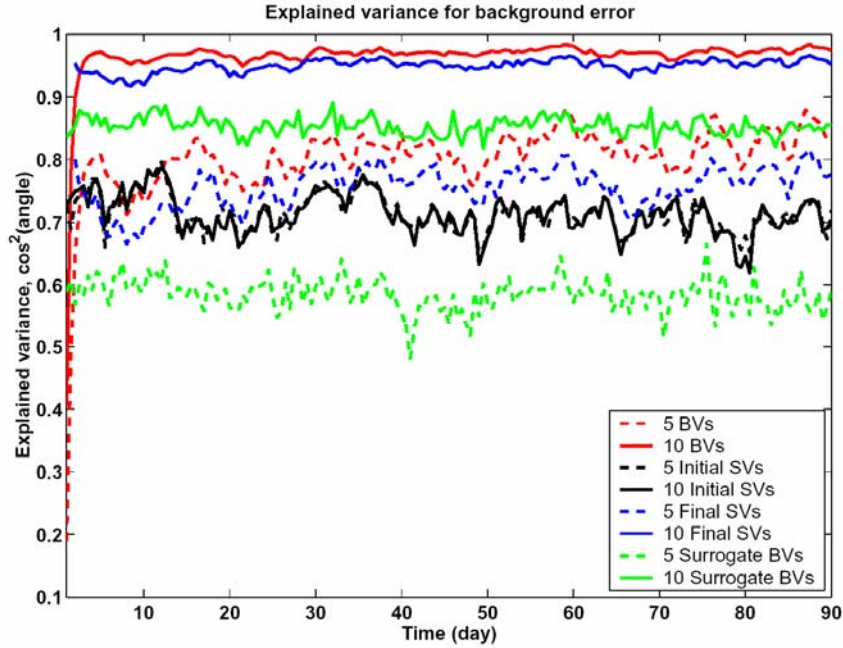
Figure B.3 The explained variance by different types of dynamic vectors. The red lines denote the BVs, the black lines denote the initial SVs, the blue lines denote the final SVs and the green lines denote the surrogate BVs. The dash lines are for the ensemble size 5 and the solid lines are for the ensemble size 10.
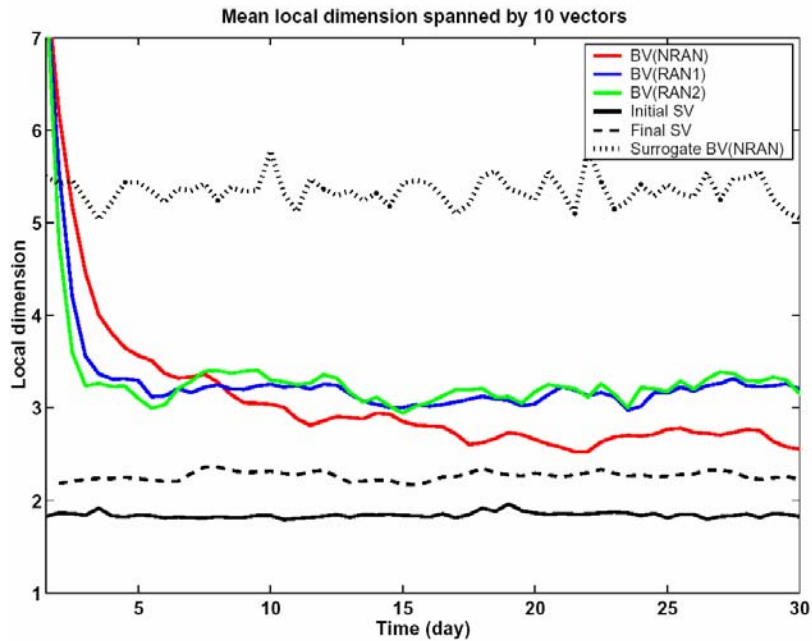
Figure B.4 The E-dimension of different types of dynamic vectors (ensemble size is 10). The red line denotes BV, the solid-black line denotes initial SV, the dash-black line denotes final SV, the blue line denotes BV refreshed with a random perturbation of size 0.1, the green line denotes BV refreshed with a random perturbation of size 0.2 and the dot-black line denotes the surrogate BV (no random perturbation).

C. Data Assimilation as Synchronization of Truth and Model: Experiments with the Three-Variable Lorenz System

S.-C. Yang, D. Baker, H. Li,

K. Cordes, M. Huff, G. Nagpal, E. Okereke, J. Villafañe, E. Kalnay and G. Duane

Submitted to Journal of Atmospheric Sciences

Abstract: The potential use of chaos synchronization techniques in data assimilation for numerical weather prediction models is explored here through experiments with the Lorenz 3-variable model. Our experiments show that synchronization takes for a wide range (over two orders of magnitude) in the coupling coefficient. We compare a coupling scheme based on coupling along the direction of either bred vector or singular vector. Our results suggest that coupling along dynamically chosen directions has the potential to improve current chaos synchronization schemes. Experiments on "generalized synchronization" (GS) were performed by letting one of the parameters in the slave equation differ from those that guide the master's dynamics. We find that GS is easier to attain than identical synchronization even with low coupling strengths but the slaves provide only partial information about the master.

A direct comparison with a standard data assimilation technique, 3-Dimensional Variational Analysis (3D-Var), demonstrates that this scheme is slightly more effective in producing an accurate analysis than the simpler synchronization scheme. We note that higher growth rates of bred vectors from both the master and the slave anticipate the location and size of error spikes in both 3D-Var and synchronization,

whereas the more advanced data assimilation method of Kalman Filtering, yielding the most accurate analyses, avoids large error spikes through the use of adaptive weights. Adaptive synchronization, with a coupling parameter proportional to the bred vector, succeeds in reducing episodes of large error growth. Our results suggest that a hybrid chaos synchronization-data assimilation approach may provide an avenue to improve and extend the period for accurate weather prediction.

Bibliography

Alves O., M. Balmaseda, D. Anderson, and T. Stockdale, 2002: Sensitivity of dynamical seasonal forecasts to ocean initial conditions. *Quart. J. Roy. Met. Soc.,* **103**, 647-667.

Arakawa, A., and M. J. Suarez. 1983: Vertical differencing of the primitive equations in sigma coordinates. *Monthly Weather Review*: **111**, 34–45.

Bacmeister, J., P. J. Pegion, S. D. Schubert, and M. J. Suarez, 2000: Atlas of seasonal means simulated by the NSIPP I Atmospheric GCM. *NASA/TM-2000-104606*, **17**.

Bacmeister, J. T., and M. J. Suarez, 2002: Wind stress simulations and the equatorial momentum budget in an AGCM. *J. Atmos. Sci.*, **59**, 3051-3073.

Battisti, D. S., 1988: The dynamics and thermodynamics of a warm event in a coupled tropical atmosphere-ocean model. *J. Atmos. Sci.,* **45**, 2889-2919.

Bengtsson, L. U. Schleese, E. Roeckner, M. Latif, T. P. Barnett, and N. Garham, 1993: A two tiered approach to climate forecasting. *Science*, **261**, 1026-1029.

Bloom, S.C., L.L. Takacs, A.M. da Silva, and D. Ledvina. 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256–1271.

Boffetta, G., A. Crisanti, F. Papatella, A. Provenzale, and A. Vulpiani, 1998: Slow and fast dynamics in coupled systems: A time series analysis view, *Physica D.*, 116, 301-312.

Boulanger, J.-P., and C. Menkes, 1995: Propagation and reflection of long equatorial waves in the Pacific and reflection of long equatorial waves in the Pacific Ocean

during 1991-1993 El Nino. *J. Geophys. Re*s., **100,** 25041-25059.

Buizza, R., and T. N. Palmer, 1995: The singular vector structure of the atmospheric global circulation. *J. Atmos. Sci.,* **52,** 1434–1456. .

Cai, E. Kalnay, and Z. Toth, 2003: Bred vectors of the Zebiak-Cane model and their application to ENSO predictions. *J. Climate*, **16**, 40-55.

Cane, M. A., S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño, *Nature*, **321**, 827-832.

Zebiak, S. E., and M. A. Cane. 1987: **A** Model El Niño–Southern Oscillation. *Monthly Weather Review*, **115**, 2262–2278.

Caplan, P., J. Derber, W. Gemmill, S.-Y. Hong, H.-L. Pan, and D. Parrish, 1997: Changes to the 1995 NCEP operational medium-range forecast model analysis-forecast system. *Wea. Forecasting*, **12**, 581-594.

Chen, D., S. E. Zebiak, A. J. Busalacchi, and M. A. Cane, 1995: An improved procedure for El Niño forecasting. *Science*, **269**, 1699-1702.

Chen, D., M. A. Cane, A. Kaplan, S. E. Zebiak, and D. Huang, 2004: Predictability of El Niño over the past 148 years, *Nature*. **428**, 733-736.

Chen, Y.-Q., D. S. Battisti, T. N. Palmer, J. Barsugli, and E. S. Sarachik, 1997: A study of the predictability of tropical Pacific SST in a coupled atmosphere-ocean model using singular vector analysis: The role of annual cycle and the ENSO cycle. *Mon. Wea. Rev.*, **125**, 831-845.

Chou, M.-D., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. *NASA Technical Memorandum*, 104606, **10**, 84pp.

Chou, M.-D., and M. J. Suarez, 1999: A solar radiation parameterization for atmospheric studies. *NASA Technical Memorandum*, 104606, **11**, 40pp.

Contreras, R. F. 2002: Long-term observations of tropical instability waves. *J. Phys. Oceanogr.* **32**, 2715–2722.

Corazza, M., E. Kalnay, D. J. Patil, S.-C. Yang, R. Morss, M. Cai, I. Szunyogh, B. R. Hunt, and J. A. Yorke, 2002: Use of the breeding technique to estimate the structure of the analysis "error of the day". Nonlinear Processes in Geophysics, **10**, 233-243.

Errico, R. M., and T. Vukicevic. 1992: Sensitivity Analysis Using an Adjoint of the PSU-NCAR mesoseale Model. *Mon. Wea. Rev.,* **120**, 1644–1660.

Fan, Y., M. R. Allen, D. L. T. Anderson, and M. A. Balmaseda, 2000: How predictability depends on the nature of uncertainty in initial conditions in a coupled model of ENSO. *J. Climate*, **13**, 3298-3313.

Gates, W.L., 1992: AMIP: The Atmospheric Model Intercomparison Project. *Bull. Am. Meteorol. Soc.*, **73**, 1962-1970.

Giering, Ralf, and T. Kaminski, 1998: Recipes for adjoint code construction. *ACM Trans. On Math. Software.* **24**, 437-474.

Gill, A. E., 1982: Some simple solutions for hear-induced tropical circulation. *Quart. J. Roy. Meteor. Soc.*, **106**, 447-462.

Gu, D., and S.G.H. Philander 1995: Secular changes of annual and interannual variability in the tropics during the past century. *J. Climate*, **8**, 864–876.

Hakkinen, S., and G.L. Mellor, 1992: Modeling the seasonal variability of the coupled Arctic ice-ocean system, *J. Geophys. Res.*, **97**, 20285-20304.

Hamill, T. M., Snyder, C., and Morss, R. E., 2000: A comparison of probabilistic forecasts from bred, singular vector and perturbed observation ensembles. *Mon. Wea. Rev.*, **128**, 1835-1851.

Ji, M., and A. Leetmaa. 1997: Impact of data assimilation on ocean initialization and El Niño prediction. *Mon. Wea. Rev.*, **125**, 742-753.

Latif, M., D. Anderson, T. Barnett, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, and E. Schneider, 1998: A review of predictability and prediction of ENSO, *J. Geophys. Res*., **103**, 14,375-14,393.

Levitus, S., World Ocean Atlas 1994: CD Rom Sets, *National Oceanographic Data Center Informal Report*, **13**.

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

Louis, J., M. Tiedke, J. Geleyn, 1982: A short history of the PBK parameterization at ECMWF. In: Proceedings, ECMWF *Workshop on Planetary Boundary Layer Parameterization*, Reading, UK, 59-80.

Kalnay, E. and Z. Toth, 1994: Removing growing errors in the analysis cycle. Preprints of the AMS 10th Conference on Numerical Weather Prediction, in Portland, OR, July 18-22 1994, 212-215.

Kanamitsu, M., 1989: Description of the NMC global data assimilation and forecast system. *Wea. Forecasting*, **4**, 334-342.

Keppenne, C.L., and M.M. Rienecker, 2002: Development and initial testing of a parallel Ensemble Kalman filter for the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.,* **130**, 2951-2965.

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

_____, Y. Tang, and A. M. Moore. 2003: The calculation of climatically relevant singular vectors in the presence of weather noise as applied to the ENSO problem. *J. Atmos. Sci.*, **60**, 2856–2868.

Koster, R. D., and M. J. Suarez, 1996: Energy and water balance calculations in the Mosaic LSM. *NASA Technical Memorandum*, 104606.

Miller, S. K., M. M. Rienecker, M. J. Suarez, P. J. Pegion and J. T. Bacmeister, 2004: The GMAO CGCM VI – climatology and interannual variability. *NASA Technical Memorandum*, 104606.

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

——, and ——, 1997b: The singular vectors of a coupled ocean–atmosphere model of ENSO. Part II: Sensitivity studies and dynamical significance. *Quart. J. Roy. Meteor. Soc.,* **123,** 983-1006.

——, and ——, 1999a: Stochastic forcing of ENSO by the intraseasonal oscillation. *J. Climate,* **12,** 1199-1220.

——, and ——, 1999b: The non-normal nature of El Niño and intraseasonal variability. *J. Climate*, **12**, 2965-2982.

——, J. Vialard, A. T. Weaver, D. L. T. Anderson, R. Kleeman and J. R. Johnson.

2003: The role of air–sea interaction in controlling the optimal perturbations of low-frequency tropical coupled ocean–atmosphere modes. *J. Climate*: **16**, 951-968.

Moorthi, S., and M. J. Suarez, 1992: Relaxed Arakawa-Schubert: a parameterization of moist convection for general circulation models. *Mon. Wea. Rev.*, **120**, 978-1002.

Morss, R. E., 1998: Adaptive observations: Idealized sampling strategies for improving numerical weather prediction. PhD thesis, Massachusetts Institute of Technology. 225pp.

Pacanowski, R. C., and S. G. H. Philander, 1981: Parameterization of vertical mixing in numerical models of the tropical oceans, *J. Phys. Oceanogr.*, **11**, 1443-1451.

____, R. C., and S. M. Griffies, 1998: **MOM 3.0 Manual**, NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, USA 08542.

Palmer, T. N., R. Gelaro, J. Barkmeijer and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633-653.

Patil, D. J. S., B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, 2001: Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.*, **86**, 5878.

Peña, M. and E. Kalnay, 2004: Separating fast and slow modes in coupled chaotic system. *Nonlinear Process. Geophys.*, **11**, 319-327.

Reynolds, R.W., N.A. Rayner, T.M., Smith, D.C. Stokes, and W. Wang, 2002: An Improved In Situ and Satellite SST Analysis for Climate, *J. Climate*, **15**, 1609-1625.

Rienecker, M., M. Suarez, D. Admec, R. Koster, S. Schubert, and J. Hansen, 2001: The NASA seasonal-to-interannual prediction project annual report.

Ropelewski, C.F., and M.S. Halpert, 1996: Quantifying Southern Oscillation-precipitation relationships. *J. Climate*, **9**, 1043-1959.

Rosati, A., K. Miyakoda, and R. Gudgel, 1997: The impact of ocean initial conditions on ENSO forecasting with a coupled model. *Mon. Wea. Rev.*, **125**, 754-772.

Rotunno, R. and Bao, J. W., 1996: A case study of cyclogenesis using a model hierarchy. *Mon. Wea. Rev.*, **124**, 1051-1066.

Sadourney, R., 1975: The dynamics of finite difference models of the shallow water equations, *J. Atmos. Sci.*, **32**, 680-689.

Schopf, P., and Loughe, A., 1995: A reduced-gravity isopycnal ocean model: hindcasts of El Niño. *Mon. Wea. Rev.*, **123**, 2839-2863.

Sterl, A., and A. Kaltenberg, 1994: Embedding a mixed layer model into an ocean general circulation model of the Atlantic: The importance of surface mixing for heat flux and temperature. *J. Geophys, Res.*, **99**, 14,139-14,157.

Stockdale, T. N., D. L. T. Anderson, J. O. S. Alves, and M. A. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature*, **392,** 370-373.

Surarez, M. J., and P. S. Schopf, 1988: A delayed action oscillator for ENSO. *J. Atmos. Sci.*, **45**, 3283-3287.

Saha, S., W. Wang, S. Nadiga, H.-L. Pan, and G. White, 2004: The new seasonal forecast model at NCEP. Extended abstract of the 13th Conference on Interactions of the Sea and Atmosphere, Portland, ME.

Suarez, M., 1996: Dynamical aspects of climate simulations using the GEOS general circulation model, *NASA Technical Memorandum 104606*, 10, p77.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, B. Hunt, and E. Kalnay et al., **2004**: A local ensemble Kalman filter for the NCEP GFS. Preprints of the 20th conference on weather analysis and forecasting/ 16th conference on numerical weather prediction. Seattle, WA, USA, Janurary 11-15, 2004.

Takacs, L. L. and M. J. Suarez, 1996: Dynamical aspects of climate simulations using the GEOS GCM. *NASA Technical Memorandum*, 104606, **10**, 56.

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

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317-2330.

___ and ___, 1996: Climate ensemble forecasts: How to create them?. *Idojaras,* **100**, 43-52.

___ and ___, 1997: Ensemble forecast at NCEP and the breeding Method. *Mon. Wea. Rev.*, **125,** 3297-3319.

Troccoli, A. and K. Haines, 1999: Use of the Temperature-salinity relation in a data assimilation context. *J. Atmos. Oceanic Technol*., **16**, 2011-2025.

___, M. Balmaseda, J. Segnschneider, J. Vialard, D. Anderson, K. Haines, T. Stockdale, F. Vitart, and A. D. Fox, 2002: Salinity adjustments in the presence of temperature data assimilation. *Mon. Wea. Rev*., **130**, 89-102.

___, M. M. Rienecker, C.L. Keppenne, and G.C. Johnson, Temperature data assimilation with salinity corrections: validation in the tropical Pacific Ocean,

1996-1998, *J. Geophys. Res.*, submitted, 2003.

Vialar, J., F. Vitart, M. A. Balmaseda, T. N. Stockdale, and D. L. Anderson, 2003: An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model. *ECMWF Technical memorandum*, **417**.

Vintzileos, A., and R. Sadourny. 1997: A general interface between an atmospheric general circulation model and underlying ocean and land surface models: Delocalized physics scheme**.** *Mon. Wea. Rev.*, **125**, 926-941.

____, M. M. Rienecker, M. J. Suarez, S. K. Miller, P. J. Pegion, and J. T. Bacmeister, 2003: Simulation of the El Niño-interannual prediction project coupled general circulation model. *CLIVER Exchanges*, **8**, 25-27.

Wu, W., M. Iredell, S. Saha, and P. Caplan, 1997: Changes to the 1997 NCEP operational MRF model analysis/forecast system. *NCEP Technical Procedures Bulletin*, **443**.

Xie, S. 1995: Interaction between the annual and interannual variations in the equatorial Pacific. *J. Phys. Oceanogr.*, **25**, 1930-1941.

Xue, Y., M. A. Cane, S. E. Zebiak, and M. B. Blumenthal, 1994: On the prediction of ENSO: A study with a low order Markov model. *Tellus,* **46A,** 512-528.

____. ____, and ____, 1997a: Predictability of a coupled model of ENSO using singular vector analysis. Part I: Optimal growth in seasonal background and ENSO cycles. *Mon. Wea. Rev*, **125**, 2043-2056.

___, ___, and ___, 1997b: Predictability of a coupled model of ENSO using singular vector analysis. Part II: Optimal growth and forecast skill. *. Mon. Wea. Rev.*, **125**, 2057-2073.

Yang, S.-C., D. Baker, H. Li, K. Cordes, M. Huff, G. Nagpal, E. Okereke, J. Villafañe, E. Kalnay, and G. Duane: Data assimilation as synchronization of truth and model: Experiments with the three-variable Lorenz system. *J. Atmos. Sci.*, submitted, 2005.

Yang, S., K. Lau, and P. Schopf, 1999: Sensitivity of the tropical Pacific Ocean to precipitation induced freshwater flux. *Clim. Dynam.*, 15, 737-750.

Zebiak, S. E. and M. A. Cane, 1987: A model El Niño-Southern Oscillation, *Mon, Wea. Rev.*, **115,** 2262-2278.