The Differences between the Optimal Perturbations of Coupled Models of ENSO

ANDREW M. MOORE

Program in Atmospheric and Oceanic Sciences, and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado

RICHARD KLEEMAN

Courant Institute for Mathematical Sciences, New York University, New York, New York

(Manuscript received 9 September 1999, in final form 6 March 2000)

ABSTRACT

The optimal perturbations (singular vectors) of a dynamical coupled model, a hybrid coupled model, and a linear inverse model of ENSO are compared. The hybrid coupled model consists of a dynamical ocean model and a statistical atmospheric model. The dynamical ocean model is identical to that used in the dynamical coupled model, and the atmospheric model is a statistical model derived from long time series of the dynamical coupled model. The linear inverse model was also derived from long time series from the dynamical coupled model. Thus all three coupled models are very closely related and all produce similar ENSO oscillations. The dynamical model and hybrid model also possess similar levels of hindcast skill. However, the optimal perturbations of the tangent linear versions of each model are not the same. The hybrid and linear inverse models are unable to recover the SST structure of the optimal perturbations of the dynamical model. The SST structure of the dynamical coupled model is a result of nonnormality introduced by latent heating of the atmosphere by deep convection over the west Pacific warm pool. It is demonstrated that standard statistical techniques remove the effects of the latent heating on the nonnormality of the hybrid and linear inverse models essentially rendering them more normal than their dynamical model counterpart. When the statistical components of the hybrid coupled model and the linear inverse models were recomputed using SST anomalies that are appropriately scaled by the standard deviation of SST variability, nonnormality was reintroduced into these models and they recovered the optimal perturbation structure of the dynamical model. Even though the hybrid and linear inverse model with scaled SSTs can recover the large-scale features of the correct optimal structure, state space truncation means that the dynamics of the resulting optimal perturbations is not the same as that governing optimal perturbation growth in the dynamical model. The consequences of these results for observed estimates of optimal perturbations for ENSO are discussed.

1. Introduction

During the last decade the interest in seasonal prediction of the coupled ocean–atmosphere system in the Tropics has increased considerably and there are now many groups that regularly issue forecasts of El Niño–Southern Oscillation (ENSO) (e.g., Barnston et al. 1994; *COLA Experimental Long Lead Forecast Bulletin*). Naturally the increasing interest in prediction has led to an increasing interest in the predictability of ENSO and numerous studies have addressed this issue in various ways (e.g., Goswami and Shukla 1991; Blumenthal 1991; Davey et al. 1994; Webster 1995; Latif et al. 1994; Chen et al. 1997; Kleeman and Moore 1997, 1999; Moore and Kleeman 1998). More recently the ENSO community have followed the lead of numerical weather prediction in an attempt to identify the most rapidly growing disturbances in the coupled system using the ideas embodied in the generalized stability framework of Farrell and Ioannou (1996a,b). These perturbations are also of considerable theoretical interest because they suggest possible precursors for ENSO events as well as dynamical mechanisms that may be particularly effective for triggering or disrupting events. The most rapidly growing perturbations of the coupled system are the singular vectors of the linearized operators that describe the system dynamics. Since the singular vectors are the most rapidly growing perturbations that can exist in the system while linear dynamics are valid, they are often referred to as the optimal perturbations (Farrell 1982).

To date several groups have computed the optimal perturbations of various different coupled models. Due to the computationally intensive nature of these calculations all but one of these studies have been limited to coupled models of intermediate complexity. The models

Corresponding author address: Dr. Andrew Moore, University of Colorado, Campus Box 311, Boulder, CO 80309-0311.
E-mail: andy@australis.colorado.edu

used include purely dynamical models (Moore and Kleeman 1996, 1997a,b; Chen et al. 1997; Thompson 1998), models with a dynamical ocean and a statistical atmosphere (Fan et al. 2000; Eckert 1999), and linear inverse (statistical) models fitted to either the output of an existing dynamical coupled model (Blumenthal 1991; Xue et al. 1994, 1997a,b; Fan 1998; Fan et al. 2000; Thompson and Battisti 2001) or fitted to observations (Penland and Sardeshmukh 1995; Y. Xue 1999, personal communication; Johnson et al. 1999, unpublished manuscript). Each model produces what one might consider reasonable simulations of ENSO and demonstrates similar levels of ENSO forecasting skill. However, their optimal perturbations, while qualitatively similar with respect to thermocline structure, show pronounced differences in sea surface temperature (SST).

For optimal growth times, ~90 days, the thermocline structure of the optimal perturbations is often characterized by an initially large displacement in the west Pacific that propagates east as an equatorial Kelvin wave initiating an ENSO episode upon its arrival in the central and east Pacific (see Fig. 17a for an example) (Moore and Kleeman 1996; Thompson 1998). Fan et al. (2000) find a similar result for longer optimal growth times depending on the perturbation growth norm chosen. Depending on the model, these changes are supplemented by long oceanic Rossby waves also present in the initial thermocline structure that reflect from the western boundary (also apparent in Fig. 17a). For longer optimal growth times the thermocline structure in some models is more uniform along the equator (Xue et al. 1997a; Thompson 1998; Fan et al. 2000). Recent linear inverse model calculations of the optimal patterns of upper-ocean heat content anomalies suggest that thermocline anomalies are largest in the west and central Pacific but do not extend to the western boundary (Johnson et al. 1999).

The SST optimal perturbation structures of each coupled model agree to a lesser degree with each other, although most do share the common feature of having largest initial amplitude in the east and/or central Pacific. There is some deviation though in the geographical location of the SST maximum. The model that stands apart from all the other studies, however, is that of Kleeman (1993), which produces an optimal SST structure that has largest amplitude primarily in the western Pacific. The perturbation heat flux of this optimal perturbation (Kleeman and Moore 1997) qualitatively resembles that observed in connection with intraseasonal variability, and it has been argued by Moore and Kleeman (1997b, 1999a,b) that this is consistent with the idea first proposed by Lau (1985) that intraseasonal variability may act as a trigger for ENSO, or act as a significant stochastic forcing of the system. Observations during the 1997/98 El Niño event in part support this idea (McPhaden 1999). Apart from perhaps the recent studies of Eckert (1999) and Y. Xue (1999, personal communication) none of the other coupled models place such a strong emphasis on the west Pacific SST as a potentially important region for SST perturbations either as a trigger for ENSO, as a source of stochastic forcing, or as a region of potentially rapid error growth in forecasts when viewed in terms of the fastest growing optimal perturbation. Despite the SST differences between the optimal perturbation structures of different models, all of the studies to date agree that the fastest growing optimal perturbation in each case always develops into an ENSO episode. Therefore the optimal perturbations of each model, despite their differences, do offer a means of understanding forecast error growth and the predictability of ENSO within a particular model. However, lack of consensus on optimal perturbation structure casts a shadow over their value as an indicator of the dynamics that may be important for triggering ENSO episodes in nature, and the consequences of this for limiting ENSO predictability.

In this paper we will explore some of the possible reasons for the aforementioned differences that exist between the optimal perturbations of different coupled models. It will be shown that the effects of important nonlinearities in the coupled system can be misinterpreted or completely ignored by standard statistical analysis techniques. This can have a considerable influence on the structure and growth rate of the resulting optimal perturbations of linear inverse models or coupled models that include a statistical atmospheric component. The optimal perturbations of such models should therefore be treated with caution unless steps have been taken to include the influence that important nonlinearities in the system have on optimal perturbation structure.

A variety of coupled models derived from the dynamical coupled model of Kleeman (1993) have been used in this study, and each model is described in detail in section 2. The optimal perturbations of each coupled model are described in section 3 and the differences between them are also explored in section 4 using the concept of nonnormal systems that is central to generalized stability theory. Other issues relating to the structure of the optimal perturbations such as state space truncation and choice of norm are explored in sections 4 and 5. A summary and discussion of the main results are presented in sections 6 and 7 along with important conclusions.

2. The coupled models

One dynamical coupled model, two hybrid coupled models, and two linear inverse models were used in this study. The hybrid and linear inverse models were derived from the dynamical coupled model so all of the models used here are very closely related.

a. The dynamical model, DYN

The dynamical coupled model, hereafter referred to as DYN, is an intermediate coupled model of the tropical

Pacific Ocean and global atmosphere (Kleeman 1993) and computes anomalies about the annual mean climatological state of the system prescribed from observations. A version of the coupled model with a seasonally varying background state is used operationally at the Australian Bureau of Meteorology Research Centre (BMRC) for seasonal forecasting (Kleeman 1989, 1991, 1993, 1994).

The atmospheric component of the model is a two-level, damped, steady-state Gill model that describes the anomalous first baroclinic mode circulation of the tropical atmosphere at 750 mb about the observed annual mean state (Kleeman 1991). The atmosphere is heated in two ways: (i) by direct thermal forcing (DTF) in the form of Newtonian relaxation, directly proportional to SST anomalies, which represents the effects of sensible heating, shallow convection, and radiation that are not explicitly parameterized in the model; and (ii) by latent heating (LH) due to deep penetrative convection in the atmosphere. DTF and LH are parameterized in the atmospheric continuity equation as follows:

$$\mathrm{DTF} = \epsilon_a \mathcal{R} T/2 \tag{1}$$

$$\begin{aligned} Q = \mathcal{R}/2\langle I_1\{|\mathbf{U}|[q_{\mathrm{diff}}(\overline{T}+T)] + |\overline{\mathbf{U}}|[q_{\mathrm{diff}}(\overline{T}+T) \\ - q_{\mathrm{diff}}(\overline{T})]\} \\ + I_2[(\overline{q}+q)\boldsymbol{\nabla}\mathbf{U} + q\boldsymbol{\nabla}\overline{\mathbf{U}} + (\overline{\mathbf{U}}+\mathbf{U})\boldsymbol{\nabla}q \\ + \mathbf{U}\boldsymbol{\nabla}\overline{q}]\rangle \end{aligned} \tag{2}$$

$$\mathrm{LH} = \begin{cases} Q & m(\overline{T}+T) \geq m_c \\ -\overline{Q} & m(\overline{T}+T) < m_c, \end{cases} \tag{3}$$

where T is the SST anomaly, $\mathbf{U}$ is the wind anomaly, q is the specific humidity, $q_{\mathrm{diff}}(\mathcal{T})$ is the air–sea specific humidity difference at temperature $\mathcal{T}$, $\mathcal{R}$ is the gas constant for dry air, and ϵ_a is a Newtonian cooling coefficient. An overbar denotes observed annual mean climatological quantities, while all other quantities represent anomalies. Equation (2) results from integrating the atmospheric moisture equation vertically, and I_1 and I_2 incorporate the constants of integration and various other physical constants (see Kleeman 1989). The first term in (2) in $\{\ldots\}$ represents latent heat release from the ocean surface due to convectively induced surface wind anomalies, while the second term in $[\ldots]$ represents latent heat release due to convectively induced moisture convergence in the atmosphere. The total specific humidity $(\overline{q}+q)$ increases exponentially with $(\overline{T}+T)$ according to the Clausius–Clapeyron relation. This has the effect of making the atmosphere more unstable to convection. This instability manifests itself in the coupled model as a moisture convergence feedback through the term $\mathcal{R}I_2(\overline{q}+q)\boldsymbol{\nabla}\mathbf{U}/2$ in (2). The atmospheric continuity equation can be written as $\epsilon_a\Phi + c_a^2(1-r)\boldsymbol{\nabla}\mathbf{U} = Q'$, where Φ is the geopotential height anomaly, c_a is the first baroclinic mode phase speed, $r = \mathcal{R}I_2(\overline{q}+q)/2c_a^2$, and $Q' = \mathrm{DTF} + \mathrm{LH} - c_a^2 r$. By Fourier decomposing this form of the continuity equation Zebiak (1986) showed that as $r \to 1$ [i.e., increasing $(\overline{T}+T)$], the atmosphere becomes more unstable to the heating Q', particularly for the smallest scales for which the wind divergence anomaly increases as $Q'/(1-r)$. The parameter r is also related to the ratio of the low-level moist static energy m and a quantity proportional to the mean difference in dry static energy between the lower and upper levels of the atmosphere (Kleeman 1989). Thus a stepwise nonlinearity (3) is used to ''trigger'' convective heating anomalies in the model when m of the boundary layer exceeds the critical value m_c, which corresponds to an SST of approximately 28.5°C, the observed threshold for deep penetrative convection (Graham and Barnett 1987). Equation (3) shows that SST perturbations T can either enhance existing deep convection when $\mathrm{LH} = Q$, or shut down existing convection when $\mathrm{LH} = -\overline{Q}$. This kind of behavior has been observed in atmospheric GCMs with imposed SST anomalies (e.g., Hoerling et al. 1997), and was also the basis for locating Tropical Ocean Global Comprehensive Ocean–Atmosphere Data Set in the west Pacific (Godfrey et al. 1998).

The ocean component of DYN describes the dynamics of the first baroclinic mode subject to the equatorial long-wave approximation. SST anomalies T evolve according to

$$\partial T/\partial t - \alpha h + \epsilon T = 0, \tag{4}$$

where h is the ocean thermocline depth anomaly, α is a constant of proportionality that varies with longitude, and ϵ is a Newtonian cooling coefficient. Equation (4) is solved only along the equator, and SST anomalies are assumed to have a Gaussian structure in the meridional direction with an e-folding scale of 10° lat, which corresponds closely with the atmospheric first baroclinic mode equatorial radius of deformation. This mimics the spreading effect that the trade-wind-induced ocean Ekman transport has on equatorial SST anomalies. Equation (4) describes the fact that vertical movements of the thermocline can create SST anomalies in the presence of equatorial upwelling along the equator. Different values are used for α in the west and east Pacific to reflect the fact that the main thermocline is deeper in the west than the east, and that in general the rate of oceanic upwelling is larger in the east and central Pacific than in the west Pacific. When the thermocline is very deep or very shallow further changes in depth of the thermocline do not influence SST anomalies. This is in accord with observations and is represented by a stepwise nonlinearity in the second term of (4).

Horizontal advection is also known to influence the development of SST anomalies (e.g., Picaut et al. 1996) but its effects are not included in this study in an effort to minimize the number of factors that influence the nonnormality of the coupled models. The absence of horizontal advection in (4) does not affect the ability of the coupled models to produce ENSO-like oscillations

(see Kleeman 1993; Moore and Kleeman 1997a,b, 1999b), and does not fundamentally alter the structure of the optimal perturbations of DYN (Moore and Kleeman 1999b).

The ability of the coupled model to support self-sustaining or damped oscillations depends on the phase speed of equatorial Kelvin waves in the ocean, c_o, and atmosphere, c_a, the coupling strength between the ocean and atmosphere γ, which enters as a mean wind speed in the linear drag law used to compute surface wind stress anomalies, the Newtonian cooling coefficients ϵ_a and ϵ, and the parameter α in (4). Unless otherwise stated the following parameter values were used: $c_o = 2.8$ m s^{-1}, $c_a = 60$ m s^{-1}, $\gamma = 10$ m s^{-1}, $\epsilon_a = (3$ days$)^{-1}$, and α varies between 6.8×10^{-9} °C m^{-1} s^{-1} in the west Pacific and 3.4×10^{-8} °C m^{-1} s^{-1} in the east. In this configuration the coupled model supports a self-sustaining oscillation with a period of approximately 3.5 yr with characteristics similar to those of the observed ENSO. Figure 1a shows a Hovmöller diagram of T along the equator for the oscillation. Principal oscillation pattern (POP) analysis reveals that the POP describing ENSO in DYN is very similar to the ENSO POP computed from observations (Kleeman and Moore 1999). The effect of the nonlinearity in (3) is apparent in Fig. 1 where it produces an asymmetry between the duration of warm and cold events although the amplitudes of warm and cold events are similar. The tangent linear and adjoint tangent linear versions of DYN were used to compute the optimal perturbations. The tangent linear version of DYN linearized about the observed annual mean state produces an ENSO oscillation with symmetric warm and cold phases of equal duration as shown in Fig. 1b. Unless otherwise indicated, in sequel all references to DYN refer to the tangent linear model.

b. The hybrid coupled models, HCM1 and HCM2

The term ''hybrid'' as used here refers to a coupled model composed of a statistical atmospheric model and a dynamical ocean model. While the term hybrid coupled model (HCM) is normally used to describe coupled models in which one component is a GCM, our use of the term here should not cause any confusion. The dynamical ocean model is identical to that described in section 2a. The statistical atmospheric model was constructed from a 100-yr time series of T and surface wind stress anomalies $\boldsymbol{\tau}$ from the self-sustaining oscillation of the nonlinear DYN (cf. Fig. 1a). The statistical atmospheric model was constructed using singular vector decomposition (SVD) as described by Bretherton et al. (1992), and more recently by Syu et al. (1995) with the objective of finding statistically significant covarying patterns of T and $\boldsymbol{\tau}$. The covarying patterns of T and $\boldsymbol{\tau}$ identified this way will be referred to as the SST SVDs and wind stress SVDs, respectively.

Figure 2 shows the structure of the first two members of the SST SVD and wind stress SVD spectra. SVD1 and SVD2 account for 99.9% of the covariance of SST and $\boldsymbol{\tau}$ in Fig. 1a and have the structure of two phases of the model ENSO oscillation in quadrature. The wind stress SVD1 and SVD2 are similar to the first 2 EOFs computed from observed $\boldsymbol{\tau}$ by Kleeman et al. (1992), and agree qualitatively with recent multivariate analyses of Xue et al. (2000).

The dynamical atmospheric component of DYN was replaced by a statistical model composed of the first two SVDs of T and $\boldsymbol{\tau}$, and the resulting HCM will be referred to as HCM1. A Hovmöller diagram of T along the equator from the tangent linear version of HCM1 is shown in Fig. 1c. HCM1 clearly captures the ENSO oscillation of the tangent linear DYN (Fig. 1b). The tangent linear ENSO oscillation is shown because this is the version of HCM1 used to compute optimal perturbations. However, the ENSO oscillation of HCM1 and its tangent linear version are very similar because the statistical atmosphere is linear, and because the thermocline nonlinearity in (4) does not influence the relative duration of warm and cold events.

The efficacy of HCM1 can be tested by comparing its ENSO hindcast skill to the skill of DYN. Hindcasts of 2-yr duration were started on the first of each month between January 1972 and December 1996 using nonlinear versions of DYN and HCM1 by first spinning up the ocean component of each model using observed Florida State University wind anomalies for the 2 yr prior to the hindcast start date, and then coupling the ocean and atmospheric models for the next 2 yr (Kleeman 1993).[1] Figure 3 shows the average anomaly correlation between the observed and hindcast Niño-3 index as a function of hindcast lag time for DYN, HCM1, and persistence forecasts. Clearly the hindcast skills of HCM1 and DYN are very similar.

The SVD analysis described above is the standard statistical approach for finding statistically significant covarying patterns of two fields. If the values of T and $\boldsymbol{\tau}$ are taken at face value then not suprisingly the SVDs that account for the largest fraction of SST variance resemble the east Pacific ENSO cycle as Fig. 2 shows because this is where the largest SST anomalies occur. However, the annual average temperture of the west Pacific warm pool is around 29°C while in the east Pacific it is only about 24°C. Therefore by virtue of the Clausius–Clapeyron relation and the moisture convergence feedback, a SST anomaly in the warm pool will have a larger impact on the atmosphere than an identical SST anomaly in the cold tongue [see also discussions by Mayer and Weisberg (1998) and Wang et al. (1999)]. This effect is naturally incorporated in nonlinear DYN by (1)–(3). Figure 4 shows the standard deviation σ of

[1] Operationally at BMRC subsurface thermal observations are assimilated during this stage (Kleeman et al. 1995), which improves the predictive skill of the model considerably. Data assimilation however was not included in the hindcasts presented here.

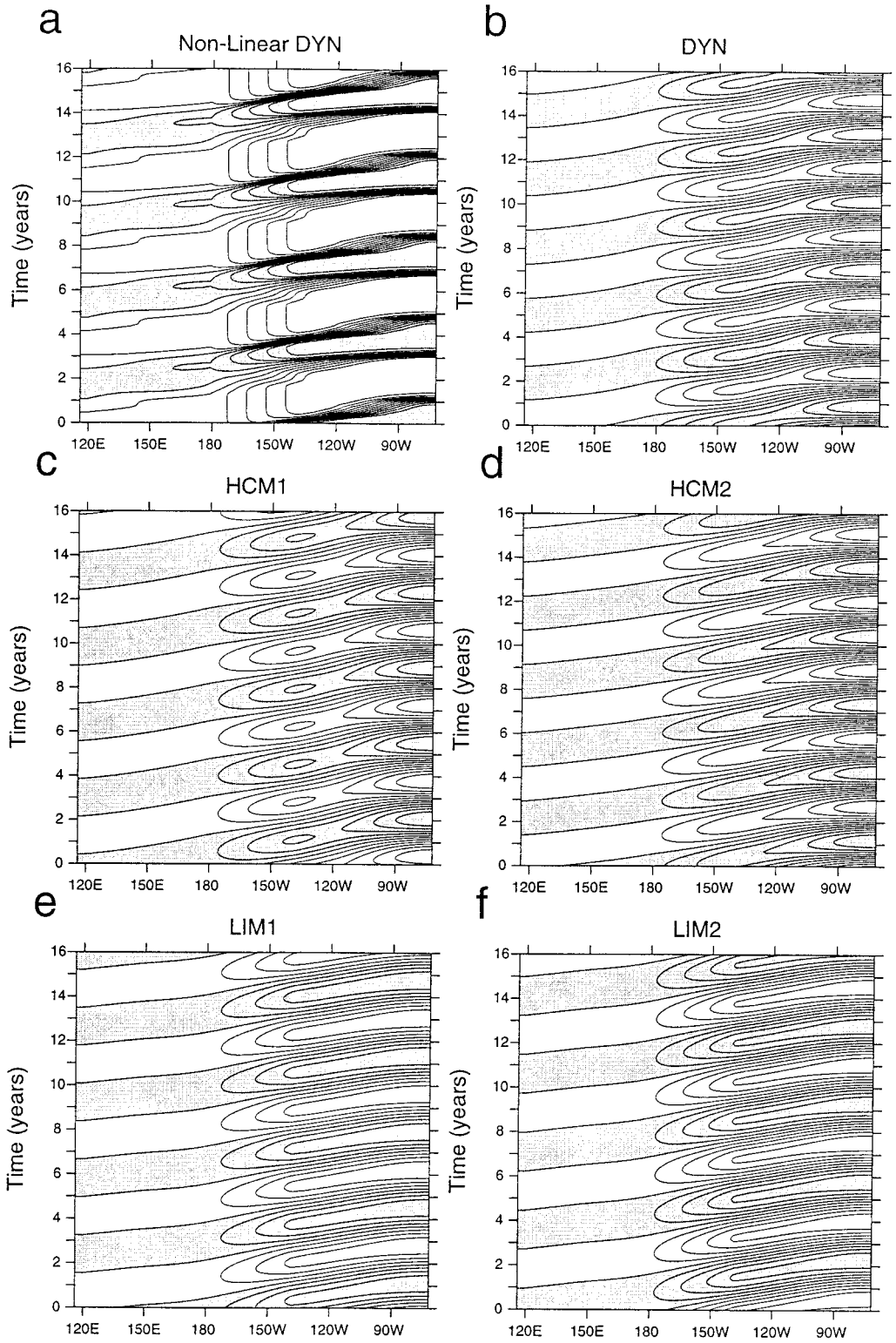


FIG. 1. Hovmöller diagrams of SST anomaly, *T*, for the ENSO mode from (a) the nonlinear DYN, and tangent linear versions of (b) DYN, (c) HCM1, (d) HCM2, (e) LIM1, and (f) LIM2. In (a) the contour interval is 0.5°C. In (b)–(f) the contour interval is arbitrary and the exponential growth/decay factor of the oscillation has been suppressed in each case. Shaded regions indicate anomalously warm SST.

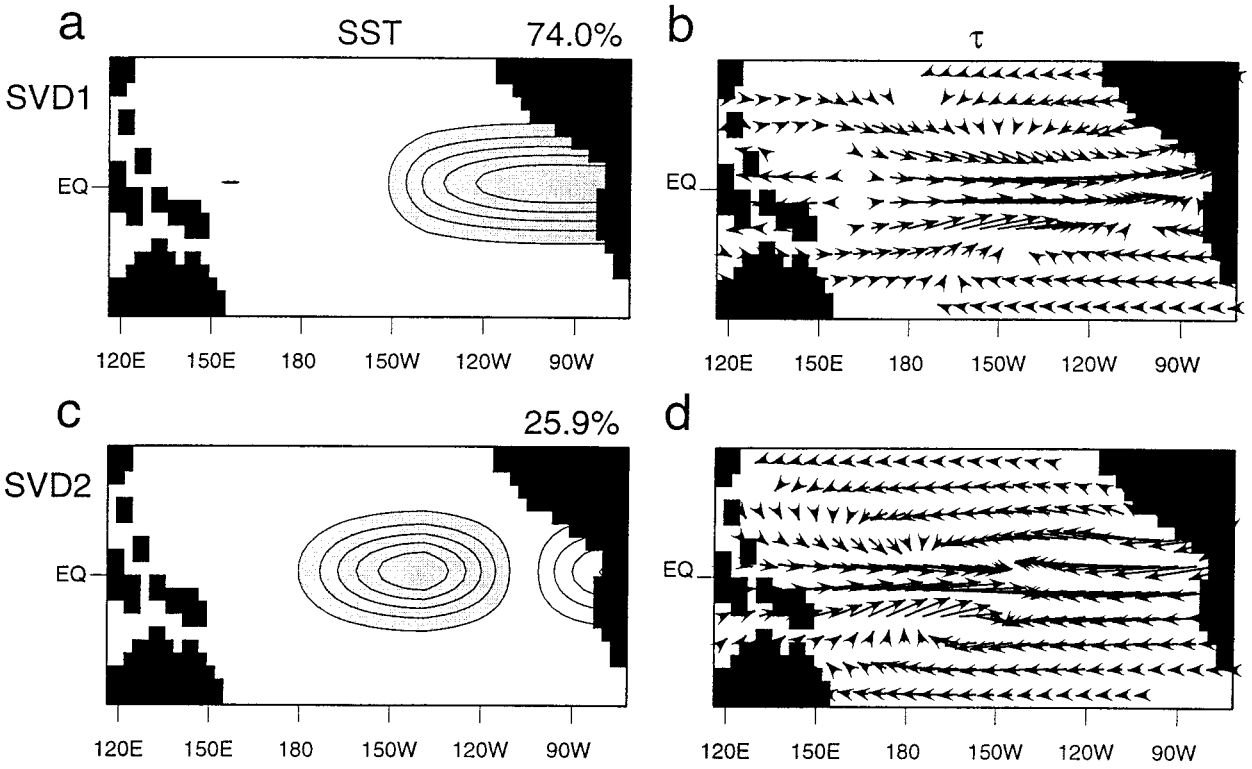


FIG. 2. The first and second SST and wind stress SVDs of the ENSO oscillation of nonlinear DYN. The contour interval and vector scaling is arbitrary, and the percentage explained covariance of T and $\boldsymbol{\tau}$ is indicated. Shaded regions represent warm SST anomalies.

SST along the equator from DYN and indicates that in the east Pacific SST anomalies are ~2.5°C while in the west Pacific they are only ~0.5°C. However, because of (3) the lower amplitude SST variability in the west Pacific will be as important as the larger amplitude variability in the east. The effects of this important nonlinearity, however, are effectively ignored in HCM1 since the smaller amplitude SST variability in the west is not captured by the dominant SVDs of Fig. 2 that are retained. The inverse of σ in Fig. 4 crudely mirrors the variation in mean SST along the equator. To allow for the effects of (3), therefore, a second HCM was constructed by scaling T by $\sigma(x)$ at each longitude x before the SVD analysis of T and $\boldsymbol{\tau}$ was performed. A SVD was performed using $\boldsymbol{\tau}$ and $\hat{T}(x) = T(x)/\sigma(x)$ where $\hat{T}$ is the "normalized SST." The first two members of the SST and wind stress SVD spectra in this case are shown in Fig. 5; they account for 99.7% of the covariance in $\hat{T}$ and $\boldsymbol{\tau}$. Figure 5a reveals that the west Pacific is now favored as well as the east by SST SVD1. The resulting HCM that retains the two SVDs of Fig. 5 will be referred to as HCM2. The ENSO oscillation of the tangent linear HCM2 is shown in Fig. 1d and is very similar to that

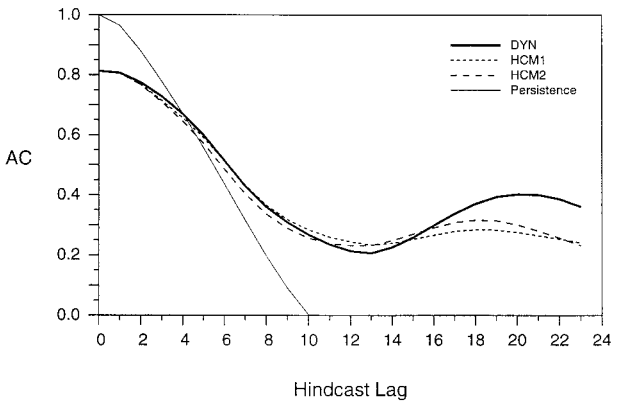


FIG. 3. The average anomaly correlation hindcast skill score (AC) for Niño-3 index as a function of hindcast lag time for DYN, HCM1, HCM2, and persistence for the period Jan 1972–Dec 1998.

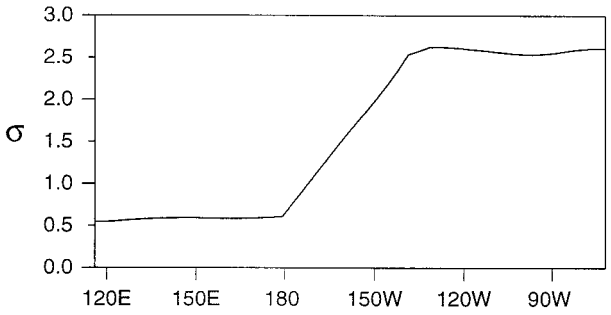


FIG. 4. The zonal variation along the equator of the standard deviation, σ, of SST anomaly for the ENSO oscillation of nonlinear DYN.

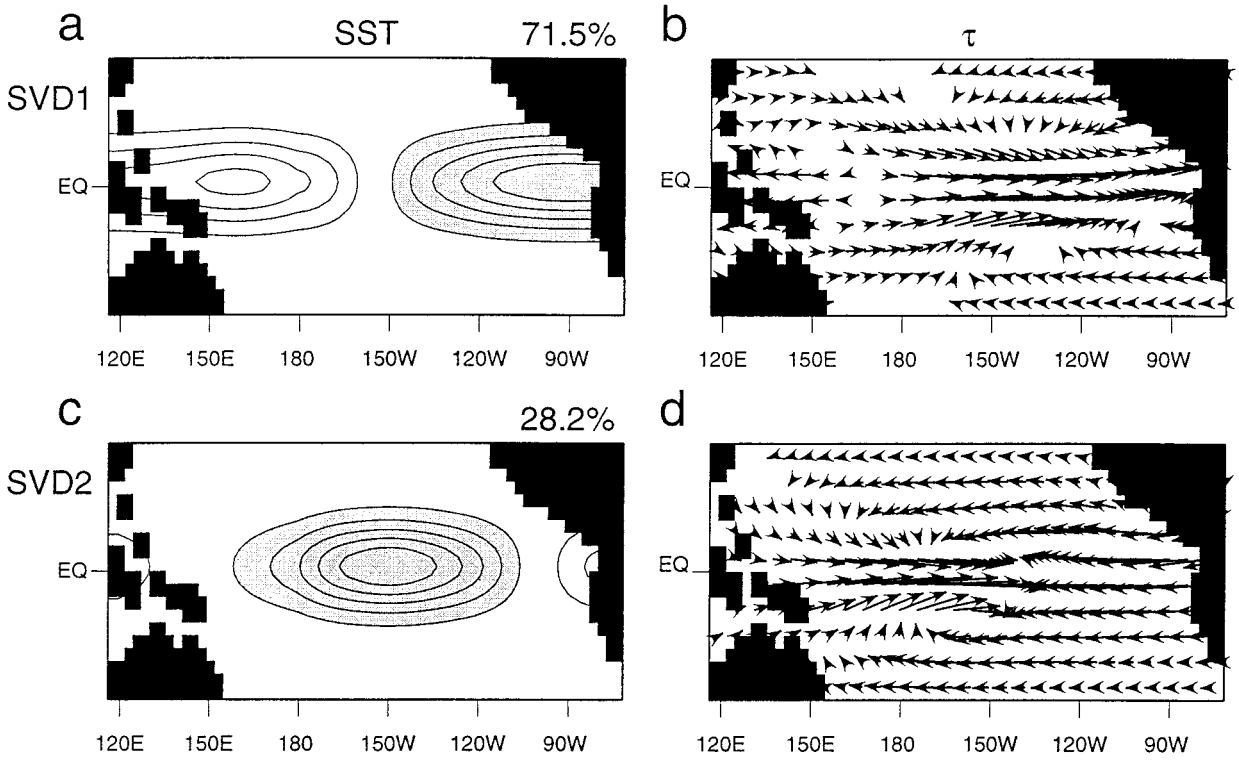


FIG. 5. Same as in Fig. 2 but SST is normalized by its standard deviation at each longitude.

of HCM1. Figure 3 shows that HCM2 hindcasts of ENSO for 1972–96 are as skillful as those of DYN and HCM1.

c. The linear inverse models, LIM1 and LIM2

As noted in section 1 some optimal perturbation calculations reported in the literature have been performed using linear inverse models (LIMs). These are either statistical reconstructions of the propagator of dynamical coupled models (Blumenthal 1991; Xue et al. 1994, 1997a,b; Fan 1998; Fan et al. 2000; Thompson and Battisti 2001), or propagator estimates of the real system using observational data (Penland and Sardeshmukh 1995; Y. Xue 1999, personal communication; Johnson et al. 1999, unpublished manuscript). In this section we will describe an LIM of DYN closely following the methodology described in detail by Penland and Sardeshmukh (1995), hereafter PS95. PS95 argued that the SST anomalies in the tropical Pacific and Indian Oceans can be modeled as a linear system driven by Gaussian white noise so that

$$d\mathbf{T}/dt = \mathbf{B}\mathbf{T} + \boldsymbol{\xi}, \tag{5}$$

where $\mathbf{T}$ denotes the vector of tropical SSTs, $\mathbf{B}$ is the matrix that controls the SST dynamics, and $\boldsymbol{\xi}$ is a white noise forcing representing the effects of the nonlinearity and external forcing. Following Penland (1989, 1996) the matrix $\mathbf{B}$ can be estimated from a time series of $\mathbf{T}(t)$, and is given by

$$\mathbf{B} = \ln[\mathbf{C}(t_2 - t_1)\mathbf{C}(0)^{-1}]/(t_2 - t_1), \tag{6}$$

where $\mathbf{C}(t_2 - t_1)$ and $\mathbf{C}(0)$ are the covariance matrices of $\mathbf{T}(t)$ at lag $(t_2 - t_1)$ and lag 0, respectively. Alternatively, a propagator $\mathbf{D}$ that advances the anomalies forward in by time by 1 month may be estimated, and for an autonomous system $\mathbf{B} = \mathbf{D}^n$, where $t_2 - t_1 = n$ months. Since the eigenmodes of $\mathbf{D}$ and $\mathbf{B}$ are identical for an autonomous system, these two approaches produce identical results in the present study since the annual average observed fields are used as the basic state.

Using $\mathbf{T}(t)$ sampled every month from a 100-yr integration of nonlinear DYN (cf. Fig. 1a), $\mathbf{B}$ was estimated using (6). First, anticipating the importance of (3) as discussed in section 2b, the normalized SST anomalies $\hat{\mathbf{T}}(t) = \mathbf{W}^{-1}\mathbf{T}(t)$ were computed. Two forms of $\mathbf{W}$ were used: $\mathbf{W} = \mathbf{I}$, which corresponds to the original PS95 formulation, and $\mathbf{W} = \text{diag}[\sigma(x)]$. Second, to reduce the dimension of the problem, the EOFs of $\hat{\mathbf{T}}(t)$ were computed and only the first two EOFs were retained, which account for 99.3% (93.1%) of the covariance in SST and $\boldsymbol{\tau}$, when $\mathbf{W} = \mathbf{I}$ $\{\mathbf{W} = \text{diag}[\sigma(x)]\}$. Equation (5) was then formulated in terms of the coefficients of the retained EOFs. The vector of EOF amplitudes $\mathbf{p}(t)$ evolves according to $d\mathbf{p}/dt = \mathbf{M}\mathbf{p}$ where $\mathbf{M}$ is given by (6) if $\mathbf{C}$ is the covariance matrix of the EOF coefficients. If $\mathbf{E}$ is the matrix whose columns are the retained EOFs of $\hat{\mathbf{T}}$, then the reconstruction $\tilde{\mathbf{T}}$ of $\mathbf{T}$ in physical space is given by $\tilde{\mathbf{T}}(t) = \mathbf{W}\mathbf{E}\mathbf{p}(t)$. Therefore,

$$\tilde{\mathbf{T}}(t_2) = \mathbf{WEHE}^{\mathrm{T}}\mathbf{W}^{-1}\tilde{\mathbf{T}}(t_1) = \exp[\tilde{\mathbf{B}}(t_2 - t_1)]\tilde{\mathbf{T}}(t_1), \quad (7)$$

where $\tilde{\mathbf{B}}$ is the estimate of $\mathbf{B}$ that results from the retained EOFs, and $\mathbf{H} = \exp[\mathbf{M}(t_2 - t_1)]$ is the propagator of $\mathbf{p}$. Unless otherwise stated $t_2 - t_1 = 3$ months, although the results presented are insensitive to this choice of time interval. If $\mathbf{W} = \mathbf{I}$ in (7) the resulting LIM is referred to as LIM1. Figure 1e shows the ENSO oscillation that results in LIM1, which is clearly very similar to that of the tangent linear DYN, HCM1, and HCM2. Choosing $\mathbf{W} = \mathrm{diag}[\sigma(x)]$ yields LIM2, and Fig. 1f shows that the ENSO oscillation of LIM2 is very similar to that of the other models.

3. Optimal perturbations

By definition (Farrell 1982), the optimal perturbations of each coupled model are the eigenvectors of $\mathbf{R}^{\mathrm{T}}\mathbf{R}$, where $\mathbf{R}$ and $\mathbf{R}^{\mathrm{T}}$ are the propagators of each tangent linear coupled model and their adjoints, respectively. The perturbation growth norm used in this section is SST^2 integrated over the entire tropical Pacific since this is the norm used in most previous optimal perturbation studies discussed in section 1. Other norms are considered in section 5. Because of differences in the formulation of each tangent linear model, the form of the state vector requires consideration. For DYN the state vector is composed of $\mathbf{T}$ and $\mathbf{q}$, where $\mathbf{q}$ is the vector of parabolic cylinder function amplitudes (see Kleeman 1993; Moore and Kleeman 1996). Since the SST^2 norm involves only $\mathbf{T}$, we will assume that at initial time $\mathbf{q} = 0$, which amounts to an optimal perturbation with initially unperturbed ocean thermocline depth and ocean currents. Thus for the SST^2 norm we need only consider a state vector composed of $\mathbf{T}$. We will consider two representations of state space. The first, described here and in section 4a, is based on statistical analyses and is defined in terms of the SVDs/EOFs of the system. This is particularly useful for understanding the behavior of the HCMs and LIMs. In this case it is instructive to express $\mathbf{T}(t)$ as $\mathbf{T}_S(t) + \mathbf{T}_R(t)$ where $\mathbf{T}_S(t)$ is the component of $\mathbf{T}(t)$ that projects only onto the subspace occupied by the retained SVDs/EOFs of unnormalized SST, and $\mathbf{T}_R(t)$ is the component of $\mathbf{T}(t)$ that projects onto the subspace not spanned by the retained SVDs/EOFs. In the HCMs and LIMs of sections 2b and 2c, only two SVDs and two EOFs were retained that describe the ENSO oscillation, so in this case $\mathbf{T}_S$ can also be interpreted as the component of $\mathbf{T}$ that projects onto ENSO. In the HCMs and LIMs we can therefore choose the state vector to be either (i) $\mathbf{T}(0) = \mathbf{T}_S(0) + \mathbf{T}_R(0)$, or (ii) the vector $\mathbf{T}_S(0)$ of retained SST SVD/EOF amplitudes only, which amounts to choosing $\mathbf{T}_R(0) = 0$. As we will demonstrate, the choice of state vector for HCM2 and LIM2 can have a significant impact on the structure of the optimal perturbations. The second representation of state space will be introduced in section 4b and is based on the eigenmodes of the linear propagator $\mathbf{R}$.

An optimal growth time of 3 months (i.e., one season) is used in all calculations described here. In all of the optimal perturbation calculations reported in the literature the optimal perturbation spectrum was found to be dominated by one singular vector. The same is true of all models used in this study so in the following we confine our attention to the fastest growing optimal perturbation of the SST^2 norm.

Fields of T, h, and $\boldsymbol{\tau}$ for the optimal perturbation of DYN are shown in Fig. 6. Initially the optimal perturbation takes the form of an SST dipole centered near the date line in the western Pacific. The structure and dynamics of this optimal perturbation are discussed in Moore and Kleeman (1997a,b). As noted by these authors the optimal perturbation evolves into a mature ENSO episode. The SST^2 norm growth factor μ is also indicated in Fig. 6.

The optimal perturbation of HCM1 using $\mathbf{T}$ as the state vector is shown in Fig. 7. The optimal perturbation that results when using only $\mathbf{T}_S$ as the state vector [i.e., $\mathbf{T}_R(0) = 0$] is identical to Fig. 7. Figures 6 and 7 reveal that HCM1 does not recover the optimal perturbation SST structure of DYN. In HCM1 both the initial and final optimal perturbation SST structures resemble a mature ENSO episode. The SST^2 norm growth factor μ is also indicated in Fig. 7, and shows that the perturbation grows faster in HCM1 than the optimal perturbation of DYN. However, the true test of optimal growth is how fast the optimal perturbation of HCM1 can grow in the tangent linear version of DYN since HCM1 is an approximation of DYN. In Fig. 7 (and all subsequent figures relating to HCMs and LIMs), the SST^2 norm growth factor μ of the HCM1 optimal perturbation as it evolves in DYN is shown in parentheses. Figure 7 indicates that the optimal perturbation of HCM1 actually decays in time when evolved in DYN, and is clearly a suboptimal perturbation of the fully dynamical coupled model.

The optimal perturbations of HCM2 are shown in Figs. 8a–d using $\mathbf{T}$ as the state vector and in Figs. 8e–h using only $\mathbf{T}_S$ as the state vector. Figures 8 and 6 reveal that HCM2 is able to recover most of the large-scale initial SST structure of the optimal perturbation of DYN, with largest amplitude in the western Pacific when $\mathbf{T}$ is used as the state vector. When only $\mathbf{T}_S$ is used as the state vector the optimal perturbation of HCM2 is similar to that of HCM1. Irrespective of the state vector used, Fig. 8 reveals that HCM2 overestimates the growth factor μ compared to the true optimal perturbation in Fig. 6. However, the optimal perturbation of Fig. 8a gives a μ close to the true value when evolved by DYN, while the perturbation in Fig. 8e is clearly suboptimal in this regard.

Figure 9 shows the initial and final SST structure of the optimal perturbation of LIM1 using $\mathbf{T}$ as the state vector. When only $\mathbf{T}_S$ is used as the state vector the

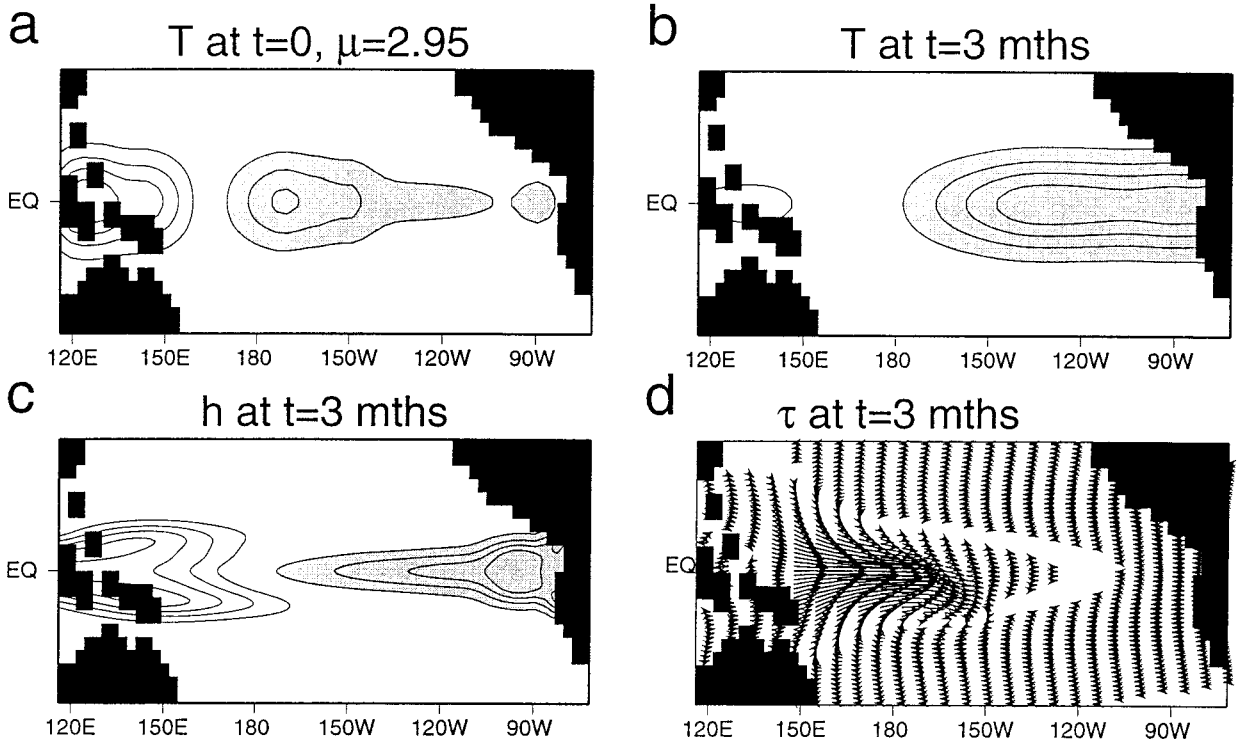


FIG. 6. The (a) initial perturbation SST, (b) final perturbation SST, (c) final thermocline depth perturbation, and (d) final perturbation wind stress for the optimal perturbation of DYN. The contour interval and vector scaling is arbitrary. Shaded regions represent warm SST perturbations and an anomalously deep thermocline. The SST^2 norm growth factor μ is indicated.

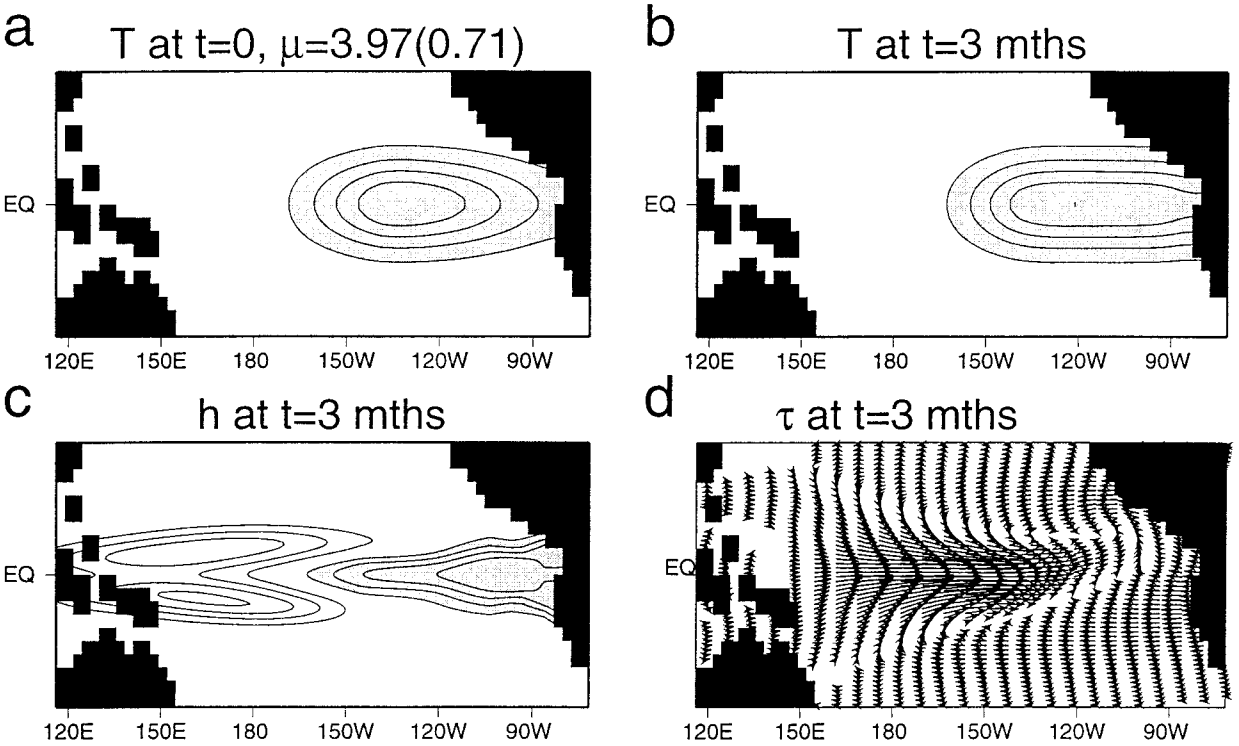


FIG. 7. Same as in Fig. 6 but for HCM1. The SST^2 norm growth factor μ for the perturbation evolved in HCM1 (DYN) is indicated.

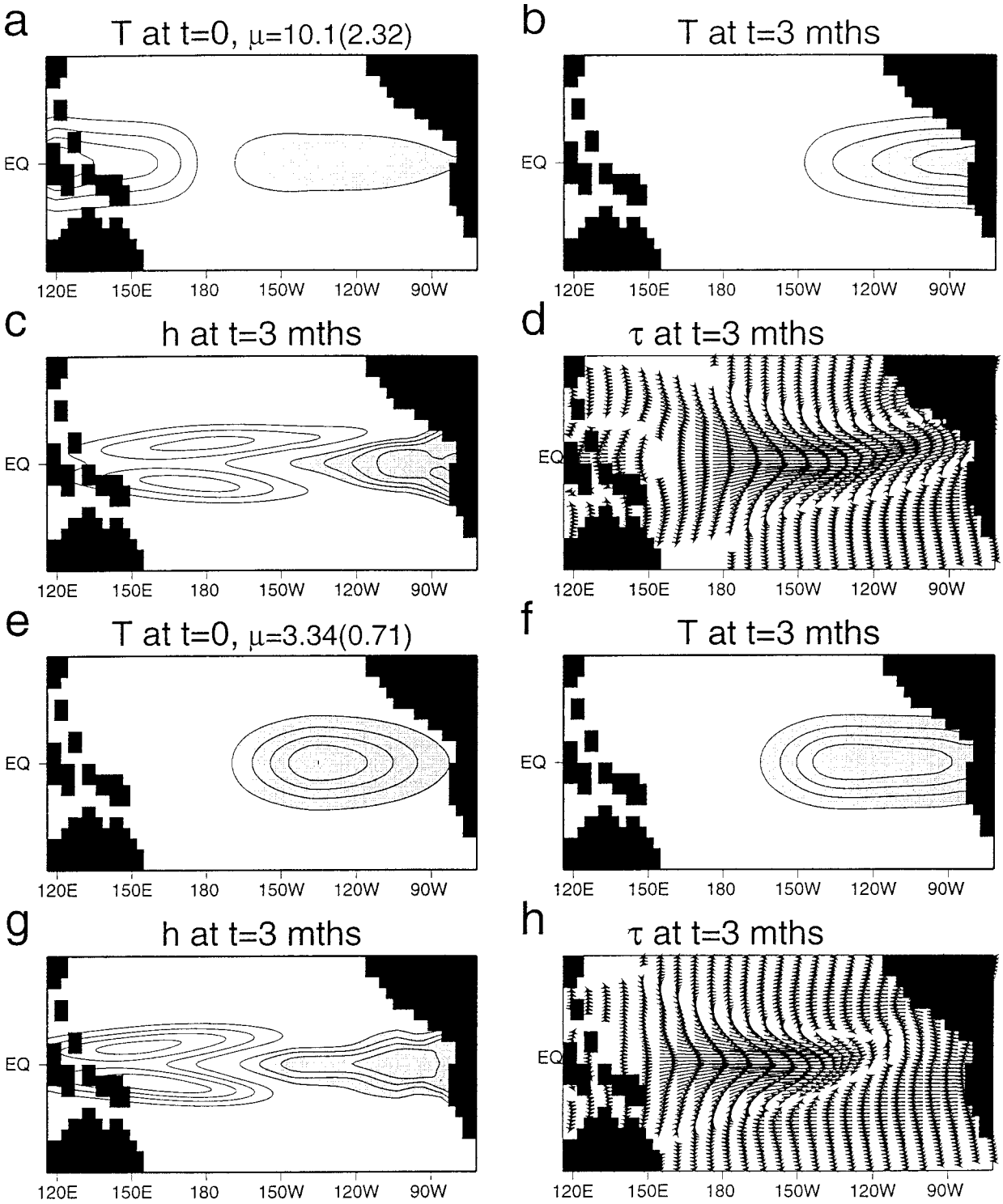


FIG. 8. Same as in Fig. 6 but for HCM2 (a)–(d) using **T** and (e)–(h) $\mathbf{T}_S$ as the state vector.

optimal perturbation of LIM1 is identical to Fig. 9. Like HCM1, LIM1 is constructed without regard for the effects of (3) on the system and is unable to recover the optimal perturbation of DYN. The optimal perturbation of LIM1 is clearly suboptimal when evolved in DYN as μ indicates. Figures 10a,b show the optimal perturbation of LIM2 using **T** as the state vector, which recovers the large-scale initial SST structure of the op-

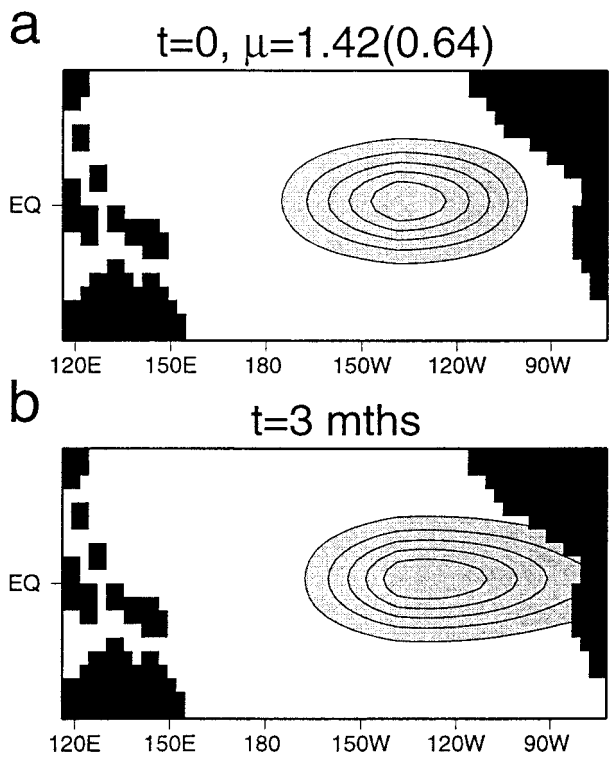


FIG. 9. The (a) initial and (b) final patterns of perturbation SST for the optimal perturbation of LIM1. The contour interval is arbitrary, and shaded regions represent warm SST perturbations. The SST^2 norm growth factor μ for the perturbation evolved in LIM1 (DYN) is indicated.

timal perturbation of DYN. The growth factor μ of the LIM2 optimal perturbation when evolved by DYN is close to that of the true optimal in Fig. 6. When $\mathbf{T}_S$ is used as the state vector Figs. 10c,d show that the optimal perturbation of LIM2 is similar to that of LIM1.

The effect on the optimal perturbation of retaining more SVDs in the HCM is illustrated in Fig. 11, which shows the initial optimal perturbation SST structure of HCM1 and HCM2 when the first 4 and 12 SVDs are used and $\mathbf{T}$ is the state vector. For HCM1 (HCM2), SVDs $3 \rightarrow 4$ and $3 \rightarrow 12$ account for 0.110% (0.291%) and 0.112% (0.297%), respectively, of the covariance of T and $\boldsymbol{\tau}$. As the number of SVDs increases, HCM1 does not recover the optimal perturbation of DYN. However the addition of more SVDs to HCM2 yields a better representation of the DYN optimal perturbation. This is true only up to a point, however, and when the number of SVDs increases beyond 20 or so the optimal perturbations in HCM1 and HCM2 become dominated by unphysical small scales. However, all optimal perturbations of HCM1 and HCM2 are suboptimal when they are evolved in DYN. The optimal perturbations of LIM1 and LIM2 as the number of EOFS retained in $\mathbf{E}$ is increased are shown in Figs. 11e–h, and exhibit the same general behavior as those of the HCMs. For LIM1 (LIM2), EOFs $3 \rightarrow 4$ and $3 \rightarrow 12$ account for 3.1% (5.5%) and 3.7% (6.9%), respectively, of the variance in T. Notice the enormous growth factors μ attained by the optimal perturbations of the LIMs when 12 EOFs are retained due to the fact that the small-scale EOFs help to conceal ENSO. However, these perturbations are extremely suboptimal and decay very rapidly when evolved in DYN.

To demonstrate the influence of (3) on the optimal perturbations of DYN, Fig. 12a shows the optimal perturbation initial SST of DYN when DTF $\neq 0$ and LH $= 0$ in (3) everywhere, which removes the nonlinear effect of the warm pool on the atmosphere. As shown by Moore and Kleeman (1999b), LH acts to destabilize the coupled system, so the coupling strength γ was increased to 20 m s^{-1} to yield a self-sustaining oscillation. In this case the optimal perturbation more closely resembles that of HCM1 and LIM1. Figure 12b shows the optimal perturbation initial SST for the case when DTF $= 0$ in (1) and LH $\neq 0$, with $\gamma = 20$ m s^{-1}—the SST dipole of DYN in the west Pacific is recovered (cf. Fig. 6a) and is clearly associated with LH.

The results above suggest that weighting the SST by its standard deviation in the computation of the SVDs for HCM2, or for estimating $\tilde{\mathbf{B}}$ in LIM2, can mimic the effect that the tangent linear form of (3) has on the latent heating of the atmosphere over the warm pool in DYN, provided that the state vector projects onto the entire space of $\mathbf{T}$ spanned by both $\mathbf{T}_S$ and $\mathbf{T}_R$. To further demonstrate the importance of allowing for the effects of (3) in the construction of the HCMs and LIMs, HCM1 and LIM1 were reconstructed using time series of T and $\boldsymbol{\tau}$ from a version of nonlinear DYN with LH $= 0$ and $\gamma = 20$ m s^{-1}. Thus the effects of the nonlinearity (3) on SST were removed from the system. In this case an ENSO oscillation similar to that in Figs. 1b–f results (not shown). The optimal perturbations of the new HCM1 and LIM1 are shown in Figs. 12c and 12d, respectively, and are a reasonable representation of the corresponding optimal perturbation of DYN shown in Fig. 12a.

4. Optimal perturbation structure: State space considerations and nonnormality

In this section we will demonstrate that the difference in optimal perturbation structure between the HCMs and LIMs with unnormalized and normalized SST is related to the differing degrees of nonnormality of the two models.

a. The hybrid and linear inverse models

The results of section 3 show that the contribution of $\mathbf{T}_R(0)$ to the optimal perturbation of HCM2 and LIM2 is crucial for recovering the optimal perturbation structure of DYN. The role played by $\mathbf{T}_R$ in controling optimal perturbation growth can be clarified by the following analysis. We will denote by $\mathbf{T}(t)$ the SST com-

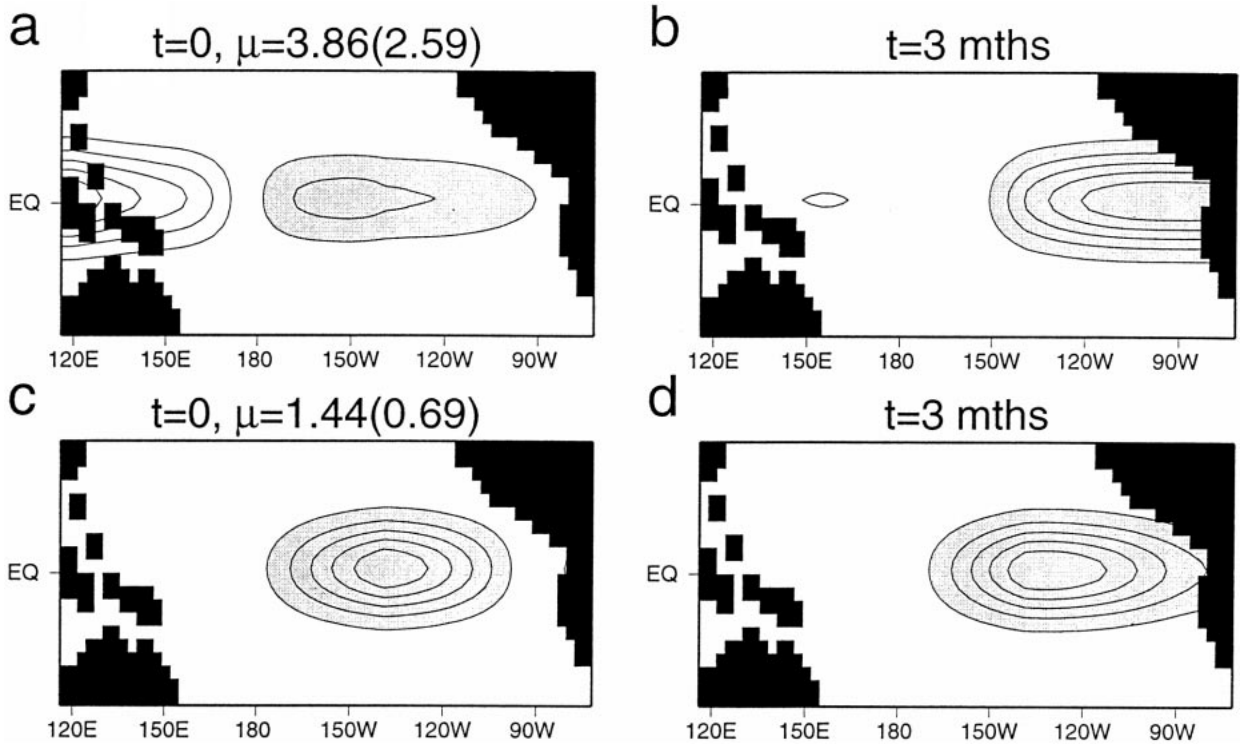


FIG. 10. Same as in Fig. 9 but for LIM2 using (a),(b) $\mathbf{T}$ and (c),(d) $\mathbf{T}_S$ as the state vector.

ponent of the state vector of the optimal perturbation, and following section 3 let $\mathbf{T}(0) = \mathbf{T}_S(0) + \mathbf{T}_R(0)$ where $\mathbf{T}_S(0) = \sum_{i=1}^{2} s_i \mathbf{t}_i$ is the component of $\mathbf{T}$ in the HCMs and LIMs that projects onto the SST SVDs/EOFs denoted $\mathbf{t}_i$, and $\mathbf{T}_R(0)$ is a residual term that does not span the space occupied by $\mathbf{t}_1$ and $\mathbf{t}_2$ so that $\mathbf{T}_S^{\mathrm{T}}(0) \cdot \mathbf{T}_R(0) = 0$.

In HCM1 and LIM1 the SST SVDs/EOFs are orthogonal and normalized so that $\mathbf{t}_i^{\mathrm{T}} \cdot \mathbf{t}_j = \delta_{ij}$. The growth factor $\mu(t)$ of the optimal perturbation SST^2 norm over the time interval t is given by

$$\mu(t) = \frac{\mathbf{T}^{\mathrm{T}}(t) \cdot \mathbf{T}(t)}{\mathbf{T}^{\mathrm{T}}(0) \cdot \mathbf{T}(0)} = \frac{\mathbf{T}^{\mathrm{T}}(0)\mathbf{R}^{\mathrm{T}} \cdot \mathbf{R}\mathbf{T}(0)}{\mathbf{T}^{\mathrm{T}}(0) \cdot \mathbf{T}(0)}, \qquad (8)$$

where as before $\mathbf{R}$ is the tangent linear propagator. Equation (8) can be reexpressed as

$$\mu(t) = \frac{\mathbf{T}_S^{\mathrm{T}}(0)\mathbf{R}^{\mathrm{T}} \cdot \mathbf{R}\mathbf{T}_S(0) + \mathbf{T}_S^{\mathrm{T}}(0)\mathbf{R}^{\mathrm{T}} \cdot \mathbf{R}\mathbf{T}_R(0) + \mathbf{T}_R^{\mathrm{T}}(0)\mathbf{R}^{\mathrm{T}} \cdot \mathbf{R}\mathbf{T}_S(0) + \mathbf{T}_R^{\mathrm{T}}(0)\mathbf{R}^{\mathrm{T}} \cdot \mathbf{R}\mathbf{T}_R(0)}{\mathbf{T}_S^{\mathrm{T}}(0) \cdot \mathbf{T}_S(0) + \mathbf{T}_R^{\mathrm{T}}(0) \cdot \mathbf{T}_R(0)}. \qquad (9)$$

1) LIM1 AND HCM1

It is instructive to consider first the growth factor $\mu(t)$ in the context of LIM1 and HCM1 since (9) reduces to a common form for both models. For LIM1 (and LIM2), (7) shows that $\mathbf{T}_R(t) = \mathbf{R}\mathbf{T}_R(0) = \mathbf{W}\mathbf{E}\mathbf{H}\mathbf{E}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{T}_R(0) = 0$. Therefore even if an initial perturbation $\mathbf{T}(0)$ spans the entire space defined by $\mathbf{T}_S$ and $\mathbf{T}_R$, at later times $\mathbf{T}(t)$ will occupy only the space spanned by the retained EOFs of $\mathbf{T}_S$. Thus (9) reduces to

$$\mu(t) = \frac{\mathbf{T}_S^{\mathrm{T}}(t) \cdot \mathbf{T}_S(t)}{\mathbf{T}_S^{\mathrm{T}}(0) \cdot \mathbf{T}_S(0) + \mathbf{T}_R^{\mathrm{T}}(0) \cdot \mathbf{T}_R(0)} = \frac{A}{1 + B}, \qquad (10)$$

where $A = \mathbf{T}_S^{\mathrm{T}}(t) \cdot \mathbf{T}_S(t)/\mathbf{T}_S^{\mathrm{T}}(0) \cdot \mathbf{T}_S(0)$ and $B = \mathbf{T}_R^{\mathrm{T}}(0) \cdot \mathbf{T}_R(0)/\mathbf{T}_S^{\mathrm{T}}(0) \cdot \mathbf{T}_S(0)$.

Consider now HCM1 and let $\mathbf{G}$ represent the autonomous propagator that advances the tangent linear solution forward Δt, where $t = N\Delta t$. Using the multiplicative property of autonomous propagators we have $\mathbf{R} = \mathbf{G}^N$. The SST SVDs and wind stress SVDs used in HCM1 are not eigensolutions of its propagator $\mathbf{R}$, so over time part of $\mathbf{T}_S$ will map into the space spanned by $\mathbf{T}_R$. Since $\mathbf{T}_R$ has no projection on $\mathbf{t}_1$ and $\mathbf{t}_2$ it produces no wind field and hence no ocean thermocline perturbation h. Thus in general from (4) $\mathbf{G}^n\mathbf{T}_S(0) = \mathbf{T}_S(n\Delta t) + \sum_{i=1}^{n} \mathbf{T}_R(i\Delta t)e^{-\epsilon(n-i)\Delta t}$ where $\mathbf{T}_S^{\mathrm{T}}(i\Delta t) \cdot \mathbf{T}_R(i\Delta t) = 0$.

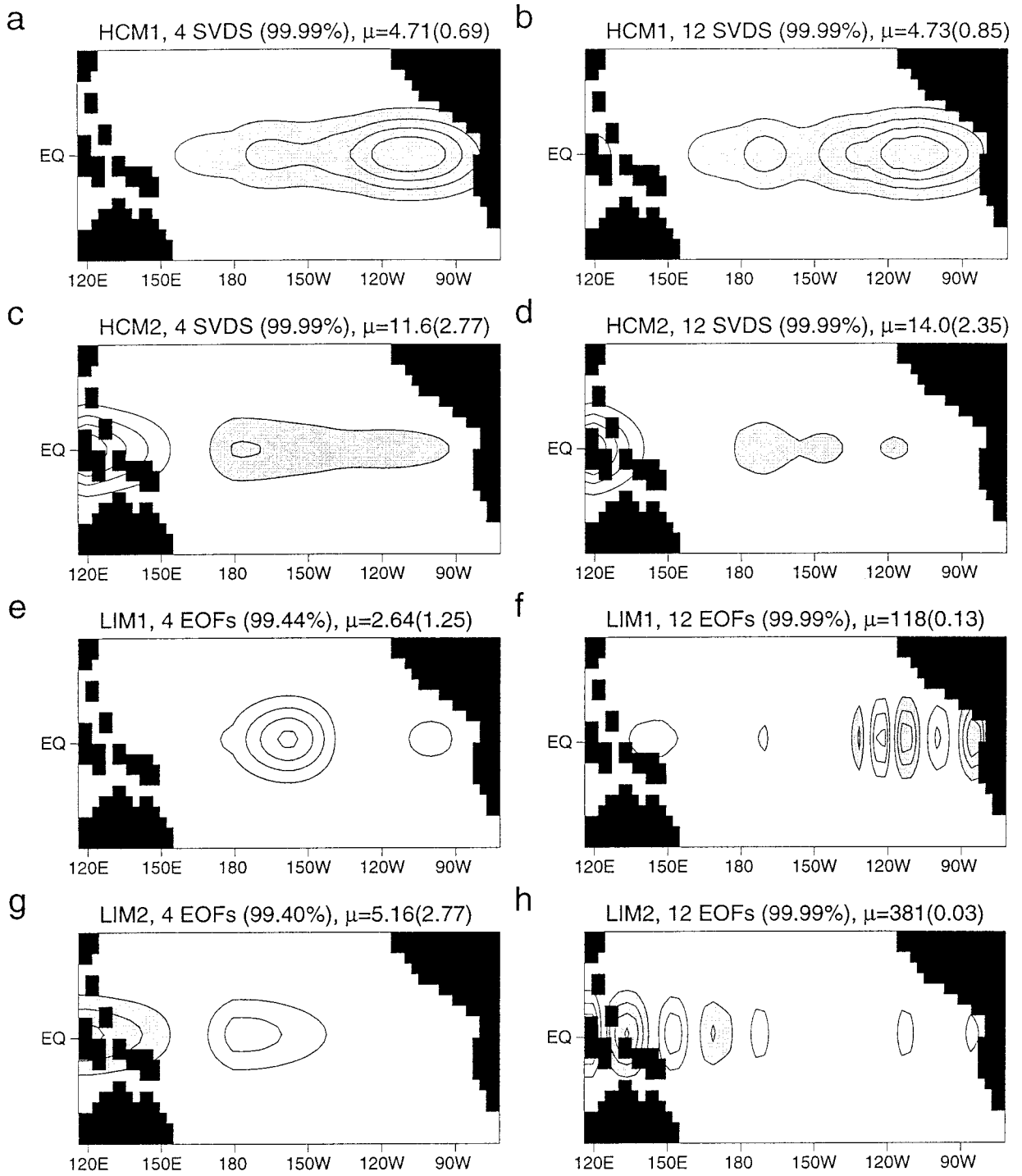


FIG. 11. The initial perturbation SST of the optimal perturbations for HCM1 and HCM2 with (a), (c) 4 SVDs and (b), (d) 12 SVDs; and LIM1 and LIM2 with (e), (f) 4 EOFs and (f), (h) 12 EOFs. The contour interval is arbitrary, and shaded regions represent warm SST perturbations. The percentage explained covariance of SST and τ of the retained SVDs is indicated in (a)–(d), while in (e)–(h) the percentage explained variance of SST of the retained EOFs is shown. The SST^2 norm growth factor μ of each perturbation evolved in each model and DYN (value in parentheses) is indicated.

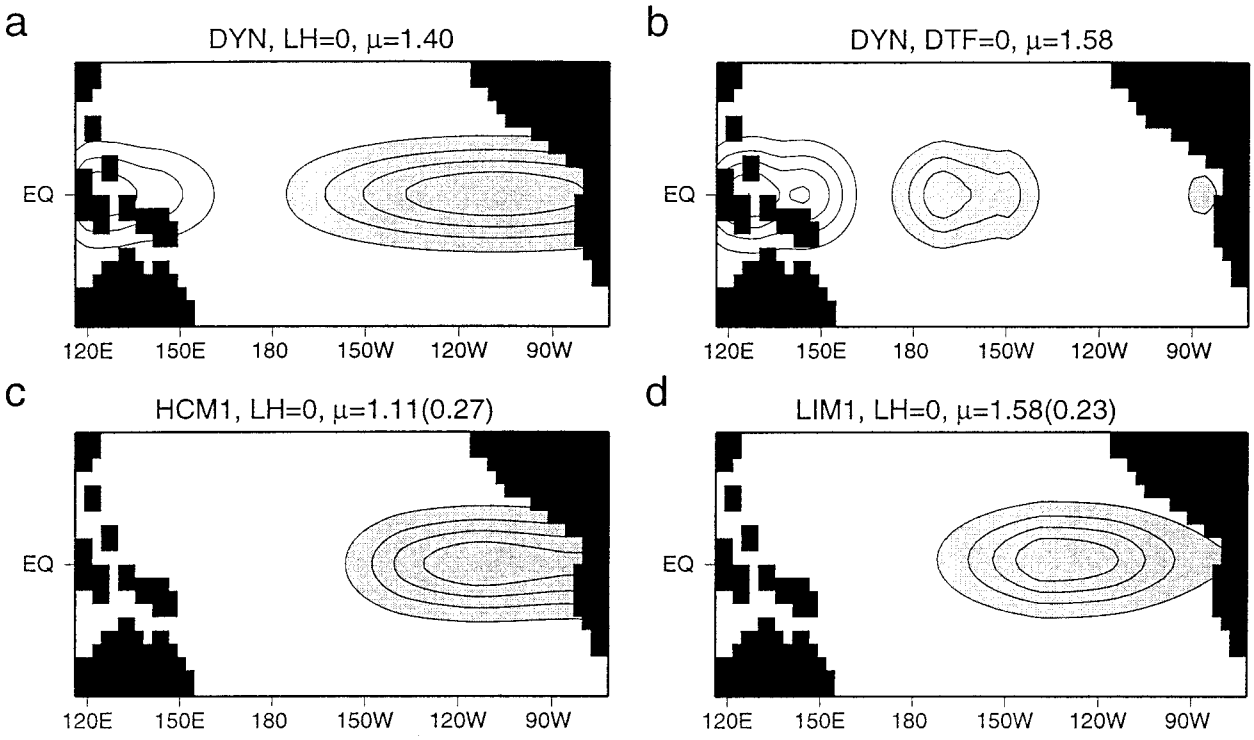


FIG. 12. The initial perturbation SST for the optimal perturbations of DYN when (a) DTF $\neq$ 0, LH = 0, (b) DTF = 0, LH $\neq$ 0. The initial perturbation SST for the optimal perturbation of (c) HCM1 and (d) LIM1 when DTF $\neq$ 0 and LH = 0. The contour interval is arbitrary, and shaded regions represent warm SST perturbations.

Similarly according to (4) $\mathbf{RT}_R(0) = \mathbf{T}_R(0)e^{-\epsilon N\Delta t}$ and, after some manipulation, we have $[\mathbf{RT}_S(0)]^{\mathrm{T}} \cdot [\mathbf{RT}_R(0)] = \sum_{i=1}^{N} \mathbf{T}_R(i\Delta t)^{\mathrm{T}}\, \mathbf{T}_R(0)e^{-\epsilon(2N-i)\Delta t}$. For $N\Delta t \sim 3$ months, $e^{-\epsilon N\Delta t} \sim 0.07$ so assuming that $\mathbf{T}_R^{\mathrm{T}}(i\Delta t) \cdot \mathbf{T}_R(0) \leq \mathbf{T}_S^{\mathrm{T}} \cdot \mathbf{T}_S$, a reasonable approximation since $\mathbf{T}_R(i\Delta t)$ is typically small, then $[\mathbf{RT}_S(0)]^{\mathrm{T}} \cdot [\mathbf{RT}_R(0)]$ and $[\mathbf{RT}_R(0)]^{\mathrm{T}} \cdot [\mathbf{RT}_S(0)]$ in (9) are negligible. Similarly, $\mathbf{T}_R^{\mathrm{T}}(t) \cdot \mathbf{T}_R(t) = \mathbf{T}_R(0)e^{-2N\epsilon t}$ is negligible and (9) for HCM1 reduces to (10).

Consider now the conditions that must be met by the $\mathbf{T}_S(0)$ and $\mathbf{T}_R(0)$ components of the optimal perturbation of LIM1 and HCM1 to yield the maximum possible value of $\mu(t)$ given by (10). Since $B \geq 0$, the maximum value of $\mu(t)$ for HCM1 and LIM1 can be achieved only if $B = 0$, which corresponds to $\mathbf{T}_R(0) = 0$, and yields $\mu(t) = A$. Therefore the initial perturbation of HCM1 and LIM1 that produces the largest growth of $\mu(t)$ must, by necessity, project only onto the subspace spanned by the retained SST SVDs/EOFs $\mathbf{t}_i$. Therefore the initial SST of the optimal perturbations in Figs. 7 and 9 are linear combinations of only $\mathbf{t}_1$ and $\mathbf{t}_2$. This explains why the optimal perturbation structures of HCM1 and LIM1 are identical irrespective of whether $\mathbf{T}$ or $\mathbf{T}_S$ is chosen as the state vector.

2) LIM2 AND HCM2

The expression (8) for $\mu(t)$ also reduces to a common form for LIM2 and HCM2. In this case the optimal perturbation SST is given by $\mathbf{T}(0) = \mathbf{W}\hat{\mathbf{T}}(0) = \mathbf{W}\sum_{i=1}^{2} \hat{s}_i\hat{t}_i$, which relates the normalized ($\hat{\mathbf{T}}$) and unnormalized ($\mathbf{T}$) SST vectors, where $\mathbf{W} = \mathrm{diag}[\sigma(x)]$, and $\hat{\mathbf{t}}_i$ are the SST SVDs/EOFs derived from $\hat{\mathbf{T}}$ and $\boldsymbol{\tau}$ such that $\hat{\mathbf{t}}_i^{\mathrm{T}} \cdot \hat{\mathbf{t}}_j = \delta_{ij}$. Let $\hat{\mathbf{T}}_S$ be the component of normalized SST $\hat{\mathbf{T}}$ that spans the subspace defined by $\hat{\mathbf{t}}_1$ and $\hat{\mathbf{t}}_2$, and $\hat{\mathbf{T}}_R$ the component and subspace of $\hat{\mathbf{T}}$ that is orthogonal to $\hat{\mathbf{T}}_S$. Following section 4a(1), $\mathbf{T}(0) = [\mathbf{T}_S(0) + \mathbf{T}_R(0)]$, where now $\mathbf{T}_S(0) = \mathbf{W}\hat{\mathbf{T}}_S(0) = \sum_{i=1}^{N} \hat{s}_i\, \mathbf{W}\hat{\mathbf{t}}_i$, $\mathbf{T}_R(0) = \mathbf{W}\hat{\mathbf{T}}_R(0)$, and $\hat{\mathbf{T}}_S^{\mathrm{T}}(0) \cdot \hat{\mathbf{T}}_R(0) = 0$. The vectors $\mathbf{W}\hat{\mathbf{t}}_i$ are not orthogonal so in general $\mathbf{T}_S^{\mathrm{T}}(t) \cdot \mathbf{T}_R(t) \neq 0$. The growth factor $\mu(t)$ given by (8) can now be written as

$$\mu(t) = \frac{A + C_t}{1 + C_0 + B}, \tag{11}$$

where A and B are defined in section 4a(1), and C_t and C_0 are the dot-product terms $C_t = 2\mathbf{T}_S^{\mathrm{T}}(t) \cdot \mathbf{T}_R(t)/\mathbf{T}_S^{\mathrm{T}}(0) \cdot \mathbf{T}_S(0)$, and $C_0 = 2\mathbf{T}_S^{\mathrm{T}}(0) \cdot \mathbf{T}_R(0)/\mathbf{T}_S^{\mathrm{T}}(0) \cdot \mathbf{T}_S(0)$. For LIM2, $C_t = 0$ since according to (7) $\mathbf{T}_R(t) = 0$, but in general $C_0 \neq 0$. For HCM2, C_t and C_0 do not in general vanish so there is no requirement for $\mathbf{T}_R(0)$ to vanish when $\mu(t)$ attains its maximum value. The optimal perturbations of HCM2 and LIM2 will therefore span the complete space occupied by $\mathbf{T}_S$ and $\mathbf{T}_R$. The $\mathbf{T}_R$ component, however, produces no dynamical response in the atmosphere or ocean and will decay exponentially in time in HCM2, while in LIM2 it vanishes immediately according to (7). The nonorthogonality of

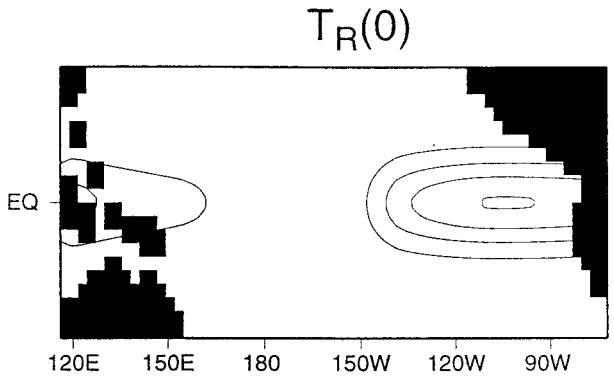


FIG. 13. The $\mathbf{T}_R$ component of the initial perturbation SST of the optimal perturbation of HCM2.

$\mathbf{T}_S(0)$ and $\mathbf{T}_R(0)$ in HCM2 and LIM2 means that it is possible for $\mathbf{T}_R(0)$ to conceal a large amplitude $\mathbf{T}_S(0)$ component. As the optimal perturbation evolves $\mathbf{T}_R$ decays revealing $\mathbf{T}_S$. To illustrate, Fig. 13 shows $\mathbf{T}_R(0)$ for the optimal perturbation of HCM2, and by adding this to $\mathbf{T}_S(0)$ shown in Fig. 8e the optimal perturbation of Fig. 8a is recovered. As Fig. 8e shows, the structure of $\mathbf{T}_S(0)$ resembles a mature ENSO episode, and has little in common with the overall initial SST structure of the optimal perturbation in Fig. 8a. The residual $\mathbf{T}_R(0)$ conceals the ENSO structure of $\mathbf{T}_S(0)$, and the latter emerges and dominates the optimal perturbation as $\mathbf{T}_R(0)$ undergoes exponential decay.

b. Quantifying nonnormality

While the SVD/EOF representation of $\mathbf{T}$ and $\hat{\mathbf{T}}$ helps us to understand the optimal perturbation structures of the HCMs and LIMs via the $\mathbf{T}_S$ and $\mathbf{T}_R$ decomposition, $\mathbf{T}_S$ and $\mathbf{T}_R$ are not eigensolutions of the linear propagator $\mathbf{R}$. A more appropriate and complete representation of state space for exploring optimal perturbation dynamics are the eigenmodes of the tangent linear propagator $\mathbf{R}$ of each model. The optimal perturbations of each model can be represented as a linear superposition of these eigenmodes. For each model the gravest eigenmode of $\mathbf{R}$ describes the ENSO oscillation (cf. Fig. 1), which we will refer to as the ENSO mode, and each model is nonnormal with nonorthogonal eigenmodes. The optimal perturbations of each coupled model evolve into a mature ENSO episode as shown in Figs. 6–10 so the optimal perturbations must have a large projection on the ENSO mode. Therefore the optimal perturbation is an attempt to initially conceal the ENSO mode using a linear combination of the other nonorthogonal eigenmodes of the system. If the eigenmodes were purely real, this state of affairs is best achieved if the ENSO mode is almost parallel (or antiparallel) to all of the other eigenmodes in phase space (Farrell and Ioannou 1996a). Another way of saying this for the more general case of complex eigenmodes is that there is a high degree of linear dependence between the ENSO mode and the other eigenmodes. This degree of linear dependence can be quantified by $\nu = |\mathbf{b}||\mathbf{e}|/(\mathbf{b}^{\mathrm{H}} \cdot \mathbf{e})$ where $\mathbf{e}$ is the ENSO eigenmode of the propagator $\mathbf{R}$, $\mathbf{b}$ is the corresponding eigenmode of $\mathbf{R}^{\mathrm{T}}$, the so-called adjoint ENSO mode, and the superscript H denotes the conjugate transpose. For normal systems $\mathbf{e}$ and $\mathbf{b}$ are identical and $|\nu| = 1$. For purely real eigenmodes $\nu = 1/\cos\theta$, where θ is the angle between $\mathbf{e}$ and $\mathbf{b}$. As $\mathbf{b}$ and $\mathbf{e}$ become more nearly orthogonal $\theta \rightarrow \pm\pi/2$ and ν becomes large. It is a fact that $\mathbf{b}$ is orthogonal to all the eigenmodes of $\mathbf{R}$ except $\mathbf{e}$, so $\theta \rightarrow \pm\pi/2$ corresponds to the case where $\mathbf{e}$ becomes more nearly parallel or antiparallel to the other eigenmodes of $\mathbf{R}$ (or more linearly dependent on the other eigenmodes in the case of complex eigenmodes). Thus ν is a useful measure of the nonnormality of the ENSO mode (Farrell and Ioannou 1999). The larger the value of ν the more nearly parallel $\mathbf{e}$ is to the other eigenmodes of $\mathbf{R}$, or the more linearly dependent $\mathbf{e}$ is on these other eigenmodes. Thus as ν increases it becomes easier to conceal the ENSO mode $\mathbf{e}$ using a linear combination of the other eigenmodes. Table 1 shows the values of $|\nu|$ for the ENSO modes of the various coupled models used and reveals (a) that the nonnormality of DYN is strongly influenced by LH, and (b) that normalizing $\mathbf{T}$ by $\sigma(x)$ in the HCMs and LIMs increases the nonnormality of these models, perhaps reintroducing some of the nonnormality associated with LH.

TABLE 1. Values of $|\nu|$ for the ENSO modes of each coupled model indicated where ν is defined in the main text.

Model	$\lvert\nu\rvert$
DYN	3.6
DYN, LH = 0	1.8
HCM1	1.2
HCM2	2.8
LIM1	1.1
LIM2	1.9

Hovmöller diagrams of SST of the adjoint ENSO mode $\mathbf{b}$ of each model are shown in Fig. 14 using $\mathbf{T}$ as the state vector in the HCMs and LIMs. The similarity between the SST structure of the adjoint ENSO mode and the corresponding optimal perturbations of each model (Figs. 6–10) provides visual confirmation of the large projection the optimal perturbations have on the ENSO mode. As $t \rightarrow \infty$ the optimal perturbation and adjoint ENSO mode of a given model become identical. Figures 14 and 6–10, however, reveal that this limit is practically reached after only a few months. Thus the optimal perturbations resemble very closely the optimal excitation of the ENSO mode itself for each model.

c. State space truncation and optimal perturbation structure

Another important issue is the degree to which the state space of the LIMs and HCMs overlap the state

space of DYN. Given that dynamical information is jettisoned in the construction of the LIMs and HCMs by discarding the SVDs/EOFs that account for a small fraction of SST variance, and in the case of LIMs by ignoring other dynamical variables, it is clear that the state space of these models is a subset of the state space of DYN. In the case of the LIMs it is a severely truncated subset. Following section 4b, a convenient set of basis functions for defining the state space of each coupled model are the eigenmodes of their respective propagators. The spectrum of DYN is described in appendix A and possesses 1082 eigenmodes, while HCM1 and HCM2 possess 1034, and in LIM1 and LIM2 only two eigenmodes, that are complex conjugates of each other, are retained that correspond to the POPs of the estimated propagator.

Recall that the optimal perturbations $\boldsymbol{\eta}$ of any model can be expressed as a linear superposition of the nonorthogonal eigenmodes $\boldsymbol{\phi}_i$, so that $\boldsymbol{\eta} = \sum_{i=1}^{N} a_i \boldsymbol{\phi}_i$. The coefficients a_i can be determined by applying the biorthogonality relation of section 4b to each eigenmode of DYN. Figure 15 shows the coefficients $|a_i|$ of the optimal perturbation of DYN (cf. Fig. 6a) for each of the first 40 eigenmodes (see also Fig. A1c). In Fig. 15 the ENSO mode is eigenmode #2. Clearly the optimal perturbation is composed of several eigenmodes and many of these are low-frequency coupled modes that are absent in HCM1 and HCM2. Only the gravest eigenmode that defines the ENSO oscillation is retained in LIM1 and LIM2.

The connection between the statistical and dynamical representations of state space can now be made. It has been demonstrated that HCM2 and LIM2 can recover the optimal perturbation of DYN only if (a) sufficient SVDs/EOFs are retained, and (b) the optimal perturbations initially project onto the entire state space of $\mathbf{T}$. A relevant question therefore is whether the dynamics of the optimal perturbations of HCM2 and LIM2 mimic those of DYN. As shown in section 4a(2), the optimal perturbations of HCM2 and LIM2 are composed of components $\mathbf{T}_S$ and $\mathbf{T}_R$, and in the cases considered $\mathbf{T}_S$ primarily describes the ENSO eigenmode. In LIM2, $\mathbf{T}_R$ vanishes immediately according to (7) since $\mathbf{E}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{T}_R(0) = 0$, leaving the ENSO mode described by $\mathbf{T}_S$. In HCM2, $\mathbf{T}_R$ is exponentially damped and produces no atmospheric or oceanic response, and so projects only onto nonoscillatory damped eigenmodes. If the SST propagator $\mathbf{R}$ of HCM2 in gridpoint space is dimension $N \times N$ it will possess two eigenmodes with eigenvalues ω and ω^* that describe the ENSO mode, and $N - 2$ eigenmodes with identical eigenvalues $e^{-\epsilon\tau}$ that can be represented as linear combinations of the SST SVDS not retained in $\mathbf{T}_S$. It is these latter eigenmodes onto which $\mathbf{T}_R$ of the optimal perturbation of HCM2 projects. Figure A1c shows the eigenfrequencies of the eigenmodes for which $|a_i| > 10^{-2}$ in Fig. 15. Clearly the optimal perturbation of DYN is composed primarily of oscillating eigenmodes and not stationary modes of which there are very few. Thus even though HCM2 recovers much of the SST structure of the optimal perturbation of DYN, the evolution of the two perturbations is described by eigenmodes with very different dynamics. The two SVDs that make up the atmospheric component of HCM2 (Fig. 5) represent the SST and wind stress of two phases of the ENSO mode in that model. Thus the ENSO mode is the only coupled ocean–atmosphere oscillating mode that HCM2 can support. As shown in the appendix the only other oscillating eigenmodes that will be present in HCM2 are those associated with the ocean alone and these are primarily decoupled from the atmosphere.

In the optimal perturbation of DYN the ENSO mode is concealed by several low-frequency coupled modes that dissipate to reveal the ENSO mode. In HCM2 this behavior is mimicked by nonoscillating, damped modes described by $\mathbf{T}_R$, while in LIM2 it is achieved by the immediate disappearance of $\mathbf{T}_R$. Clearly the dynamics of the optimal perturbations of the HCMs and LIMs differ from their DYN counterpart.

d. LIMs constructed from SST and h

In addition to SST, some of the LIM calculations reported in the literature have used thermocline depth h or upper-ocean heat content anomalies and wind stress anomalies to compute optimal perturbations. (Blumenthal 1991; Xue et al. 1994, 1997a,b; Y. Xue 1999, personal communication; Johnson et al. 1999, unpublished manuscript). Since h and heat content are probably not strongly influenced by nonlinearity, and heat content anomalies have a large ampltiude in the west and central Pacific, it is seems logical to suppose that LIMs constructed from these fields may be able to reproduce the optimal perturbation structure of DYN. To test this idea two variants of LIM1 and LIM2 were constructed using time series of T and h from a 100-yr integration of nonlinear DYN. We will refer to the LIMs that use unnormalized and normalized SSTs as LIM1A and LIM2A, respectively. As shown in section 4b the adjoint ENSO mode is a good surrogate for the optimal perturbation for the 3-month optimal growth time used here. Figure 16 shows Hovmöller diagrams of T and h along the equator for the adjoint ENSO mode of LIM1A and LIM2A. Also shown in Fig. 16a is a Hovmöller diagram of the adjoint ENSO mode h field (see Fig. 14a for the corresponding structure of T). Figure 16b reveals that when unnormalized SSTs are used as in LIM1A the adjoint ENSO mode SST structure is incorrect and similar to that of LIM1 (Fig. 14d). However, the adjoint ENSO mode h structure (Fig. 16c) is reasonably well captured by LIM1A. When normalized SSTs were used as in LIM2A both the SST and h structure of the adjoint ENSO mode (Figs. 16d,e) were recovered reasonably well. Therefore Fig. 16 reveals that the use of h field data in the construction of the LIMs does not guarantee that the SST structure of the resulting optimal perturbations will be correct. Inter-

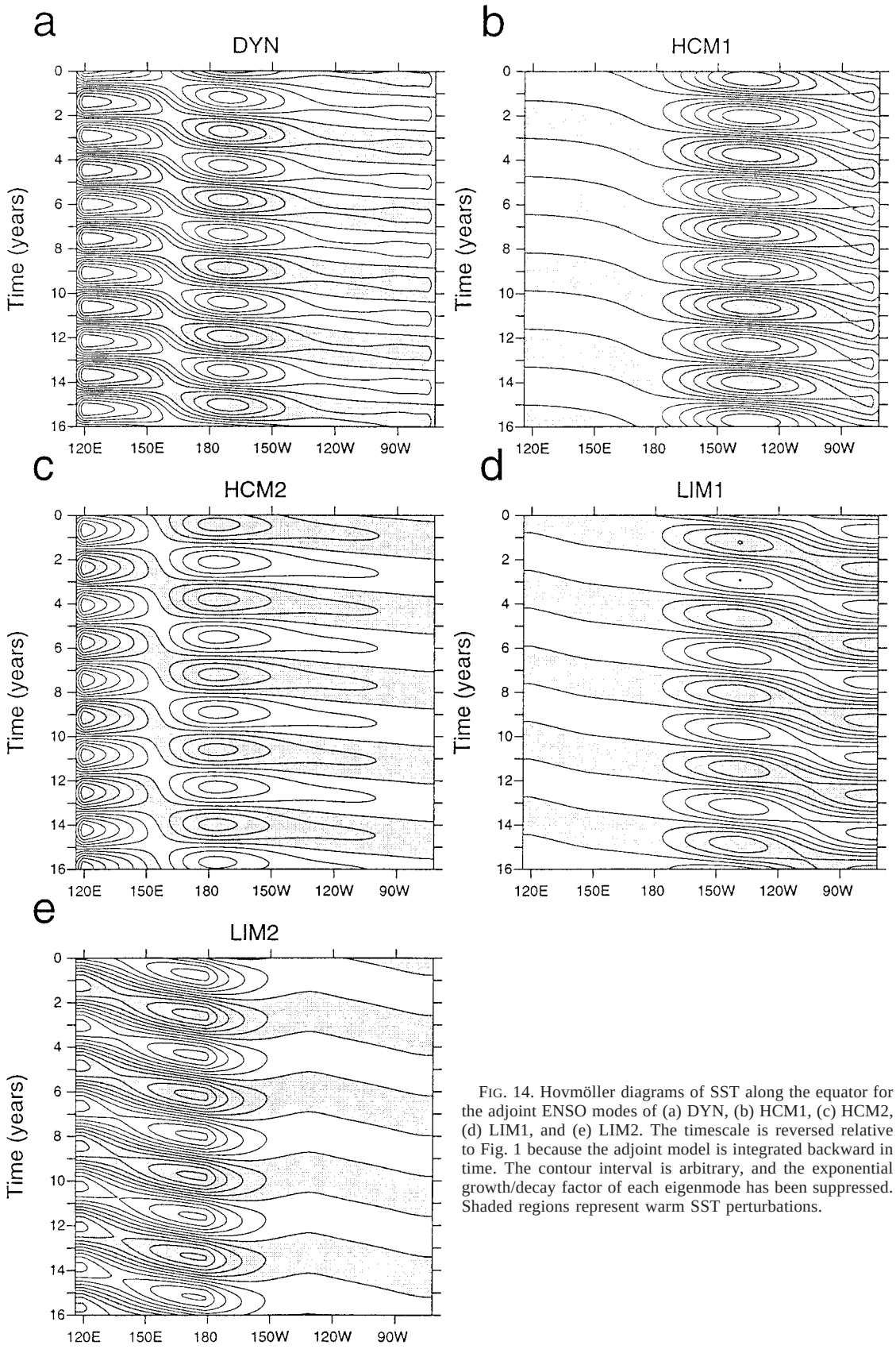


FIG. 14. Hovmöller diagrams of SST along the equator for the adjoint ENSO modes of (a) DYN, (b) HCM1, (c) HCM2, (d) LIM1, and (e) LIM2. The timescale is reversed relative to Fig. 1 because the adjoint model is integrated backward in time. The contour interval is arbitrary, and the exponential growth/decay factor of each eigenmode has been suppressed. Shaded regions represent warm SST perturbations.

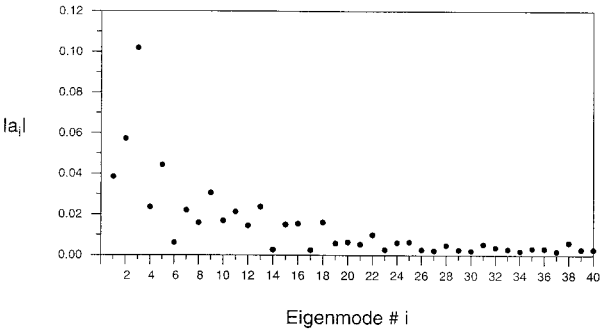


FIG. 15. The magnitude of the eigenmode projection coefficient $|a_i|$ of the optimal perturbation of DYN on the first 40 eigenmodes with largest eigenvalue modulus of the tangent linear coupled model.

estingly LIM1A and LIM2A are not able to recover all of the details of the optimal perturbation *h* structure of DYN. This appears to be due to severe state space truncation of the LIMs discussed in section 4c. However in common with the optimal perturbations of DYN, Figs. 16c,e show that the LIMs do recover the thermocline maximum displacement in the central Pacific (analogous to a heat content maximum or minimum), which is an observed feature of the tropical Pacific prior to El Niño episodes. Since $\boldsymbol{\tau}$ is slaved to T in DYN, using wind stress in the construction of the LIMs does not contribute any additional information.

5. Norm dependence

So far our attention has been confined to the SST^2 norm. In this case it was found that when SST is suitably normalized a HCM and LIM can recover the gross features of the optimal perturbation of the dynamical model from which the HCM and LIM are derived. It is pertinent to ask whether these same results carry over to other norms.

In this section we will consider three generic types of norms. Using the notation of section 4, these are

1) norms that do not depend on $\mathbf{T}$,
2) norms that depend only on the $\mathbf{T}_S$ component of $\mathbf{T}$,
3) norms that depend on both the $\mathbf{T}_S$ and $\mathbf{T}_R$ components of $\mathbf{T}$,

where $\mathbf{T}_S$ and $\mathbf{T}_R$ refer to the components of the HCMs and LIMs only. A specific example of each of type of norm will be considered. The perturbation energy of the ocean $\mathcal{E}_O$ was used as an example of the form 1 under the assumption that $\mathbf{T} = 0$ at initial time for optimal perturbations of this norm. The combined perturbation energy of the atmosphere and ocean $\mathcal{E}_T$ was used as an example of the form 2 since $\mathbf{T}_R$ does not contribute to the energy of either component of the coupled system. A mixed norm $\mathcal{N} = a\mathbf{T}^{\mathrm{T}} \cdot \mathbf{T} + b\mathcal{E}_O$ was used as an example of the form 3, where the weighting coefficients *a* and *b* control the relative contributions of SST and ocean energy to the norm. The initial structures of the optimal perturbations of each norm from DYN, HCM1, and HCM2 are shown in Fig. 17. For norms that can be applied to the linear inverse models LIM1 and LIM2 we anticipate that the same general results described below will apply. The results shown in Fig. 17 are summarized below and in each case $\mathbf{T}$ was used as the state vector for the HCMs and LIMs.

a. Ocean energy norm, $\mathcal{E}_O$

Since a norm of the form 1 does not explicitly involve SST, the thermocline depth structures of the optimal perturbations of each model by and large agree as shown in Figs. 17a–c. Recall the assumption that $\mathbf{T}(0) = 0$.

b. Total energy norm, $\mathcal{E}_T$

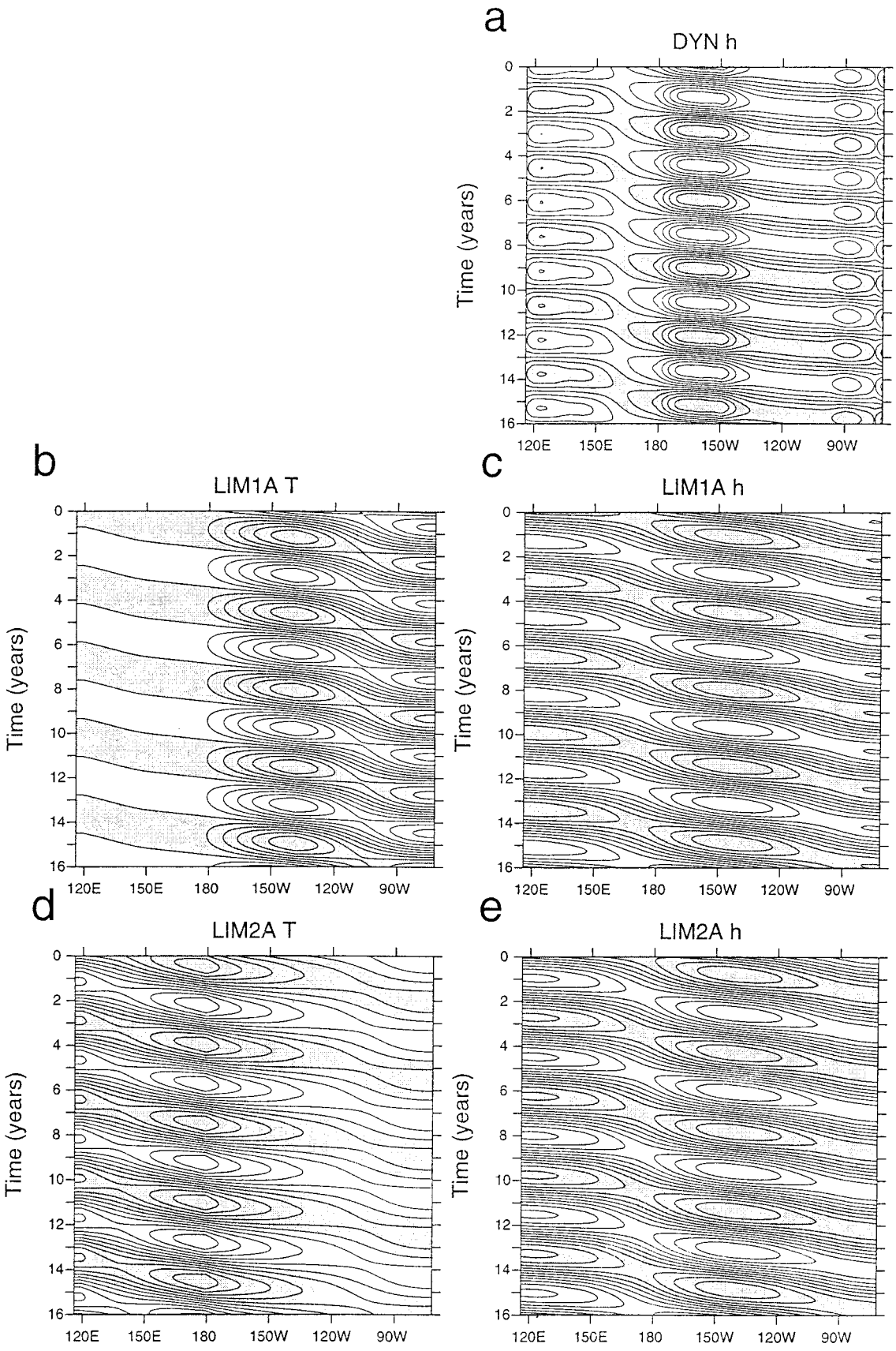
Only $\mathbf{T}_S$ contributes to norms of the form 2 so the optimal perturbations of HCM1 and HCM2 are similar to each other as shown in Figs. 17e,h and 17f,i since $\mathbf{T}_R = 0$ for both. The HCMs can recover the thermocline structure of the optimal perturbation of DYN but not its SST structure. The latter is due to two factors. First, the $\mathcal{E}_T$ norm renders the atmosphere of each HCM essentially normal so that only $\mathbf{T}_S$ contributes to their optimal perturbations (see section 4c). Second, the state-space truncation of the HCMs means that the high wavenumber eigenmodes that are so prominent in the optimal perturbation of DYN (Fig. 17d) are absent in the HCMs (see section 4d). The origin of these high wavenumber eigenmodes is discussed in Moore and Kleeman (1996).

c. Mixed norm

In the case of the mixed norm, HCM2 is capable of recovering all of the gross features of the thermocline depth and SST of the optimal perturbation of DYN provided enough SVDs are retained in its atmospheric component (Figs. 17l,o). The nonnormality reintroduced into the system by $\mathbf{W} = \mathrm{diag}[\sigma(x)]$ allows $\mathbf{T}_R \neq 0$ to conceal the ENSO eigenmode signature in the initial SST structure of the optimal perturbation of HCM2. For the optimal perturbation of HCM1 (Fig. 17k,n) $\mathbf{T}_R = 0$ at initial time as discussed in section 4a(1). Figures 17j–o show the case for $a = b = 1$ but the same general results hold for other combinations of *a* and *b* (not shown).

6. Summary and conclusions

We have examined the idea that a shortcoming in the ability of statistically based coupled models and linear inverse models may be one reason why the SST structures of the optimal perturbations of various different coupled models reported in the literature differ from one another. This idea has been explored using a dynamical coupled model and various approximations to it in the form of HCMs, that replace the dynamical



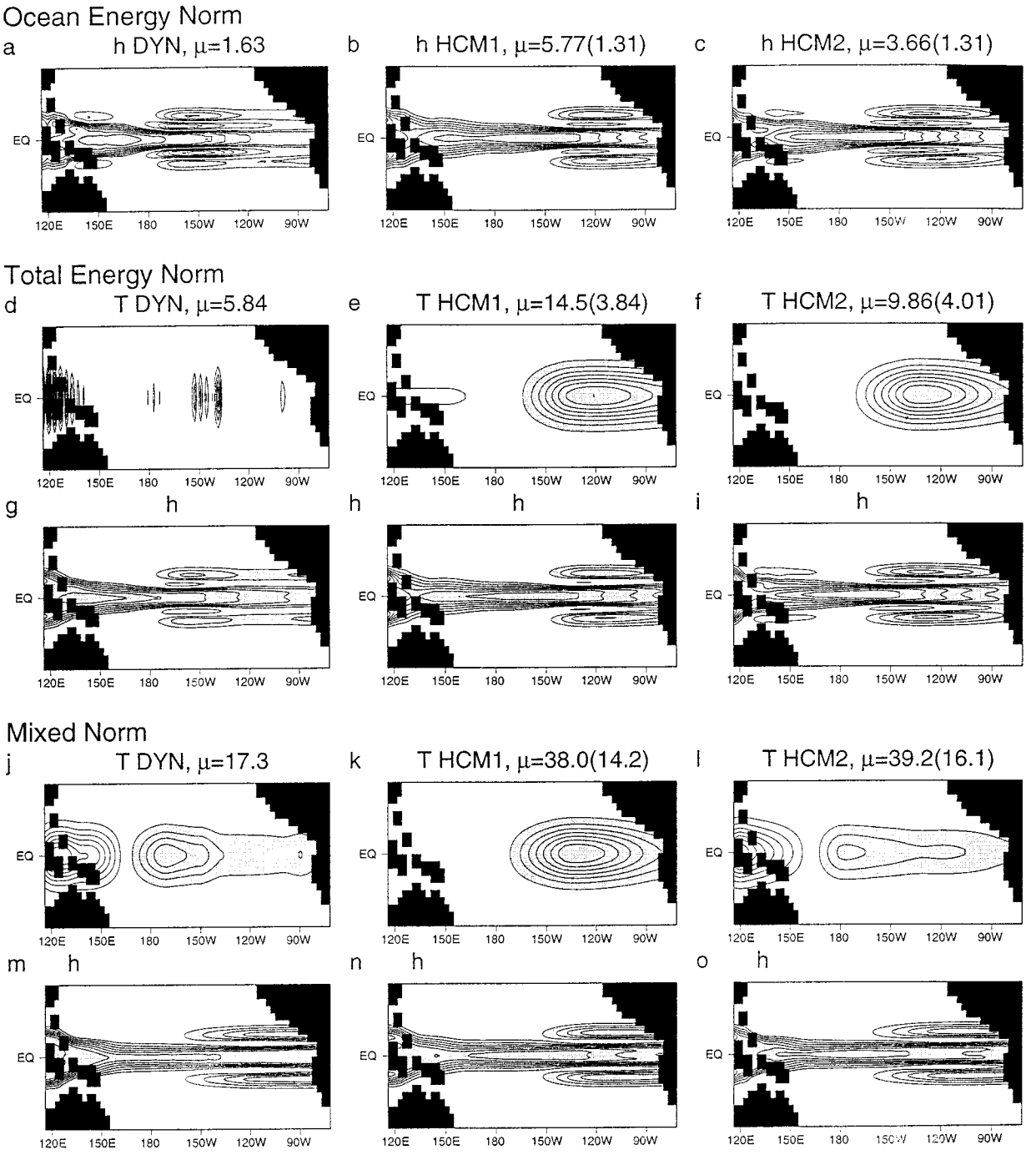


FIG. 17. The initial perturbation SST T and/or thermocline depth perturbation h for the optimal perturbations of DYN, HCM1, and HCM2 using (a)–(c) an ocean energy norm, $\mathcal{E}_O$; (d)–(i) the total energy (atmosphere energy + ocean energy) norm, $\mathcal{E}_T$; and (j)–(o) the mixed norm $a\mathbf{T}^{\mathrm{T}} \cdot \mathbf{T} + b\mathcal{E}_O$, with $a = b = 1$. The contour interval is arbitrary, and shaded regions represent warm SST anomalies and an anomalously deep thermocline. The growth factor μ of the corresponding norm of each perturbation evolved in each model and DYN (value in parentheses) is indicated.

←

FIG. 16. Hovmöller diagrams along the equator of the adjoint ENSO modes of (a) h for DYN, (b) T for LIM1A, (c) h for LIM1A, (d) T for LIM2A, and (e) h for LIM2A. The timescale is reversed relative to Fig. 1 because the adjoint model is integrated backward in time. The contour interval is arbitrary, and the exponential growth/decay factor of each eigenmode has been suppressed. Shaded regions represent warm SST anomalies and an anomalously deep thermocline.

atmospheric model with a statistical approximation, and LIMs based on time series derived from the dynamical model. The main finding of this paper is that if there are important nonlinearities in the dynamical model that exert a considerable influence on the model variables used in the construction of the HCMs and LIMs, and that also influence the structure of the optimal perturbations in the dynamical model, then this must be accounted for in some way during the construction of the HCMs and LIMs. If not the HCMs and LIMs may be unable to recover the correct optimal structures of the dynamical model. In addition the state-space truncation of the HCMs and LIMs will prevent them from recovering the optimal structure unless the optimal perturbations project onto the entire state space of the system.

The important nonlinearity discussed here is the relative influence of equal amplitude SST anomalies in the west and east tropical Pacific on the atmospheric circulation, as governed by the Clausius–Clapeyron relation and moisture convergence feedback by virtue of the warm average SSTs that prevail in the warm pool of the western tropical Pacific. As a result of this, the atmosphere is generally close to a state of convective instability over the warm pool, and SST anomalies of only a few tenths of a degree can initiate large changes in the surface wind stress due to changes in deep penetrative convection. This nonlinearity is included in the dynamical coupled model of Kleeman (1993) used here. The growth of perturbations in the tangent linear dynamical coupled model is strongly influenced by the tangent linear convective process, which is one factor that controls the nonnormality of the ENSO mode of the system (Moore and Kleeman 1999b). The nonnormality of the ENSO mode translates into the ability to construct perturbations that excite the ENSO mode, and at the same time grow rapidly even if the ENSO mode is stable. As the nonnormality of the ENSO mode increases it becomes easier to conceal it using a linear combination of the other nonorthogonal eigenmodes. In the Kleeman coupled model this leads to an optimal perturbation with an initial SST structure in the form of a dipole with largest amplitude over the warm pool due to the important role played by convectively induced latent heating in this region. Using a standard approach to the construction of HCMs and LIMs in which time series of raw SST anomalies are used, these models fail to recover the optimal perturbation SST structure of the dynamical model. This is due to the fact that the ENSO modes of the resulting HCMs and LIMs are more normal than their dynamical model counterpart. As a result it is difficult to conceal the ENSO mode, and the initial optimal perturbation SST structures of the HCM and LIM resemble the mature phase of the ENSO mode. This result is not suprising given the fact that for a normal system the optimal perturbation for the ENSO mode would be the ENSO mode itself. The HCM and LIM are nearly normal systems because the use of raw SST time series used in their construction gives most weight to regions where absolute SST variability is largest, namely the eastern Pacific. The lower amplitude variability in the west, that is just as important as the variability in the east in the dynamical model, is relegated to SVDs or EOFs that account for a small fraction of variability of the raw data and that are typically discarded. If instead of using raw SST data we use SST that is suitably scaled to reflect the importance of SST variability in the west, then nonnormality is reintroduced into the system and the HCM and LIM can recover the gross features of the optimal perturbation SST of the dynamical model. It has been demonstrated that normalizing the SST yields optimal perturbations that may project onto the entire state space of SST.

The ability of the HCMs and LIMs to recover the correct optimal structures also depends on the state-space truncation of these models. In the case of the LIMs this truncation can be severe. In a dynamical model, the optimal perturbation may be a linear superposition of many eigenmodes, but many of these modes may have been filtered out of the HCMs and LIMs by the statistical approximations they employ. Thus even if DYN, HCM, and LIM produced ENSO modes that were nonnormal to the same degree, their optimal perturbations could still be different and evolve differently. In practice we have seen that a HCM and LIM that are constructed using suitably weighted SST anomalies are capable of recovering the optimal perturbation structure of the dynamical model. However given the different dimensions of state space in the HCM and LIM compared to the dynamical model, the dynamics of the eigenmodes that conceal the ENSO mode component of the optimal perturbation need not be the same in each model. In the cases examined here we found that the optimal perturbation of DYN was composed primarily of low-frequency coupled eigenmodes. The optimal perturbation of the HCM2, however, while very similar in structure to that of DYN, was composed primarily of damped, nonoscillating modes. In the case of LIM2, the components of the initial optimal perturbation that do not project into the ENSO mode vanish immediately. Therefore HCM2 and LIM2 provide misleading information about the dynamics of optimal perturbation growth even though their optimal perturbation structures are by and large correct.

7. Discussion

There are two primary reasons why the optimal perturbations of ENSO models have received so much attention. First, for a particular model, they offer a means of understanding and quantifying the growth of initial condition errors during ENSO forecasts. The problem of error growth is clearly model specific, so differences in SST structure between the optimal perturbations of different coupled models is of interest only because it may help explain the difference in forecast skill displayed by different models for specific ENSO events.

Second, the optimal perturbations suggest possible dynamical mechanisms by which ENSO episodes may be triggered or affected by perturbations that arise in nature. An extension of the optimal perturbation theory used here demonstrates the role that these perturbations can also play in the presence of stochastic forcing (Kleeman and Moore 1997), a subject of considerable debate at the present time with regard to the relationship between the Madden–Julian oscillation and ENSO (Kerr 1999). Lack of consensus on the correct optimal perturbation for ENSO is therefore clearly a problem in this regard because different models may display sensitivity to different kinds of perturbations arising from quite different physical phenomena. In the light of these comments we will consider the relevance of our results to previous work with coupled models, and to the real coupled system.

a. Relevance of results to other coupled model studies

It is of interest to place the results of this paper in the context of the optimal perturbation calculations reported elsewhere in the literature by various investigators using different coupled models. As noted in section 1 the models used by other investigators, with the exception of Eckert (1999), can all be considered as intermediate coupled models of ENSO. Thus each model consists of an equatorial ocean model and some form of simple atmospheric model, either a dynamical model or a statistical model. The model used by Eckert (1999) consists of an ocean GCM coupled to a statistical atmospheric model. Linear inverse techniques have also been used to compute the optimal perturbations of either dynamical coupled models or of the real coupled system. In the case where they have been computed (see Blumenthal 1991; Xue et al. 1994, 1997a,b; Moore and Kleeman 1996, 1997a,b; Chen et al. 1997; Thompson 1998; Fan 1998; Fan et al. 2000; Eckert 1999), the thermocline depth structures of the optimal perturbations of each coupled model and LIM are qualitatively similar to each other. The results of section 5 suggest that this similarity is due primarily to the role played by linear equatorial wave dynamics, which are by and large the same in each model. On the other hand the SST structures of the optimal perturbations differ between coupled models and LIMs (see Blumenthal 1991; Xue et al. 1994, 1997a,b, 2000; Penland and Sardeshmukh 1995; Moore and Kleeman 1996, 1997a,b; Chen et al. 1997; Thompson 1998; Fan 1998; Fan et al. 2000; Eckert 1999; Thompson and Battisti 2001). The results of this study suggest that these differences are most likely due to differences between the atmospheric components of the coupled models.

An interesting comparison can also be made between the results presented here and those from the Zebiak and Cane model that has been discussed extensively in the literature. The optimal structures of the Zebiak and Cane (1987) (ZC) coupled model were computed by Blumenthal (1991) and Xue et al. (1994, 1997a,b) using a linear inverse approximation of the dynamical model. Computation of the optimal structures have also been made directly using a dynamical model by Chen et al. (1997) and Thompson (1998) using the Battisti (1988) version of ZC. The optimal perturbation SST and thermocline depth structures from the dynamical model and linear inverse approximation of ZC agree with each other, which at first sight may appear to conflict with the results presented here. To resolve this apparent conflict we must consider the dynamics of the atmospheric component of the ZC and Kleeman models. The ZC model does not possess a latent heating nonlinearity like (3) over the warm pool, but instead includes a nonlinear total wind-convergence heating anomaly in the east and southwest Pacific associated with the ITCZ and SPCZ, respectively. The SPCZ related heating anomalies, however, turn out to be relatively unimportant in the model (Zebiak 1986). As shown in Fig. 12a, if $LH = 0$ in the Kleeman model, the location of the optimal perturbation SST maximum moves into the eastern and central Pacific and is more similar to that of the ZC and Battisti models than the case when $LH \neq 0$. Figures 12a, 12c, and 12d also show that a HCM and LIM constructed from the ENSO oscillation with $LH = 0$ recovers the gross feature of the optimal perturbation SST structure of the dynamical model. It is our belief therefore that the similarity in structure of the optimal perturbations of the ZC model and its linear inverse equivalent is due to the fact that the region where the nonlinear SST effects are important (i.e., under the ITCZ) is also within the region where SST variance is largest, namely the central and east Pacific. This idea was verified by the present authors using time series from the Kleeman model, and artificially increasing the weight given to SST anomalies in the east Pacific when constructing HCMs and LIMs to simulate the presence of a nonlinear atmospheric response to SST there. In this case the SST structure of the optimal perturbations of the HCMs and LIMs become more tightly confined to the east Pacific (not shown).

The results of this study offer a plausible explanation for some of the differences that exist in the SST structures of optimal perturbations of various different coupled models and LIMs. In particular they explain why the SST structure of the optimal perturbations of the Kleeman model differ from those of other models. Clearly these ideas deserve more attention since they strongly suggest that the integrity of optimal perturbation calculations for ENSO based on LIMs or coupled models that possess statistical atmospheres may be questionable. The importance of this issue in relation to the real coupled system can be appreciated by performing the following thought experiment. Suppose for a moment that DYN represents the real coupled system. Figure 6 would then represent the true optimal perturbation in nature. The HCMs and LIMs, however, represent at-

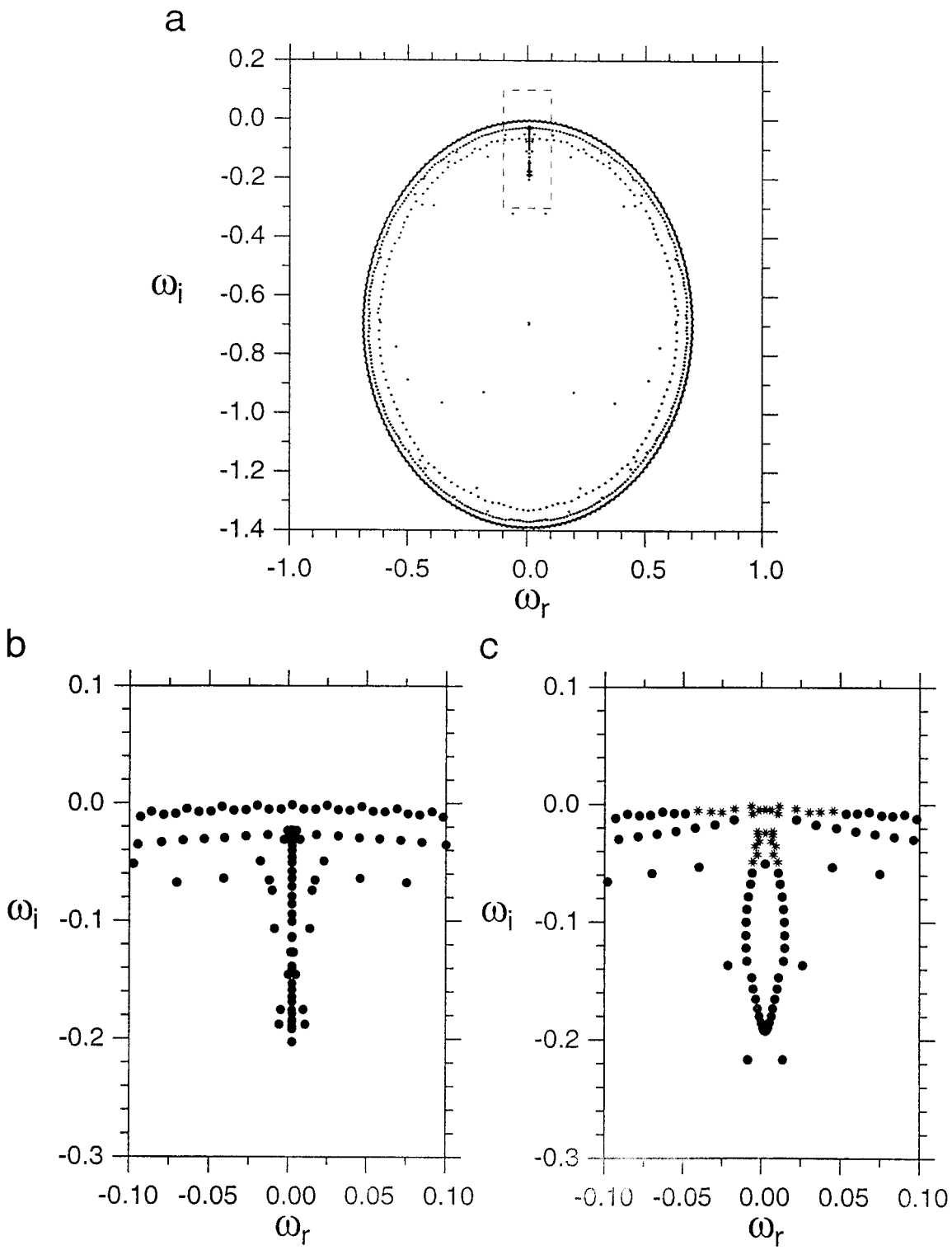


FIG. A1. (a) The complete eigenvalue spectrum of the ocean model and SST equation. The eigenmodes evolve in time as $e^{-i\omega t}$ where $\omega = \omega_r + i\omega_i$ in units of 2π (days)$^{-1}$. (b) A close-up view of the ocean model eigenvalue spectrum in the vicinity of the origin within the dashed box of (a). (c) Same as (b) but for the coupled model DYN. The points denoted by stars in (c) are those eigenmodes for which $|a_i| > 0.01$ in Fig. 15.

tempts to construct approximations of the real system from observations that sample its variability. Figures 7–12 indicate that incorrect estimates of the true optimal perturbation for ENSO will be obtained using HCMs and LIMs unless steps are taken to account for the important effects of nonlinearity in nature.

b. Relevance of results to the real coupled system

It is of interest to consider the relevance of our results to the real coupled system since clearly the dynamical coupled model used here is very simple in many respects. We have demonstrated that statistical approximations of a dynamical coupled model that is known to possess important nonlinearities may yield unreliable estimates of the optimal perturbations, unless steps are taken to incoporate the known influence of nonlinearities on the system response in the statistical models themselves. The important nonlinearities discussed in this paper were included in the dynamical coupled model used here because they are known to exist in the nature. Therefore the results of this paper are pertinent to optimal perturbation estimates that have been made by Penland and Sardeshmukh (1995), Y. Xue (1999, personal communication), and Johnson et al. (1999, unpublished manuscript) using LIMs constructed from observed SSTs. The results of this study suggest that unless the influence of the west Pacific warm pool SST nonlinearity is incorporated into a LIM, the resulting optimal perturbation estimates may be significantly in error. To our knowledge, however, the effects of SST nonlinearties on the optimal perturbation structures have not been explored in the studies cited above. This, however, raises the important issue of how to correctly incorporate the nonlinear influences of SST on the system into LIMs. In the work presented here we used the somewhat ad hoc, yet physically justifiable method of weighting the SST anomalies at each longitude by the SST standard deviation. This works well in the Kleeman model because the SST is computed only at the equator so only a 1D variation in the nonlinear influence of SST need be considered (i.e., warm west vs cold east). However, it is conceivable that other choices of weight function may perform just as well in terms of recovering the correct optimal structure. If actual SST observations from the Pacific Ocean are used to construct a LIM then the 2D character of SST-related nonlinearities in every region of the Pacific basin must be considered. Weighting SST by its standard deviation everywhere is clearly inappropriate in this case since this will emphasize the importance of SST anomalies not only over the warm pool where they are known to be important, but also in regions that may be cold in the mean and have small SST variance, and where obviously there is no reason to expect SST anomalies to be any more important than similar anomalies in a region that in the mean is cold but characterized by large SST variance. Thus a more physically based weighting is required; otherwise, by necessity, the weighting must be based on a priori assumptions about the influence of SST on the atmospheric response in different parts of the basin. The latter is clearly undesirable since all of the important SST-related nonlinearities in the Pacific may not yet have been identified. Clearly the issue of how to correctly weight observed SST anomalies when constructing LIMs requires careful consideration. Some of these problems may perhaps be circumvented using principal interaction patterns, which are the basis functions of nonlinear, reduced state space, models of dynamical systems (Hasselmann 1988). These, and the other issues discussed above, are currently being explored by the authors and will be presented elsewhere in due course.

Acknowledgments. This work was supported in part by Grant ATM9809790 from the NSF Climate Dynamics Division. The authors are grateful to Prof. Dan Sorensen for making the ARPACK libraries used for many of our calculations freely available. Helpful discussions with Brian Farrell and Petros Ioannou are also gratefully acknowledged.

APPENDIX

Eigenspectra

The optimal perturbations can be represented as a linear superposition of the nonorthogonal eigenmodes of the coupled models. While the eigenvalue spectra of DYN, HCM1, and HCM2 are different, the basic structure of their spectra is the same and determined by the ocean component of the coupled models, which is common to each model. Figure A1a shows the complete eigenvalue spectrum of the dynamic and thermodynamic components of the ocean component of the models when uncoupled from the atmosphere. Note the presence of a branch of eigenmodes with zero real frequency, which are nonoscillating modes. The eigenvalue spectrum of the coupled model DYN is very similar to that shown in Fig. A1a and a close up view of the origin of the DYN spectrum is shown in Fig. A1c since this is where differences exist in the eigenspectra between the ocean model and coupled model. Fig. A1b shows a close-up view of the ocean model spectrum in Fig. A1a. The branch of stationary modes in the ocean model has changed in the coupled model in which many of these modes oscillate with low frequency. The eigenvalue spectra of HCM1 and HCM2 are very similar to those of DYN (not shown). A number of stationary modes exist in HCM2, which form the basis for the residual component $\mathbf{T}_R$ of the optimal perturbations of HCM2 (see section 4a).

REFERENCES

Barnston, A. G., and Coauthors, 1994: Long-lead seasonal forecasts—Where do we stand? *Bull. Amer. Meteor. Soc.,* **75,** 2097–2114.

Battisti, D. S., 1988: Dynamics and thermodynamics of a warming

event in a coupled tropical atmosphere–ocean model. *J. Atmos. Sci.,* **45,** 2889–2919.

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

Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate,* **5,** 541–560.

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 the annual cycle and the ENSO cycle. *Mon. Wea. Rev.,* **125,** 831–845.

Davey, M. K., S. Ineson, and M. A. Balmaseda, 1994: Simulation and hindcasts of tropical Pacific Ocean variability. *Tellus,* **46A,** 433–447.

Eckert, C., 1999: On predictability limits of ENSO. A study performed with a simplified model of the tropical Pacific ocean–atmosphere system. Examensarbeit Nr. 55, 76 pp. [Available from Max-Planck-Institut für Meteorologie, Bundesstr. 55, D-20146 Hamburg, Germany.]

Fan, Y., 1998: ENSO prediction and predictability in an intermediate coupled model. Ph.D. dissertation, University of Oxford, 241 pp. [Available from Dept. of Atmospheric, Oceanic, and Planetary Physics, University of Oxford, Parks Rd., Oxford OXl 3PU, United Kingdom.]

——, 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.

Farrell, B. F., 1982: The initial growth of disturbances in baroclinic flow. *J. Atmos. Sci.,* **39,** 1663–1686.

——, and P. J. Ioannou, 1996a: Generalized stability theory. Part I: Autonomous operators. *J. Atmos. Sci.,* **53,** 2025–2040.

——, and ——, 1996b: Generalized stability theory. Part II: Nonautonomous operators. *J. Atmos. Sci.,* **53,** 2041–2053.

——, and ——, 1999: Perturbation growth and structure in time-dependent flows. *J. Atmos. Sci.,* **56,** 3622–3639.

Godfrey, J. S., R. A. Houze Jr., R. H. Johnson, R. Lukas, J. L. Redelsperger, A. Sumi, and R. Weller, 1998: Coupled ocean–atmosphere response experiment (COARE): An interim report. *J. Geophys. Res.,* **103,** 14 395–14 450.

Goswami, B. N., and J. Shukla, 1991: Predictability of a coupled ocean–atmosphere model. *J. Climate,* **4,** 3–22.

Graham, N., and T. P. Barnett, 1987: Observations of sea surface temperature and convection over tropical oceans. *Science,* **238,** 657–659.

Hasselmann, K., 1988: PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.,* **93,** 11 015–11 021.

Hoerling, M. P., A. Kumar, and M. Zhong, 1997: El Niño, La Niña, and the nonlinearity of their teleconnections. *J. Climate,* **10,** 1769–1786.

Kerr, R. A., 1999: Does a glode-girdling disturbance jigger El Niño? *Science,* **285,** 322–323.

Kleeman, R., 1989: A modeling study of the effect of the Andean Mountains on the summertime circulation of tropical South America. *J. Atmos. Sci.,* **46,** 3344–3362.

——, 1991: A simple model of the atmospheric response to ENSO sea surface temperature anomalies. *J. Atmos. Sci.,* **48,** 3–18.

——, 1993: On the dependence of hindcast skill on ocean thermodynamics in a coupled ocean–atmosphere model. *J. Climate,* **6,** 2012–2033.

——, 1994: Forecasts of tropical Pacific SST using a low-order coupled ocean–atmosphere dynamical model. *NOAA Experimental Long-Lead Forecast Bulletin* (since June 1994). [Available online at http://nic.fb4.noaa.gov/products/predictions/experimental/bulletin/index.html.]

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

——, and ——, 1999: A new method for determining the reliability of dynamical ENSO predictions. *Mon. Wea. Rev.,* **127,** 694–705.

——, M. Latif, and M. Flügel, 1992: A hybrid tropical atmosphere ocean model: Sensitivities and hindcast skill. Max-Planck-Institut für Meteorologie Rep. 76, 37 pp. [Available from Max-Planck-Institut für Meteorologie, Bundesstr. 55, D-20146 Hamburg, Germany.]

——, A. M. Moore, and N. R. Smith, 1995: Assimilation of subsurface thermal data into an intermediate tropical coupled ocean–atmosphere model. *Mon. Wea. Rev.,* **123,** 3103–3113.

Latif, M., T. P. Barnett, M. A. Cane, M. Fügel, N. E. Graham, H. von Storch, J.-S. Xu, and S. E. Zebiak, 1994: A review of ENSO prediction studies. *Climate Dyn.,* **9,** 167–179.

Lau, K.-M., 1985: Elements of a stochastic dynamical theory of the long-term variability of the El Niño–Southern Oscillation. *J. Atmos. Sci.,* **42,** 1552–1558.

Mayer, D. A., and R. H. Weisberg, 1998: El Niño–Southern Oscillation-related ocean–atmosphere coupling in the western tropical Pacific. *J. Geophys. Res.,* **103,** 18 635–18 648.

McPhaden, M. J., 1999: Genesis and evolution of the 1997–98 El Niño. *Science,* **283,** 950–954.

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 ——, 1997a: The singular vectors of a coupled ocean–atmosphere model of ENSO. Part I: Thermodynamics, energetics and error growth. *Quart. J. Roy. Meteor. Soc.,* **123,** 953–981.

——, 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 ——, 1998: Skill assessment for ENSO using ensemble prediction. *Quart. J. Roy. Meteor. Soc.,* **124,** 557–584.

——, 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.

Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.,* **117,** 2165–2185.

——, 1996: A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D,* **98,** 534–558.

——, and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate,* **8,** 1999–2024.

Picaut, J., M. Ioualalen, C. Menkes, T. Delcroix, and M. J. McPhaden, 1996: Mechanisms of the zonal displacements of the Pacific warm pool: Implications for ENSO. *Science,* **274,** 1486–1489.

Syu, H.-H., J. D. Neelin, and D. S. Gutzler, 1995: Seasonal and interannual variability in a hybrid coupled GCM. *J. Climate,* **8,** 2121–2143.

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

——, and D. S. Battisti, 2001: A linear stochastic dynamical model of ENSO. Part II: Analysis. *J. Climate,* **14,** 445–466.

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

Webster, P. J., 1995: The annual cycle and the predictability of the tropical coupled ocean–atmosphere system. *Meteor. Atmos. Phys.,* **56,** 33–55.

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.

——, A. Leetmaa, and M. Ji, 2000: ENSO prediction with Markov models: The impact of sea level. *J. Climate*, **13,** 849–871.

Zebiak, S. E., 1986: Atmospheric convergence feedback of a simple model for El Niño. *Mon. Wea. Rev.*, **114,** 1263–1271.

——, and M. A. Cane, 1987: A model El Niño–Southern Oscillation. *Mon. Wea. Rev.*, **115,** 2262–2278.