A New Method for Determining the Reliability of Dynamical ENSO Predictions

RICHARD KLEEMAN

Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia

ANDREW M. MOORE

Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder, Colorado

(Manuscript received 10 April 1997, in final form 8 May 1998)

ABSTRACT

Determination of the reliability of particular ENSO forecasts is of particular importance to end users. Theoretical arguments are developed that indicate that the amplitudes of slowly decaying (or growing) normal modes of the coupled system provide a useful measure of forecast reliability. Historical forecasts from a skillful prediction model together with a series of ensemble predictions from a ''perfect model'' experiment are used to demonstrate that these arguments carry over to the practical prediction situation. In such a setting it is found that the amplitude of the dominant normal mode, which strongly resembles the observed ENSO cycle, is a potentially useful index of reliability. The fact that this index was generally lower in the 1970s than the 1980s provides an explanation for why many coupled models performed better in the latter decade. It does not, however, explain the low skill of some coupled models in the early 1990s as the index defined here was then moderate.

1. Introduction

Our understanding of the El Niño–Southern Oscillation (ENSO) phenomenon has increased substantially over the past 15 yr thanks to intensive observational, modeling, and theoretical studies. As a result, several research groups now routinely produce forecasts of the eastern/central equatorial Pacific. An excellent compilation of these may be found in the Experimental Long-Lead Forecast Bulletin (NOAA 1997). When tested on a large sample of different initial conditions all the models display some degree of skill and typically correlations with the observations are better than 0.6 for lead times out to 12 or so months. As might be expected, however, there are times in which all models produce inaccurate forecasts. Indeed, there is evidence that there are large variations in skill on a decadal timescale with, for example, Balmaseda et al. (1995) and Chen et al. (1995) showing that the 1970s were in general less predictable than the 1980s. From a user perspective it would be useful if it was possible to identify when less (or more) confidence could be placed on particular predictions.

In the area of weather prediction this problem has been appreciated for quite some time and there are now methods being tested in various operational centers to routinely assess the reliability or otherwise of forecasts. These techniques rely on the generation of a suitable ensemble of predictions that reflect the inevitable uncertainties involved in defining the initial conditions (the analysis) of the weather forecast. The basic idea of the technique is that when the ensemble shows large divergence then less reliance is to be placed on the forecast. To produce an ensemble that accurately reflects the uncertainties of the prediction, certain sophisticated techniques [see Mureau et. al. (1993) and Toth and Kalnay (1993) for two different methods] must be used to perturb the initial conditions. In simple terms this method relies on being able to identify when the flow is particularly unstable (stable) because when it is, error growth (from the initial condition uncertainties) is particularly rapid (slow).

Ensemble techniques have been used in ENSO forecasting but mainly in a fairly ad hoc and unverified fashion. Typically an ensemble of 3–6 predictions separated by a month are examined and the spread is then used as an indication of reliability. Recently Moore and Kleeman (1998) have directly applied one of the sophisticated numerical weather prediction ensemble techniques to the ENSO problem. These authors show, using a large set of historical forecasts from the Bureau of Meteorology Research Centre (BMRC) intermediate coupled model (Kleeman et al. 1995), that there is a direct relationship between ensemble spread and fore-

Corresponding author address: Dr. Richard Kleeman, Bureau of Meteorology Research Centre, GPO Box 1289K, Melbourne, Victoria 3001 Australia.
E-mail: rzk@bom.gov.au

© 1999 American Meteorological Society

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC

cast skill. In particular, they show that when spread is small, skill is almost invariably high, whereas when it is large, skill can range from poor to good. Other authors such as Xue et al. (1994, 1997) have used singular vector methods to examine the predictability of skill.

Clearly, although the above relationships may prove useful, more information on reliability would be desirable. To obtain further reliability indicators, we explore the implications of an interesting paradigm for ENSO, which has recently received much attention in the literature. This views the ENSO phenomenon as the result of the stochastic forcing of a basically nonchaotic dynamical system. The stochastic forcing is provided by the internal variability of the atmosphere such as synoptic transients and the Madden–Julian oscillation. Such atmospheric phenomena are essentially unpredictable beyond a month; therefore, if they significantly project onto the slowly evolving ENSO part of the climate system (as is posited by the paradigm), then a fundamental limitation on predictability emerges. The nature and magnitude of this limitation has been discussed by Kleeman and Power (1994), Kleeman and Moore (1997), and Blanke et al. (1997). The paradigm has also been used as the basis of a skillful statistical prediction scheme by Penland and Magorian (1993) (see also Penland and Sardeshmukh 1995). The observed irregularity of the ENSO cycle has also been explained by this mechanism by Battisti (1988), Blanke et al. (1997), Eckert and Latif (1997), and Moore and Kleeman (1999). In addition, Moore and Kleeman and Blanke et al. have shown that the observed spectrum of the ENSO index NINO3 can be quite well reproduced by the proposed paradigm. Similar results have also been obtained by Kestin et al. (1998) using very long reliable ENSO time series and a simple autoregressive model

What are the implications of this paradigm for variation in predictability? The simplest conceptual picture is that of a noise-forced linear oscillator with a number of normal modes that damp (or possibly even grow) at varying rates. If the first few of these modes are dominant in the sense that they dissipate much more slowly than the other modes, then the amplitude of these modes in the initial conditions of a particular forecast will be especially important for predictability. This follows because these modes (unlike the others) will not be quickly damped back to the ''noise'' level of the system. Thus whether these modes have large amplitudes or not within the initial conditions will determine whether the signal to noise ratio (and therefore skill) is high during the course of a forecast. In a number of intermediate coupled models of ENSO analyzed in the literature, the leading normal modes do indeed seem to dominate the spectrum.[1] In this paper we shall argue therefore that the amplitude of the leading normal mode of the tropical Pacific, which strongly resembles the usual ENSO cycle, should be a useful indicator of forecast skill. In addition we shall give empirical evidence for such a contention. It is worth noting that Roads (1990) has applied similar ideas to the midlatitude atmospheric problem.

The paper is organized as follows: section 2 provides a mathematical justification for using normal mode amplitude as a reliability measure. Section 3 studies both a sample of historical forecasts from an intermediate coupled model together with a much larger sample from a ''perfect model experiment'' to show that the leading normal mode amplitude is useful as a reliability indicator. Finally, section 4 discusses and summarizes the work and suggests further avenues of research.

2. Theoretical framework

For the purposes of developing a theoretical framework we shall assume that the paradigm for ENSO described in the previous section of a stochastically forced dynamical system is a good one. A consequence of this picture is that at any particular time and set of initial conditions for the coupled system there are an infinite set of possible future trajectories for the system. The actual one that occurs depends on the realization of the stochastic forcing.

Consider an ensemble of possible coupled system trajectories commencing from a particular initial condition at time t_0. Suppose that one of the trajectories k is the one actually realized in the future. Match each trajectory from this *basic* ensemble with k for the purposes of calculating a correlation coefficient. Expand these pairings (or basic ensemble) by considering all possible choices for k and refer to this as the *intermediate* ensemble. Finally, expand the pairings further by including all the sets of initial conditions that have the same magnitude for each normal mode of the system[2] and refer to the consequent set of pairings as the *grand* ensemble. These ideas are illustrated schematically in Fig. 1. Consider a particular variable y of dynamical interest defined on the trajectories that is a linear combination of normal modes (an example might be NINO3). The correlation skill of the coupled system for this variable is given by

$$r(y) = \frac{\overline{y_{ji}y_{jk}} - (\overline{y_{ji}})(\overline{y_{jk}})}{\sigma(y_{ji})\sigma(y_{jk})}, \tag{2.1}$$

where the subscript j indexes the set of initial conditions that have equal normal mode amplitudes. These are used to launch the forecasts making up the grand ensemble. Here i refers to the indexing of the basic ensemble and k refers to the indexing of the intermediate ensemble. The overbars are grand ensemble means and σ is the

[1] See section 4 for an analysis of the BMRC model; see Thompson (1998) for a commentary on the spectrum of the Battisti and Cane and Zebiak models; see also Hirst (1986) for a more wide-ranging and theoretical analysis.

[2] Of course we are assuming that the *phase* of the normal mode can be arbitrary.

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC

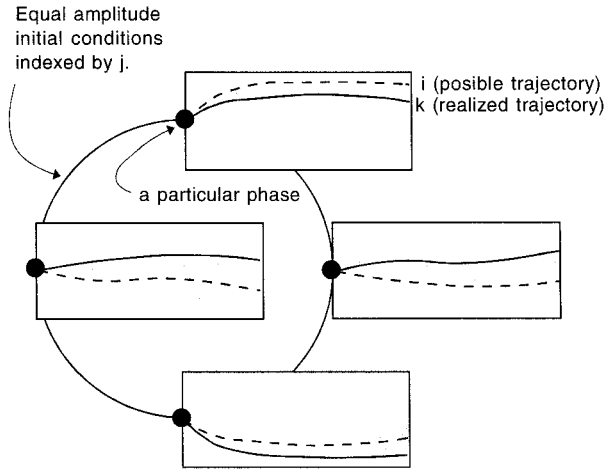

FIG. 1. A schematic to illustrate the meaning of the basic intermediate and grand ensemble of possible trajectories in a stochastically forced dynamical system (see text).

standard deviation of the grand ensemble. The values of k range over the same values as i; therefore, the standard deviation y for all realized trajectories $[\sigma(y_{jk(j)})]$ is identical to the standard deviation of y for all possible trajectories $[\sigma(y_{ji})]$. In addition because of the nature of the grand ensemble in consisting of all possible phases of the normal modes in the initial conditions, it is relatively straightforward to show (see below) that the grand ensemble mean of y is zero. This therefore allows us to rewrite Eq. (2.1) as

$$r(y) = \frac{\overline{y_{ji}y_{jk}}}{\sigma^2(y_{ji})}.$$

If we assume there are N possible values for i and M for j, then we can rewrite the denominator as (using again the fact that the grand ensemble mean of y is zero)

$$\sigma^2(y_{ji}) = \frac{1}{M}\sum_{j=1}^{M}\left[\frac{1}{N}\sum_{i=1}^{N} y_{ji}y_{ji}\right]. \tag{2.2}$$

Also it is easily shown that

$$\sigma_j^2 = \frac{1}{N}\sum_{i=1}^{N} y_{ji}y_{ji} - \mu_j^2, \tag{2.3}$$

where μ_j and σ_j^2 are, respectively, the mean and variance of y for a particular set of initial conditions denoted by j.

If we perform the basic ensemble mean on both the denominator and the numerator and then the mean with respect to k for the numerator, we obtain after use of Eqs. (2.2) and (2.3)

$$r(y) = \frac{\overline{\mu_j^2}}{\overline{\mu_j^2} + \overline{\sigma_j^2}}, \tag{2.4}$$

where we now use the overbar to denote an average with respect initial conditions (i.e., over all values of j). Note that as the mean variance approaches zero the correlation coefficient approaches 1 as it should intuitively since all trajectories converge. Conversely, as this quantity increases the correlation skill drops toward zero, again as expected intuitively.

Consider now a stochastically driven linearized dynamical system in which, to avoid unnecessary complication, the basic state exhibits no temporal variation. Such a system has a complete basis consisting of its normal modes, which are eigenvectors of the propagator matrix A. Such an operator translates the states of the system in time. In this basis it is easily shown that

$$A = A_1 \otimes A_2 \otimes \cdots \otimes A_n,$$

where $\otimes$ denotes the tensor product and where

$$A_j = e^{-\varphi_j}\begin{pmatrix}\cos\phi_j & -\sin\phi_j\\ \sin\phi_j & \cos\phi_j\end{pmatrix} \tag{2.5}$$

if the corresponding eigenvalue is complex or where

$$A_j = e^{-\varphi_j}$$

if it is real. Note that $\varphi_j = t/T_j$ and $\phi_j = 2\pi t/\tau_j$, where $|T_j|$ is the exponential decay/growth timescale (depending on whether T_j is positive/negative) and τ_j is the oscillation period of the normal mode. Note also that the 2D matrix in Eq. (2.5) is simply the rotation matrix. For the case of decaying modes, this formalism will be familiar to users of Principal Oscillation Pattern (POP) analysis (Hasselmann 1988) as this latter technique attempts to model multivariate time series by the system we are considering here.

Suppose now we can write the variable of interest y as a linear combination of the normal modes components:

$$y = a_1^1 e_1^1 + a_1^2 e_1^2 + a_2^1 e_2^1 + a_2^2 e_2^2 + \cdots + a_p e_p + \cdots, \tag{2.6}$$

where e_i^1 and e_i^2 are the components of the real and imaginary parts of the ith normal mode in the state vector of the dynamical system (the superscript is dropped in the case of a real normal mode). Note that purely for argument's sake we are assuming that the first two eigenvalues are complex and the pth is real. Further, as we are interested in initial conditions in which the magnitude of each normal mode is invariant, the initial condition state vector may be written as

$$(r_1\cos\alpha_1,\ r_1\sin\alpha_1,\ r_2\cos\alpha_2,\ r_2\sin\alpha_2,\ \cdots,\pm r_p,\ \cdots),$$

where r_i is the amplitude of the ith normal mode in the initial conditions and α_i is the corresponding phase. Now the ensemble mean at time t is simply the matrix product of the propagator and the initial conditions (see Kleeman and Moore 1997). It follows therefore that

$$\begin{aligned}\mu_j = {} & r_1\exp(-\varphi_1)[a_1^1\cos(\alpha_1+\phi_1) + a_1^2\sin(\alpha_1+\phi_1)]\\ & + r_2\exp(-\varphi_2)[a_2^1\cos(\alpha_2+\phi_2) + a_2^2\sin(\alpha_2+\phi_2)]\\ & + \cdots + r_p\exp(-\varphi_p)a_p + \cdots.\end{aligned}$$

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC

Integrating this expression over a full cycle for each complex normal mode and summing over the ± cases for each real mode, we obtain the unsurprising result (used above) that the grand ensemble mean of y is zero. On the other hand, the grand ensemble mean of μ_j^2 is nonzero and is easily shown to be given by

$$\overline{\mu_j^2} = 0.5r_1^2 \exp(-2\varphi_1)[(a_1^1)^2 + (a_1^2)^2] + 0.5r_2^2 \exp(-2\varphi_2)[(a_2^1)^2 + (a_2^2)^2] + \cdots + r_p^2 \exp(-2\varphi_p)(a_p)^2 + \cdots. \quad (2.7)$$

Note that this calculation is implicitly assuming that the phases of each normal mode are uncorrelated, otherwise one would not be able to perform independent integrations with respect to each full cycle of every normal mode (the integrations effectively represent sample summations). We shall test this assumption using a ''perfect model'' experiment in section 3 below.

The calculation of $\overline{\sigma_j^2}$ is more complex, in general, because it depends on the nature of the forcing noise and how it projects onto the normal modes. Kleeman and Moore (1997) demonstrate however that it *does not* depend on the initial conditions, that is, on the r_j and α_j. Moreover for a skillful ENSO forecast model and for NINO3 it tends to exhibit fairly characteristic behavior—a rapid increase over 6 months and quite slow growth subsequent to this. Such a conclusion has also been reached by Blanke et al. (1997), Eckert and Latif (1997), and Stockdale (1996, personal communication), who use quite different models and estimates of the stochastic forcing of the coupled system.

Now we may rewrite Eq. (2.4) as

$$r(y) = \frac{1}{\dfrac{\overline{\sigma_j^2}}{\overline{\mu_j^2}} + 1}, \quad (2.8)$$

which shows that skill varies monotonically with $\overline{\mu_j^2}$. Consideration of Eq. (2.7) allows us therefore to draw three significant conclusions.

1) Skill depends on the amplitude of the normal mode present in the initial conditions. The larger this amplitude, the higher the skill.
2) As the length of the forecast increases, the least-damped normal modes become more important to the skill. The damping time of a normal mode sets a timescale for how long it usefully[3] contributes to forecast skill.
3) Skill depends on the degree to which a particular normal mode present in the initial conditions projects onto the index used to assess skill.

It should be stressed that these conclusions are dependent on the simple conceptual model used here and the assumptions upon which it is based (e.g., that of uncorrelated phase).

Equation (2.8) provides a convenient reference point to discuss the complementary nature of the present study to our earlier work (Moore and Kleeman 1998). The previous study effectively relied on variations in the instability of the initial conditions to define a measure of the reliability of forecasts. In the present conceptual model this corresponds to using σ_j^2 as a measure of skill since this measures the spread of an ensemble of possible trajectories from a particular set of initial conditions specified by the subscript j. Here we are ignoring such variations and instead concentrating, roughly speaking, on the size of significant coupled oscillations present in the initial conditions. This amplitude is measured by μ_j^2. Our earlier studies show that we may expect σ_j^2 to depend on the phase of ENSO (at least in the model used) and thus because the dominant normal mode resembles ENSO (see below) it is quite possibly a function of j. Varying the amplitude of this normal mode intuitively does not a priori change this instability [see Battisti and Hirst (1989) where modeling evidence is presented that ENSO is weakly nonlinear] even though varying the phase may influence it. Thus we seek to capture additional information on predictability that is not available through studies of the instability of the coupled system.

3. Evidence from observations and model simulations

The major conclusion from the previous section is that there should be a direct relationship between a coupled model's forecast skill and the amplitude of the dominant ENSO normal mode present in the initial conditions of the forecast. As a first test of this hypothesis, we examined the hindcasts from the BMRC intermediate coupled model (see Kleeman et al. 1995). This model is described in considerable detail in the cited source and in publications cited within this source. Briefly, the model consists of shallow water dynamical formulations of both atmosphere and ocean. The atmospheric model has a simple convection parameterization using a moisture equation, whereas the oceanic SST responds only to thermocline perturbations. Coupling is via the exchange of wind stress and SST only. Hindcasts are currently available for the period 1971 through to the present commencing every 3 months. The amplitude of the normal mode was obtained directly from observations (see Smith 1995) of subsurface thermal data (specifically, the upper 400-m averaged temperature anomaly for 1970–present) by use of POP analysis (see Hasselmann 1988). To eliminate high-frequency noise in the data (due predominantly but not exclusively to missing observations) an 8-month running mean was applied to the data. We chose to consider only the heat content portion of the coupled system as this is where the pre-

[3] This might be defined as the the point at which the modal contribution to skill [as exemplified by Eq. (2.8)] falls below a predefined level.

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC

TABLE 1. POP temporal characteristics.

POP no.	Period (months)	Decay time (months)
1	39.437	69.1
2	55.83	61.67
3	75.74	57.69
4	124.43	51.50
5	300.24	121.51

dictability of the system mainly lies and in any case it is not completely obvious how one should combine other variables (e.g., wind and SST) with heat content. The results we obtained are, however, very similar to those obtained in a combined analysis by Latif et al. (1998) and are also quite similar to the coupled model normal mode (see below).

An important (and often overlooked) aspect of normal modes (and POPs) in most realistic dynamical systems is their nonorthogonality. This complicates the calculation of their amplitude. Fortunately, if the adjoint normal modes are available this is easily overcome because adjoint and conventional normal modes form a biorthogonal set. We therefore calculated in addition to the POPs the so-called Adjoint Principal Oscillation Pattern (APOPs) and then calculated the amplitude of the corresponding POP in the observations by taking the inner product of the APOP with the particular observation. The period and decay time for the first five POPs are displayed in the table. They are grouped according to the ratio of decay time to period, which is seen as a measure of their veracity (see von Storch et al. 1993).

Limitations of the POP analysis are discussed in detail elsewhere (von Storch et al. 1993), but there are clearly sampling problems in defining accurately some of the long period modes here given that the time series is only of order 300 months. The dominant ENSO POP (no. 1) can be easily picked out as the mode having a period of around 3–4 yr and by far the slowest decay timescale relative to its period. Figure 2 displays both real and imaginary components of this POP together with a time series of its amplitude in the observations. Also displayed for reference in Figs. 2d and 2e are the spatial patterns of the APOP, which were used to calculate the POP amplitudes—note the lack of similarity between the POP and its adjoint. A number of features from Fig. 2 are noteworthy: first, Fig. 2a [the real part of the POP corresponds well with the usual ENSO EOF of upper-level heat content (see Kleeman et. al. (1995)] with a dipolar zonal structure involving a western Pacific ''Rossby-like'' response of one sign and an eastern Pacific ''Kelvin-like'' response of the opposite sign. The imaginary part, Fig. 2b, has a large-amplitude signal of the one sign located mainly in the equatorial central Pacific. A similar pattern has also been noted by Latif et al. (1998), who used SST and wind stress as well as upper-level heat content in their analysis. This pattern corresponds with the idea of a heat reservoir buildup prior to El Niño as postulated by Wyrtki (1975). The amplitude of the POP may be seen in Fig. 2c and it is evident that there is some large and interesting variability: the periods 70–75, 81–83, and 87–88 were periods of generally high amplitude, whereas the late 1970s and parts of the 1990s showed small amplitudes. We return to this significant temporal variability below.

To obtain reasonably large forecast samples and hence statistically reliable results for skill, we chose to stratify hindcasts fairly coarsely by POP amplitude. In addition we considered the first 12 months of the hindcast as a block in order to further increase sample size.[4] The particular sample chosen does not qualitatively affect the results obtained. The period 6–12 months as well as 12–24 months produced similar results albeit with somewhat lower skill.

Results are shown in Fig. 3, which displays the anomaly correlation and rmse skill for NINO3 as a function of POP amplitude. Also displayed is the sample size for each stratification. As can be seen, a clear relationship with correlation skill is evident with poor forecasts associated with low-amplitude initial conditions and conversely for good skill. The monotonic nature of this relationship is also quite striking. There is a similar improvement in rmse with increasing POP amplitude although the effect here is significantly less pronounced for reasons explained in Kleeman (1993). These results potentially have practical implications for ENSO forecasting since the rmse provides error bars for forecasts. This is significant for the following reason: when the amplitude of the dominant normal mode is large in the initial conditions then because certain phases of this mode project strongly onto NINO3, we may expect a rather large NINO3 value (positive or negative) at some time into the forecast. Conversely, a small initial modal amplitude will imply (everything else being equal) a small NINO3 signal during a forecast. Since the error bars are relatively constant with modal amplitude, it is clear that more importance will be attached by a forecaster to the former forecast than the latter.

It is also interesting to stratify the hindcasts by time. We chose the 4-yr periods 1971–74, 1977–80, 1981–84, 1986–89, and 1990–93. Sample sizes here are a little small so caution needs to be exercised in interpreting results. Results are displayed in Fig. 4, where it is evident that the early 1970s and the 1980s were periods of high skill *and* high POP amplitude. The late 1970s, conversely, had low skill and low POP amplitude. The 1990s were somewhat contradictory in showing low skill but moderate–high POP amplitude.

These results may therefore help to explain why there is significant variation from decade to decade in ENSO forecast skill. A viewing of Fig. 2c shows clearly that the 1970s were overall a period of significantly smaller

[4] All 12 months from a given hindcast are far from independent so the effective sample size is perhaps doubled or tripled rather than increased by an order of magnitude by this strategy.

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC

ENSO POP amplitude than the 1980s and consistent with this a number of ENSO coupled models have significantly higher skill for the latter decade.

The results discussed so far are limited by the relatively small sample of real hindcasts. To gain more detailed information and also to see how the skill–mode amplitude relationship might hold in a ''perfect model'' situation, it was decided to use simulated observations by forcing the coupled model with stochastic input. As discussed in Kleeman and Moore (1997), such forcing represents the effects of atmospheric transients on the coupled model.[5] As has been noted by Kleeman and Moore (1997) and Moore and Kleeman (1999), the forcing can be applied most efficiently by forcing with spatial patterns corresponding to the dominant ''stochastic optimal'' of the coupled model. When this is done only very modest wind stress and heat flux perturbations are required in order to produce very realistic long term behavior. This is illustrated in Fig. 5, which shows NINO3 for an extended integration of 73 yr together with the corresponding spectrum. Both compare very well with the real observations (see Blanke et al. (1997). It should be noted that without stochastic input the coupled model exhibits decaying regular oscillations. The behavior in Fig. 5 is relatively insensitive to the size of the forcing applied: If the magnitude is doubled or halved there is little change in character.

To gain an estimate of the skill of forecasts at any particular starting point in the integration displayed in Fig. 5, an ensemble of 11 possible forecasts was calculated at 3-month intervals for the first 52 yr. Ensemble members were obtained by forcing 12-month integrations of the coupled model with different realizations of the stochastic forcing. Note that all ensemble members have identical initial conditions. The ensemble so constructed corresponds to the *basic* ensemble discussed in the previous section. Physically it represents the set of all possible evolutions from a particular initial condition. Skill was calculated under the assumption that one of the ensemble members (the original before generation of the ensemble) corresponded to the ''real world realization'' (the ensemble member labeled by k in the notation of the previous section).

In the present context it was decided to use the model ENSO normal mode amplitude rather than the estimation of the real normal mode amplitude, which is the POP amplitude. This was considered more consistent with the ''perfect model'' nature of the experiment. The relationships displayed below still held when the POP amplitude was used but were somewhat weaker.

The ENSO normal mode and its adjoint were calculated using a linearization of the coupled model about an annually averaged basic state. Increased coupling strength was used to clearly separate the mode from other high wavenumber ocean-only modes present.[6] A mode of almost identical spatial structure could be ''traced back'' into these modes as the coupling strength was reduced. Figures 6a and 6b show the spatial patterns for this mode and these should be compared with Figs. 2a and 2b. As can be seen the real part of the model normal mode is very similar to the observed POP whereas the imaginary part is qualitatively similar in having a strong central Pacific warming, but some differences are evident particularly in off-equatorial regions and the far western Pacific.

The ENSO normal mode amplitude was calculated using its adjoint and may be seen in Fig. 6c, where we note as in the real case discussed above there are significantly long periods in which this amplitude is both small and large (see, e.g., years 58–64 and 65–70). Note that generally speaking, the amplitude of the mode is high when a regular oscillation in NINO3 occurs and low when this is absent. Using the basic ensemble, it is possible to calculate an estimate of ''maximal'' possible skill for predictions from a particular initial condition. Maximal here means this is the skill we would expect in the real world if the model was perfect. Skill was calculated as outlined in the previous section.

The normal mode amplitude and skill (calculated as described above) are plotted in a scatter diagram in Fig. 7. Such a plot gives us some idea of how strong the relationship between these two quantities may be in ideal circumstances. Clearly there is a strong direct relationship between the two quantities with the linear correlation coefficient being 0.62. The relationship is less than perfect because of the presence of other factors influencing skill such as the varying stability of the background state (see Moore and Kleeman 1998) and also because of the still fairly small sample size used here. Interestingly, there are indications that the relationship displayed is nonlinear with the skill rapidly reaching a high value and then saturating as amplitude increases. This effect is currently under investigation.

Finally, we use the perfect model experiment discussed here to demonstrate the reasonableness of the assumption made in the previous section that the phases of the normal modes are correlated only to a small and unimportant degree. This was done by calculating the phases of the first 50 normal modes at intervals of one month through the 52 yr of perfect model integration. The phases were then intercorrelated and the results examined for normal modes having a significant projection onto NINO3 [cf. the coefficients a_j^i in Eq. (2.6)]. None of the correlations exceeded 0.08, which is approximately the level at which a correlation would occur by chance in a sample of 50.

[5] They are not included in the coupled model intrinsically because of its low-order nature (only the equilibrium response of the tropical atmosphere to SST anomalies is modeled).

[6] These modes are predominantly internal to the ocean and contribute little to SST and even less to NINO3 and so are of little relevance to the ENSO prediction problem.

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC

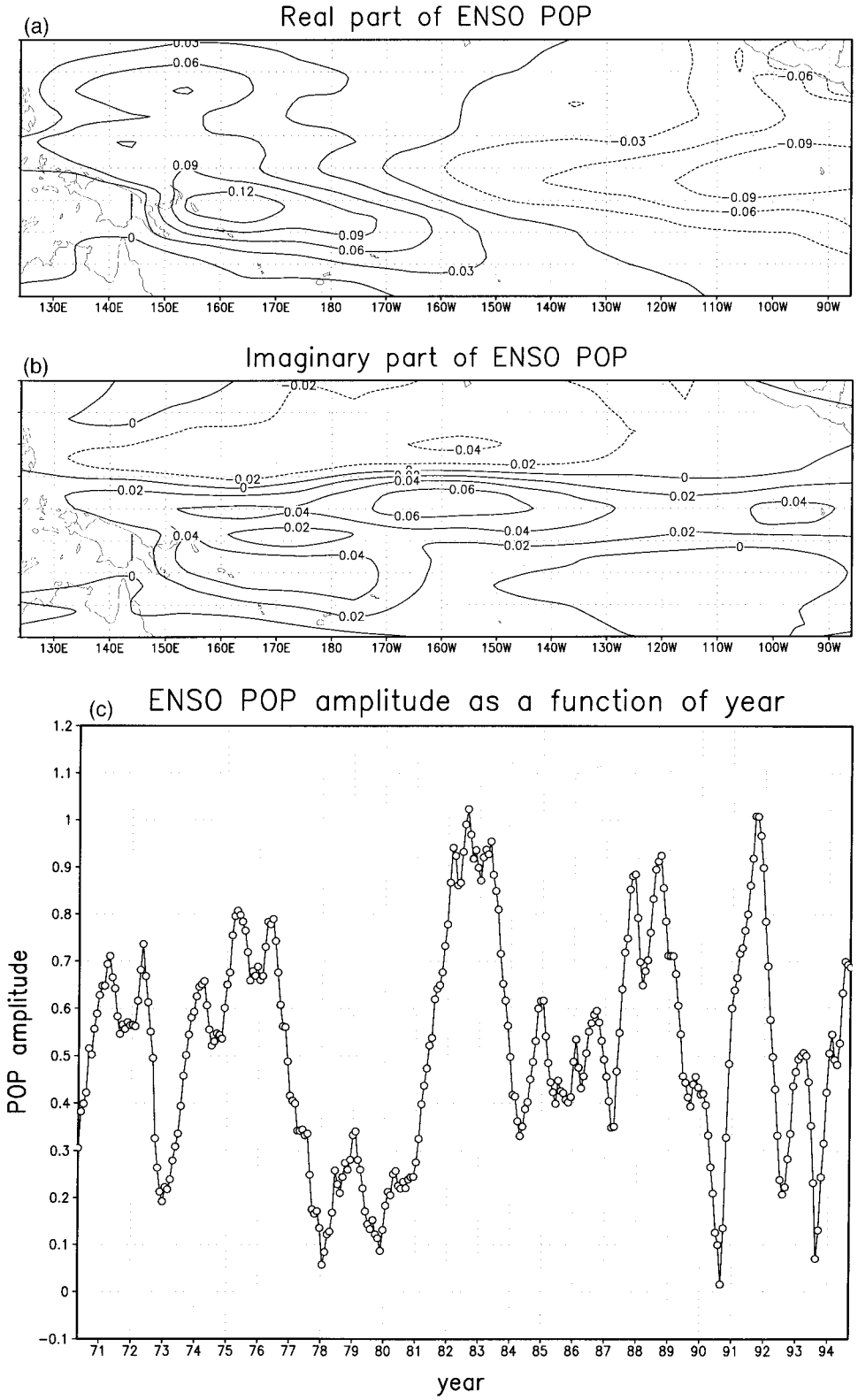

FIG. 2. (a) The spatial pattern of the real part of the ENSO POP. Contour interval is 0.03 dimensionless units. (b) Identical to (a) but for the imaginary part of the POP. Contour interval is 0.02. (c) Amplitude of

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC

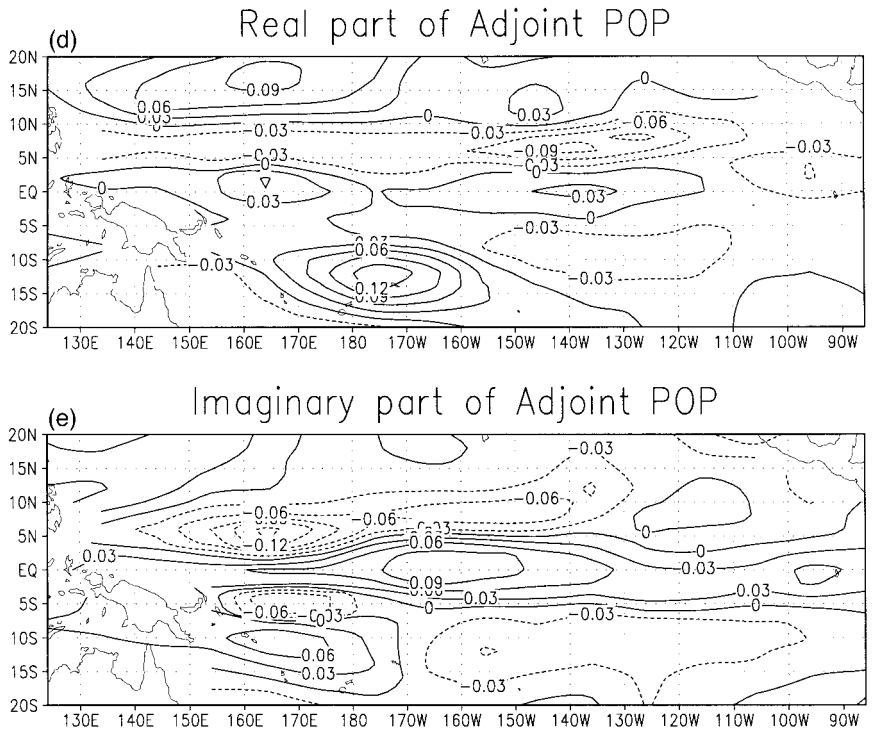

FIG. 2. (*Continued*) the ENSO POP as a function of year. Units are degrees Celsius. (d) The same as (a) but for the APOP. (e) The same as (b) but for the APOP.

It is worth observing that the uncorrelated nature of normal mode phases demonstrated here for the intermediate coupled model may not hold as well in the real dynamical system, and so we need to be cautious in interpreting the observed (as opposed to the perfect model) skill relationships using the developed theoretical framework. It is not possible to perform the same correlation analysis on the observed POPs (as opposed to the model normal modes) because of sampling issues (see above), so we cannot directly address this possible limitation to our theoretical assumption. Nevertheless the fact that it holds in a model that has demonstrated skill in predicting ENSO both historically and operationally and in addition produces variability with strong resemblance to that observed gives credence to our framework.

4. Summary and discussion

Theoretical and empirical evidence has been presented that indicates that the amplitude of the ENSO normal mode or POP in the initial conditions of a particular forecast is a good indicator of skill. The basic reason for this situation appears to be that this mode

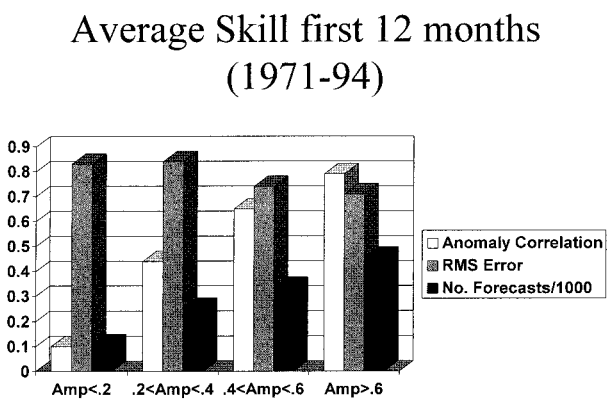

FIG. 3. Prediction skill measures as a function of POP amplitude (see text).

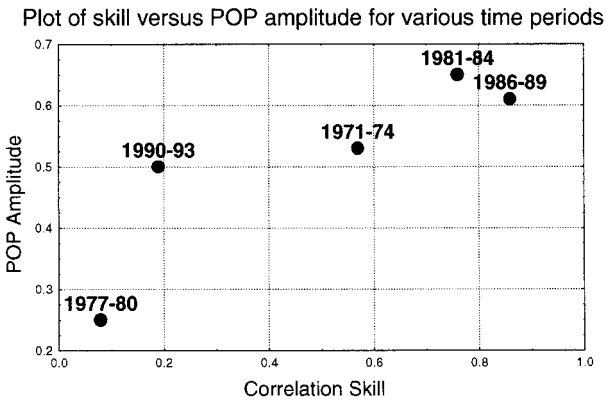

FIG. 4. Skill and POP amplitude for hindcasts from particular time periods.

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC

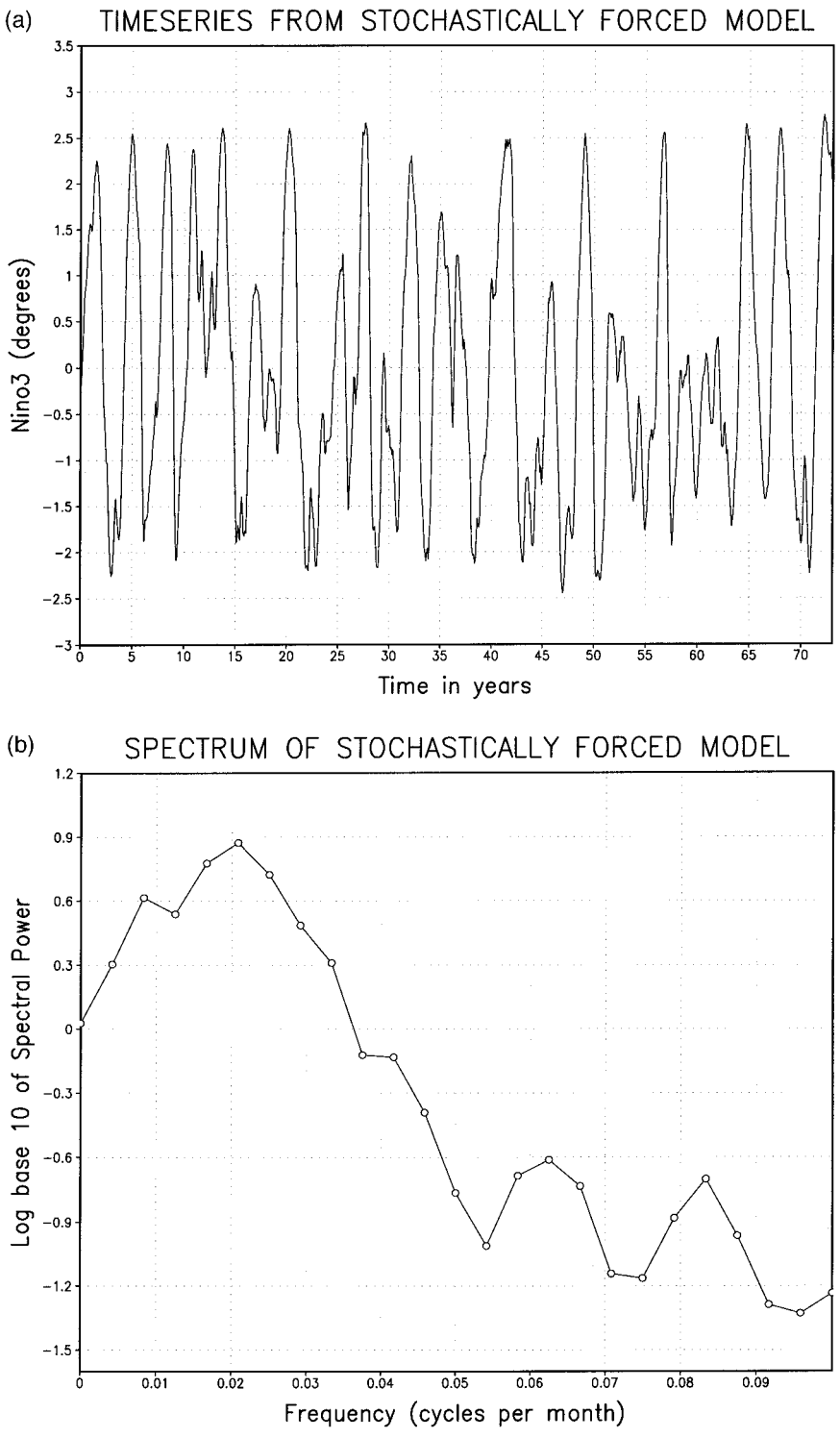

FIG. 5. (a) A time series of NINO3 from a stochastically forced run of the BMRC intermediate coupled model. Units are degrees Celsius. (b) Power spectrum of the time series from (a).

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC

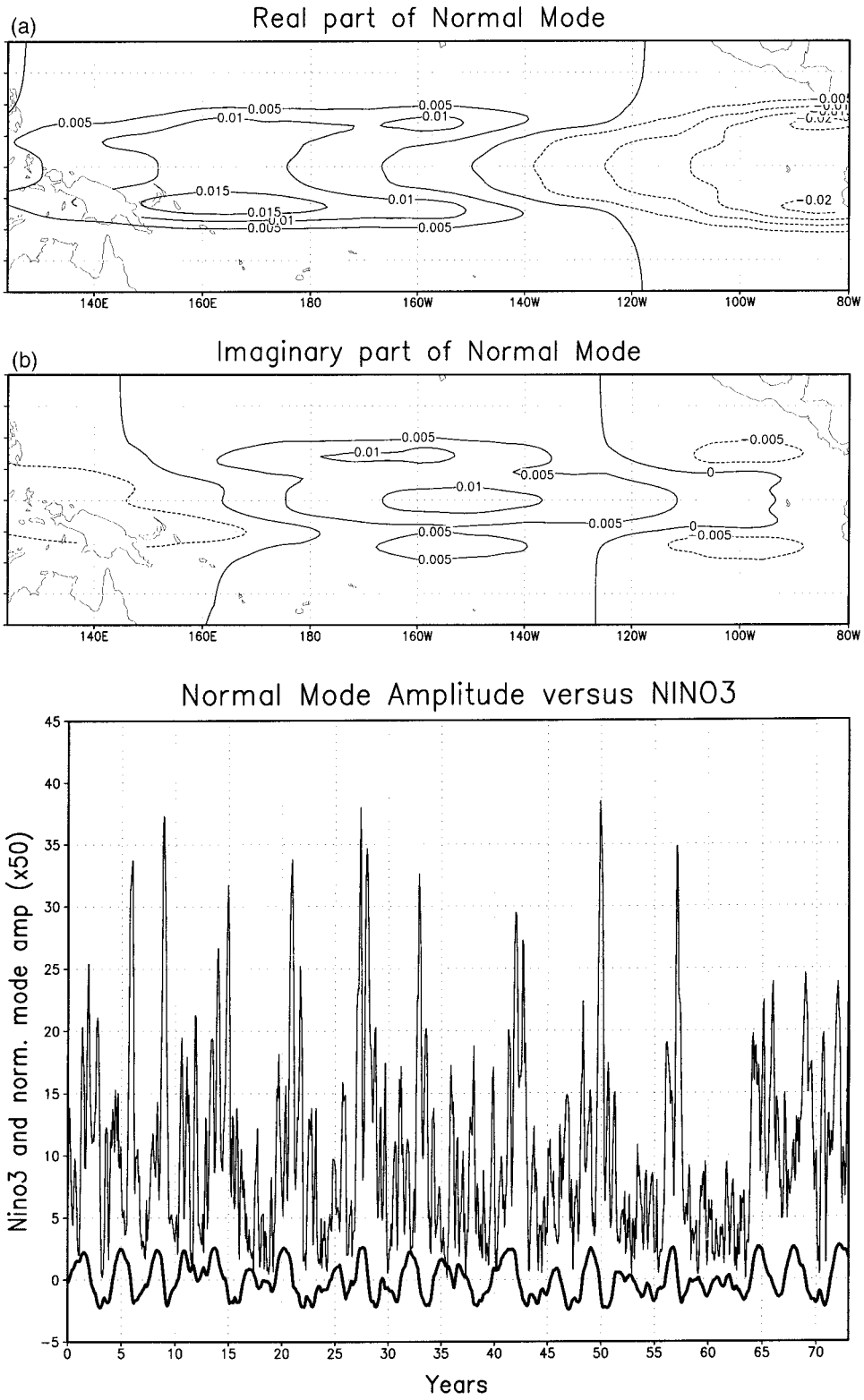

FIG. 6. (a) The spatial pattern of the real part of the dominant model normal mode. Contour interval is 0.005 dimensionless units. (b) Identical to (a) but for the imaginary part of the normal mode. (c) The amplitude (multiplied by 50) of the dominant normal mode as a function of time (thin line). Also plotted is the NINO3 values for the same time (thick line).

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC

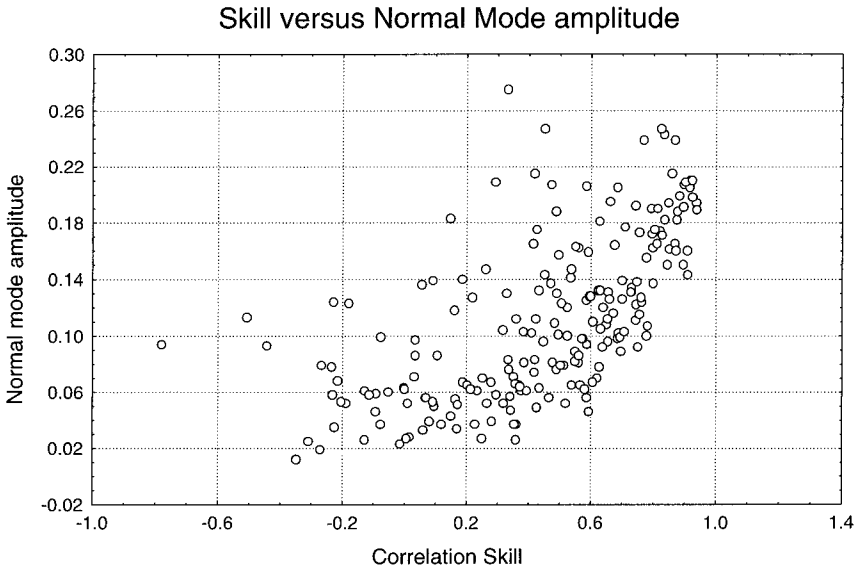

FIG. 7. A scatterplot of the ensemble skill and initial normal mode amplitude of forecasts launched every 3 months (see text).

decays slowly in the coupled system and thus the ''signal to noise ratio'' is maintained for quite some time into a forecast. Other modes present in the initial conditions that may be significant in causing SST anomalies are apparently quickly damped and hence are not useful to skill.

What evidence is there from observations regarding the leading normal mode and its damping timescale? Penland and Magorian (1993) conducted a POP analysis of SST data (1950–90) alone and concluded that POP modes with structure similar to the ENSO cycle had a damping timescale of 9 months or so. On the other hand Latif et al. (1998) conducted a POP analysis using SST, upper-ocean heat content, and wind stress (1982–92) and found that the ENSO POP only damped on a timescale of 36 months or so. This was also roughly the decay time obtained in the POP analysis conducted in section 3. The latter results would seem more consistent with the very high levels of skill shown in the second year by dynamical models during the 1980s. A possible reason for the discrepancy is that the inclusion of oceanic memory in the form of heat content acts to reduce the decay of the signal.

Results from coupled models have shown that for realistic choices of parameters, quite a number are close to what might be loosely termed a primary bifurcation (see, e.g., Kleeman 1993; Perigaud et al. 1997) (intermediate models), Blanke et al. (1997) (hybrid model), and Stockdale (1997, personal communication) (coupled GCM). By primary bifurcation we mean the point in parameter space (e.g., coupling strength) at which the model starts exhibiting self-sustaining oscillations. The coupled models mentioned either exhibit slowly decaying oscillations or self-sustaining oscillations, which are easily converted to slowly decaying oscillations by minor model parameter retuning. Such a result provides evidence that the leading ENSO normal mode is marginally stable or weakly unstable. This is consistent with the theory discussed here, because if the mode was sharply decaying we would not expect coupled models to exhibit much skill at ENSO prediction beyond this short timescale.

Evidence was also presented that the ENSO normal mode amplitude varies significantly on decadal timescales and thus the effect discussed here may be a significant contributor to the decadal variations in ENSO predictive skill previously noted in the literature. In particular it was seen that there were three periods of high skill and high normal mode amplitude (the early 1970s, the early 1980s, and the late 1980s), whereas there was one period of low amplitude and low skill (the late 1970s). The early 1990s, which has been a period in which skill in many coupled models (including the one used here) has been low, showed moderate to high average amplitude. This may, however, be due to the systematic improvement in the observing system throughout the 26 yr of subsurface data. This bias occurs because the analysis method relaxes toward climatology (and hence toward zero anomaly) in the absence of data. Thus when there are fewer observations than desirable, the amplitude of the normal mode in the analysis will tend to be underestimated relative to a case where there are sufficient observations to clearly define the normal mode. It should be noted also that there were significant periods during the 1990s when ENSO modal amplitude was very small by historical standards. Other reasons for the drop in skill may be the significant influence of unmodeled decadal variability [see Kleeman et al. (1996) and Latif et al. (1997)].

A problem that requires further investigation is the

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC

more precise definition of normal mode amplitude in the initial conditions of real forecasts. Although the POP method was useful in demonstrating the potential of the method proposed here, it clearly has limitations. These include the fact that sampling issues limit its accuracy (see previous section) and the fact that an 8-month filter was required to remove spurious noise from the observations. This latter operation means that more sophisticated estimation techniques will be required for real-time forecast analysis. All this reflects the difficulty of estimating a complex dynamical quantity such as normal mode amplitude from limited real-time data.

It is worth comparing the reliability indicator developed here with that previously developed by Moore and Kleeman (1998). This previous indicator relied on calculating the ensemble spread, which is clearly related to the degree of instability in the initial conditions of the forecast. Intuitively, the indicators are measuring two quite different aspects of the dynamical system (signal size and instability), and so one might expect that their combined use would further enhance reliability estimates. Preliminary results using the perfect model design described in the previous section confirm that there is sufficient independent information to significantly improve on the individual usage of either indicator. Further experiments are planned in this vein and it is hoped that an optimal set of reliability indicators can be developed. A closely related issue is the generalization of the formalism described in section 2 to the case of a varying background state (we have assumed a constant state for simplicity in this paper). This generalization is currently the subject of intensive investigation.

Acknowledgments. One of the authors (R.K.) wishes to thank Jorgen Frederiksen for useful feedback. We would also like to thank Prof. Dan Sorensen at Rice University for providing the software for calculating the normal mode and its adjoint described in section 3.

REFERENCES

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

Battisti, D. S., 1988: Dynamics and thermodynamics of a warming event in a coupled tropical atmosphere–ocean model. *J. Atmos. Sci.,* **45,** 2889–2929.

——, and A. C. Hirst, 1989: Interannual variability in the tropical atmosphere–ocean system: Influence of the basic state and oceanogeometry. *J. Atmos. Sci.,* **46,** 16–44.

Blanke, B., J. D. Neelin, and D. Gutzler, 1997: Estimating the effect of stochastic wind stress forcing on ENSO irregularity. *J. Climate,* **10,** 1473–1486.

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

Eckert, C., and M. Latif, 1997: Predictability of a stochastically forced hybrid coupled model of El Niño. *J. Climate,* **10,** 1488–1504.

Hasselmann, K., 1988: PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.,* **93** (D9), 11 015–11 021.

Hirst, A. C., 1986: Unstable and damped equatorial modes in simple coupled ocean–atmosphere models. *J. Atmos. Sci.,* **43,** 606–630.

Kestin, T. S., D. J. Karoly, J.-I. Yano, and N. A. Rayner, 1998: Time-frequency variability of ENSO and stochastic simulations. *J. Climate,* **11,** 2258–2272.

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

——, and S. B. Power, 1994: Limits to predictability in a coupled ocean–atmosphere model due to atmospheric noise. *Tellus,* **46A,** 529–540.

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

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

——, R. Colman, N. R. Smith, and S. B. Power, 1996: A recent change in the mean state of the Pacific basin climate: Observational evidence and atmospheric and oceanic responses. *J. Geophys. Res.,* **101,** 20 483–20 499.

Latif, M., and Coauthors, 1998: A review of the predictability and prediction of ENSO. *J. Geophys. Res. (Oceans),* **103,** 14 375–14 393.

——, R. Kleeman, and C. Eckert, 1997: Greenhouse warming, decadal variability, or El Niño? An attempt to understand the anomalous 1990s. *J. Climate,* **10,** 2221–2239.

Moore, A. M., and R. Kleeman, 1998: Skill assessment for ENSO using ensemble prediction. *Quart. J. Roy. Meteor. Soc.,* **124,** 557–584.

——, and ——, 1999: Stochastic forcing of ENSO by the intraseasonal oscillation. *J. Climate,* in press.

Mureau, R., F. Molteni, and T. N. Palmer, 1993: Ensemble prediction using dynamically-conditioned perturbations. *Quart. J. Roy. Meteor. Soc.,* **119,** 299–323.

NOAA, 1997: Experimental long-lead forecast bulletin. Rep. 1, 78 pp. [Available from Climate Prediction Center, W/NMC51 Room 604, Washington, DC 20233.]

Penland, C., and T. Magorian, 1993: Prediction of NINO3 sea surface temperature using linear inverse modeling. *J. Climate,* **6,** 1067–1076.

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

Perigaud, C., S. E. Zebiak, F. Melin, J.-P. Boulanger, and B. Dewitte, 1997: On the role of meridional wind anomalies in a coupled model of ENSO. *J. Climate,* **10,** 761–773.

Roads, J. O., 1990: Linear model predictions of time averages. *J. Climate,* **3,** 317–336.

Smith, N. R., 1995: An improved system for tropical ocean subsurface temperature analyses. *J. Atmos. Oceanic Technol.,* **12,** 850–870.

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

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

von Storch, H., A. Bürger, R. Schnur, and J.-S. von Storch, 1993: Principal oscillation patterns. Max Planck Institut für Meteorologie Rep. 113, Hamburg, Germany.

Wyrtki, K., 1975: Fluctuations of the dynamic topography in the Pacific Ocean. *J. Phys. Oceanogr.,* **5,** 450–459.

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

Unauthenticated | Downloaded 12/09/24 05:52 PM UTC