Stochastic theories for the irregularity of ENSO

BY RICHARD KLEEMAN *

Courant Institute of Mathematical Sciences, New York, NY 10012, USA

The El Niño/Southern Oscillation (ENSO) phenomenon is the dominant climatic fluctuation on interannual time scales. It is an irregular oscillation with a distinctive broadband spectrum. In this article, we discuss recent theories that seek to explain this irregularity. Particular attention is paid to explanations that involve the stochastic forcing of the slow ocean modes by fast atmospheric transients. We present a theoretical framework for analysing this picture of the irregularity and also discuss the results from a number of coupled ocean–atmosphere models. Finally, we briefly review the implications of the various explanations of ENSO irregularity to attempts to predict this economically significant phenomenon.

Keywords: ENSO; stochastic; irregular; predictability; climate oscillation

1. Observational background

On time scales ranging from the annual to the decadal, the dominant form of variability within the climate system is the El Niño/Southern Oscillation (ENSO) phenomenon whose centre of action is in the equatorial central and eastern Pacific. The variability is highly coherent spatially in both ocean and atmosphere and tends to be dominated by a relatively small number of large-scale patterns in both media.

Observations of ENSO have increased significantly in the past century as global shipping has become more prevalent and specialized oceanographic platforms have been put in place to monitor this important climatic effect. Prior to ca 1940, observations of both atmosphere and ocean in the equatorial Pacific were generally quite sparse and historical reconstructions of variables such as sea surface temperature (SST) for that period rely heavily on observed historical global teleconnections in order to infer the equatorial patterns of variation. The standard EOF analysis reveals that, for ENSO, only a small number of large-scale patterns are required to explain much of the variance for many climatically important variables. Furthermore, the fact that such patterns are highly coherent in time between different climatic variables can assist considerably in the historical reconstruction of the dominant aspects of ENSO.

Based on the data from the more recent and reliable era, it is clear that ENSO is a broadband phenomenon with a spectral peak around 4 years. Figure 1 displays a spectral analysis by Blanke et al. (1997) of a widely used SST index (NINO3) of the eastern Pacific for the period 1950–1993.

*kleeman@cims.nyu.edu

One contribution of 12 to a Theme Issue ‘Stochastic physics and climate modelling’.
The SST data can be extended back in time to the mid-nineteenth century using global EOFs to fill data-sparse regions that are prominent in the deep tropics before 1950 (e.g. Rayner et al. 2003). While the basic qualitative structure of the spectrum does not change much (see below), it is interesting to note, however, that there are significant decadal variations in the details. This is illustrated in figure 2, adapted from Torrence & Webster (1999), where a wavelet analysis technique has been used to extract time variations in the spectrum over the past 100 years or so. Given the apparently rather significant decadal variations, one must show a little caution in assigning the reliable but short data in figure 1 a definitive status. Note also that well-defined, rather narrow spectral peaks can occur in the data from time to time but are not persistent over the entire record.

A particularly robust method for extracting the dominant spectral character of ENSO has been performed by Kestin et al. (1998). Here, three completely independent data sources have been carefully compiled back to the late nineteenth century. These are derived from respectively reconstructed SST

![Figure 1](image1.png)

Figure 1. Spectrum of the monthly mean NINO3 SST index for ENSO for the period 1950–1993. Adapted from the work of Blanke et al. (1997).

![Figure 2](image2.png)

Figure 2. Wavelet analysis of the time-varying spectrum of the NINO3 SST index. Adapted from the work of Torrence & Webster (1999).
data, station measurements of atmospheric pressure (the Southern Oscillation Index) as well as rain gauge data from selected equatorial islands. These indices exhibit very high temporal coherence with a 21-year moving window correlation coefficient fluctuating between 0.6 and 0.8 during the length of the time series. Interestingly, the low-correlation periods correspond to quiescent periods for ENSO. The three time series and their spectra\(^1\) are displayed in figures 3 and 4, respectively, and, while there are some minor differences, the overall common spectral shape is quite clear and qualitatively close to the spectra calculated using 45 years of recent reliable data.

Another feature of ENSO, which is very well established from many observational studies (e.g. Rasmusson & Carpenter 1982), is the phase locking to the seasonal cycle of warm events. These tend, in general, to occur towards the end of the calendar year. While there has been much discussion in the literature concerning the details of this phase locking (including the individual character of events and potential decadal variability), this basic aspect of the ENSO cycle appears beyond doubt.

In summary, both the spectrum and phase locking of the dominant mode of ENSO variability appear reasonably well established from the existing data. A theory of the irregularity of ENSO should be able to explain these important features. One would also hope that it should explain some of the decadally varying features of the phenomenon.

\(^1\)The latter is unpublished but provided to the author by T. Kestin (2000, private communication).

Figure 3. Long time series of important ENSO indices (see text). SOI, Southern Oscillation Index. Adapted from Kestin et al. (1998).
2. Theories for ENSO irregularity

The present theoretical understanding of ENSO is that it derives fundamentally from a coupling of the tropical ocean–atmosphere system. A comprehensive (but somewhat dated) review of the dynamical understanding of ENSO and coupled variability may be found in the excellent review by Neelin et al. (1998). The tropical atmospheric variability on interannual time scales in the Pacific can, in this view, be considered to be in approximate equilibrium with the underlying ocean temperature (SST). On shorter time scales, however, moist convective instabilities are widely believed to play a crucial role in generating atmospheric variability. It seems natural therefore to divide the tropical dynamical system into fast and slow time scales associated with these two atmospheric forcing mechanisms. The slow time scale derives from the relaxation time of the equatorial ocean basin (modified by the atmospheric coupling), while the fast time scale has its genesis in the life-cycle times of various atmospheric convective disturbances.

An interesting feature of the model development in the ENSO area has been the historical importance of simple models. In general, they have given good simulations of important aspects of ENSO and have also exhibited useful levels of prediction skill when tested on historical data (e.g. Cane et al. 1986; Barnett et al. 1993; Kleeman 1993). Such simple models have been characterized for many years by atmospheric components that are in equilibrium with the SST produced by their ocean component. Thus, in terms of the terminology introduced above, such models are entirely ‘slow’. They have traditionally taken two forms: ‘intermediate’ and ‘hybrid’. The difference being that intermediate models typically have an ocean with linear shallow water equations, while hybrid models have a fully

---

2 It forces the ocean in turn by the generation of various baroclinic (internal) waves that affect SST and hence the coupled paradigm.

3 These can vary from hours to weeks depending on the nature and scale of the disturbances.
nonlinear primitive equation ocean (an OGCM). Both types of model use highly simplified atmospheric components that use either shallow water equations or linear statistical models of the wind/SST relationship.

This conceptual time-scale division introduced above facilitates the explanation of the two leading theories for ENSO irregularity. The first theory posits that the irregularity derives primarily from a chaotic interaction of various slow modes and therefore assumes implicitly that the fast modes are of secondary importance. The second theory contends instead that the slow time-scale modes are not chaotic but are instead ‘disrupted’ in a stochastic fashion by the fast modes.

In this article, we shall concentrate heavily on detailing the second explanation, but it should be borne in mind that the first explanation has many adherents among climate scientists and so the area of ENSO irregularity is still not a settled question. We review in turn the basic ideas behind each of the two theories.

(a) Slow-mode chaos

This theory was advanced by a number of investigators in the early 1990s (e.g. Chang et al. 1994; Jin et al. 1994; Tziperman et al. 1994). The principal idea is that the slow manifold of the tropical ocean–atmosphere system has a number of important time scales and the ocean is sufficiently nonlinear that modes with such time scales are able to interact and generate chaotic behaviour. Most focus has been on the interaction between the annual cycle and the most unstable coupled mode, which in most realistic models has a time scale of approximately 3–5 years. Less attention has been focused on the interaction between coupled (unstable) modes of differing time scales (see, however, Tziperman et al. (1995), who analyses the model of Zeami & Cane (1987)).

The annual cycle has been found in many coupled models to show a significant interaction with the leading coupled mode. This often, but not always, manifests itself in the form of locking of the interannual variability to a fixed integral multiple of the annual cycle. A detailed and careful analysis of this effect may be found in Jin et al. (1996). The important physical parameters of this dynamical systems problem are the ratio of the time scales of the forced and unstable modes as well as the growth rate of the latter mode. The authors mentioned above studied large parameter regions of a simplified (but realistic) coupled model in which these two factors showed variation over all realistically possible values. Overall, they found that the locking of the variability to various multiples of annual period was the most common behaviour. When the model was tuned to be between such multiples of the annual cycle, then chaos appeared. The route taken in the parameter space in approaching these chaotic regimes was one often seen in the transition to turbulence in simple dynamical models. These chaotic regions of the parameter space were interpreted as overlapping resonances of the system to the annual cycle forcing. Such resonances were the most commonly seen model behaviour, i.e. oscillations with periods of some integral multiple of the annual period. The power spectra associated with this form of chaos tend to show the overall pattern mentioned in §1; however, it usually also often displayed peaks at periods that were a (low) rational multiple of the annual period. This preference for low rational multiples of the forcing period is
a common feature of these kinds of dynamical systems. The particular peaks emphasized in the chaotic spectrum varied according to the chosen model parameters.

(b) **Stochastic forcing of slow modes by fast modes**

Many intermediate or hybrid coupled models have been developed in the past 25 years and have been consistent with the parameter space analysis of the particular simple coupled model discussed in §2a, which usually exhibit a regular oscillation. Such an oscillation may be either self-sustained or damped depending on how the particular model is tuned. Motivated partially by this empirical observation, a different paradigm for irregularity to that discussed above was suggested in the late 1980s and 1990s (e.g. Battisti 1988; Kleeman & Power 1994; Penland & Matrosova 1994; Penland & Sardeshmukh 1995; Kleeman & Moore 1997; Moore & Kleeman 1999).

This new paradigm explicitly separated the fast and slow modes in the atmosphere by adding the former to simple models as a stochastic forcing term. The additional fast-mode random forcing was found to robustly disrupt the slow scales and convert the original regular oscillation to an irregular one. It was also observed that damped regular oscillations could be transformed by this mechanism to self-sustained irregular oscillations, i.e. made to much better resemble the observed ENSO behaviour. This latter effect occurs through the flow of energy from the fast to slow modes. It is interesting to observe how the nature of early ENSO modelling with slaved atmospheric components and unrealistic regular oscillations facilitated this theoretical development.

As we shall see in §3, an important feature of the fast-mode stochastic forcing is its required spatial coherency. The dominant ENSO modes are typically rather large scale and require for their excitation forcing with a similarly large scale. It was realized therefore that the efficient stochastic forcing of ENSO could arise only from large-scale synoptic atmospheric transients such as the Madden Julian Oscillation (MJO) and associated westerly wind bursts (WWB) or inter-tropical convergence zone easterly waves. Whether or not any particular coherent convective phenomena is responsible for ENSO irregularity still remains somewhat uncertain since the form of the most efficient stochastic forcing can vary from model to model. However, it is worth noting that for a range of particular coupled models, this issue has been directly addressed (see Zavala-Garay *et al.* in press) and the conclusion is that the coherent aspects associated with the MJO, at least, can indeed be very important.

When simple models are stochastically forced in the manner described above, it is usually rather easy to produce an irregular oscillation with a spectrum qualitatively matching that shown in figure 4. It is also possible to produce a phase locking of warm events to the annual cycle similar to observations. The physical mechanism behind this latter effect is easy to discern and appears physically plausible. Much more detail on all these points may be found in §§3 and 4.

---

4 This statement is based on extensive consultations by the author with a large range of ENSO modellers.
3. Analysis of the stochastic forcing of complex dynamical systems

As we noted previously, a central question in investigating the plausibility of stochastic forcing as a mechanism for ENSO irregularity is in determining exactly what kinds of random forcing are most efficient in disrupting the slow manifold of the dynamical system. After all, if the plausible candidates for random forcing are very inefficient in such disruption, then the theory becomes less attractive as an explanation of ENSO irregularity.

This question has been analysed comprehensively in the context of systems in which the growth of variance of important dynamical variables can be explained by linearized dynamics. This appears to be a reasonable assumption for ENSO where linearized analysis has historically proved very useful (e.g. Hirst 1986; Battisti & Hirst 1989). It also appears appropriate on empirical modelling grounds for the analysis of variance growth over short intervals (three to six months). Finally, a posteriori forcing of nonlinear models with the results of the linearized analysis confirms the validity of the approximation used (e.g. Moore & Kleeman 1999).

The stochastic forcing of high-dimensional linear dynamical systems has been exhaustively investigated in Farrell & Ioannou (1995), Kleeman & Moore (1997) and Gardiner (2004). We present the second approach here as it is adapted to suit the needs of the problem presently under consideration.

Consider a multi-dimensional stochastic differential equation. If we time discretize the solution, we obtain

\[ u_{\mu+1} = R(\mu + 1, \mu)u_\mu + \Delta t f_\mu, \tag{3.1} \]

where \( u \) is a vector in the sense that it may represent spatial variation and also many physical variables. Time indices are denoted by Greek subscripts. The operator \( R \) is the so-called propagator that shifts a state vector forward in time and, finally, \( f \) is a stochastic forcing term whose statistics are assumed to satisfy

\[
\begin{align*}
\langle f_{j_2} \rangle &= 0, \\
\langle f_{\mu} f_{j_\nu} \rangle &= C_{ij}^{\mu\nu},
\end{align*}
\tag{3.2}
\]

where we are using Latin subscripts to denote the vector indices. If we iterate equation (3.1) from some initial time \( \mu = 0 \), then we obtain

\[ u_\mu = R(\mu, 0)u_0 + \Delta t \sum_{\lambda=0}^{\mu-1} R(\mu, \lambda + 1) f_\lambda. \tag{3.3} \]

If we now further assume for simplicity that the noise is white in time (a reasonable assumption for our purposes since the atmospheric synoptic time scale is very short for climate problems), then we may write

\[ \langle f_{\mu} f_{j_\nu} \rangle = \frac{1}{\Delta t} \delta_{\mu\nu} C_{ij}. \]

From equation (3.3), we may now easily write down expressions for the first and second moments (mean and covariance) of \( u_\mu \) as

\[ \langle u_\mu \rangle = R(\mu, 0)u_0, \]

\[ \langle u_\mu, u_\mu \rangle = \Delta t \sum_{\lambda=0}^{\mu-1} R(\mu, \lambda + 1) CR^*(\mu, \lambda + 1), \]

\[ \text{Phil. Trans. R. Soc. A (2008)} \]
where the asterisk denotes the transpose or adjoint operator. Taking the continuous limit, we obtain

$$\langle u(t) \rangle = R(t, 0)u(0),$$

$$\langle u(t), u(t) \rangle = \int_0^t R(t, t') C R^*(t, t') \, dt'.$$

Let us now consider the variance with respect to some index (e.g. average eastern equatorial SST)

$$\text{var}(t) = X_{ij} \langle u^i(t), u^j(t) \rangle,$$

where we are assuming the summation convention for repeated Latin indices and $X_{ij}$ can be considered the ‘metric’ matrix for our index of interest. It is now easy to show that

$$\text{var}(t) = \text{trace}\{ZC\},$$

$$Z \equiv \int_0^t R^*(t, t') X R(t, t') \, dt'.$$

Note that we have completely separated the stochastic forcing represented by the covariance matrix $C$ from the dynamics represented by the operator $Z$. It is easy to show that both these operators are positive (and hence Hermitian) and therefore have positive eigenvalues. We can therefore write

$$\text{trace}\{ZC\} = \sum_{n,m} p_n q_m (P_n, Q_m)^2,$$

where the lower case $p$ and $q$ are the eigenvalues of, respectively, $Z$ and $C$, while the upper case $P$ and $Q$ are the corresponding eigenvectors. The inner product squared here can be interpreted as vector projection. Thus, if the eigenvectors of the noise forcing covariance matrix project onto the dynamical eigenvectors (we call these stochastic optimals), then there will be significant variance (or uncertainty) growth in our index of interest. Obviously, if the noise eigenvectors (often called EOFs) resemble the stochastic optimals with largest eigenvalues, then maximal variance growth will occur. We have therefore a very convenient framework for analysing the susceptibility of dynamical systems to disruption by noise.

### 4. Stochastically forced model results

The spectrum (and eigenvectors) of the operator $Z$ discussed in §3 have been evaluated for a range of linearized ENSO intermediate and hybrid models (see Kleeman & Moore 1997; Tang et al. 2005; Moore et al. 2006), and, in the models studied, it is highly peaked with most of the variance growth being caused by the first two eigenvectors (i.e. the $p_1$ and $p_2$ values are much greater than the others). These stochastic optimals are therefore crucial for whether large variance growth can occur. Figure 5 shows the spatial patterns of heat and momentum flux associated with these optimals for a particular intermediate coupled model.

Patterns of forcing such as this within the coupled model quickly grow into SST and wind stress disturbances resembling the so-called observed westerly wind burst, as shown in figure 6.
This signature of a disturbance often associated with the MJO suggests that this large-scale pattern of internal atmospheric variability may be favourably configured to disrupt the ENSO dynamical system. It also states that only noise with large-scale spatial coherency will be effective at disruption. These results are from one coupled model only but the general qualitative conclusions hold for
many models (for other results with different models, see Moore & Kleeman 2001). The detailed nature of the stochastic optimal can show some variation from model to model and also depends on the linearization background state used.

If the ENSO intermediate model above is forced by white noise with the spatial coherency of the stochastic optimals, then an irregular oscillation is induced. Figure 7 shows the result of such forcing on the above intermediate coupled model with such noise. Without the noise, a perfectly regular decaying oscillation is observed.

The filled circles show December of each year, showing that the observed seasonal synchronization of large warm events is also achieved. The reason for the synchronization is related to the seasonal and ENSO cycle variation of the instability of the system to small perturbations. In the model displayed, this tends to be greatest in the northern spring and in the lead up to warm events. At that time, convective anomalies in the atmosphere are able to develop with the greatest zonal extent since the mean SST is high at the mentioned phases of both the seasonal and ENSO cycles. It is also worth noting that this synchronization scenario has recently been confirmed in the careful observational study of Hendon et al. (2007). The irregular oscillation in figure 7 shows a clear resemblance to the observed pattern in figure 3 and the spectrum displayed in figure 8 is qualitatively the same as the observed spectrum in figure 4.

The irregular behaviour noted is particularly robust as one can vary the amplitude of the forcing by some orders of magnitude without qualitative effect. It has now been observed by many other investigators using a range of different models (examples include, but are not restricted to, Blanke et al. (1997), Eckert & Latif (1997) and Zhang et al. (2003)). In addition, long integrations of this stochastic model show that there is a decadal variation of its spectrum qualitatively similar to the observations seen in figure 2. Such agreement is not surprising as it may also be observed in very simple stochastically forced oscillators as demonstrated in figure 9 of Kestin et al. (1998) for the case of an AR(3) process.

Figure 7. A plot of the NINO3 SST index from a stochastically forced coupled model. The filled circles indicate December. Note the frequent locking of warm event peaks to this time of year. Adapted from Moore & Kleeman (1999).
The variance growth in time predicted by this theory of ENSO indices has the form of rather rapid growth for the first six to nine months with considerably slower growth subsequently. It is interesting that this pattern is also observed in physically complete coupled models (coupled general circulation models, CGCM; see Stockdale et al. 1998) that, of course, have inbuilt atmospheric transients as part of their atmospheric components.

A further detailed comparison of the simple model results discussed above with those from CGCMs is difficult owing to the complexity of the latter and the fact that their overall simulation of the observed seasonal cycle, interannual variability and moist convection is often still suboptimal. Nevertheless, such models are improving and a recent model of Lengaigne et al. (2004) shows reasonable simulations of all these features. This particular model shows strong sensitivity of predictions to disruption by wind anomalies with the structure of WWB and, in that respect, resembles the simple model behaviour discussed above.

This latter model is also able to capture with surprising accuracy the qualitative high-frequency evolution of strong ENSO events. The detailed evolution of zonal wind, SST and zonal currents is very well depicted when compared with the good observational record of the 1997 warm event. Ensemble experiments with very closely aligned initial conditions reveal the strong sensitivity of predictions of warm events. Additionally, they show that such ensembles may be shifted markedly towards strong warm events by the insertion of a short-lived westerly wind anomaly with a spatial structure resembling that detailed earlier in this section.

5. Predictability implications

The accurate prediction of the ENSO phenomenon is of great societal value. The ENSO warm and cold events are often responsible for huge economic dislocation due to the induced global large-scale precipitation and temperature anomalies (Ropelewski & Halpert 1987; Halpert & Ropelewski 1992). For this reason,
considerable energy has been devoted to accurately forecasting ENSO (Kirtman et al. 2002) and many operational models with historical skill are currently used to make routine forecasts.

Evidently, the nature of the irregularity of ENSO is central to the issue of potential limits to the predictability of this climatic variation. Different physical mechanisms for irregularity have probably very different implications for a fundamental predictability limitation. At present, it remains unclear the degree to which present forecast errors are due to such limiting factors or to the more prosaic model error. Certainly, clear limitations in practical prediction were seen in the recent large warm event of late 1997, as documented by Barnston et al. (1999).

Predictability implications of the two mechanism outlined in §2 have been addressed using idealized twin experiments to investigate the potential error growth due to uncertain initial conditions (e.g. Goswami & Shukla 1991; Kleeman & Moore 1997). In general, the error growth patterns in the two cases have very different characteristics with chaotic (and unstochastically forced) models exhibiting much longer time scales than those seen in stochastic models (years in the first case versus months in the second). In addition, an important sensitivity has been found in the predictability results from stochastic models by Thompson & Battisti (2001). The limitation time scale was found to be a strong function of the instability time scale of the coupled mode within the system. More unstable coupled modes lengthened the predictability because it made such modes (which are evidently central to any prediction) less susceptible to disruption by random forcing. The greater the coupled mode instability, the more pronounced and persistent the resulting oscillation. Such oscillations are by their nature less able to have their phase shifted by random perturbations (a nice illustration of this idea may be found in simple climate models studied in Kleeman 2002).

The results from §3 suggest that large-scale convective fluctuations such as the MJO (and associated westerly wind events) may be important factors in limiting predictability. By their very nature, the phase and amplitude of such episodes are not predictable beyond a month at most. Thus, if they are efficient in forcing the coupled system, then they are clear candidates for the disruption of predictions. There is some evidence from an operational prediction model (documented in Kleeman 1994) that this may have indeed occurred in the case of the 1997 warm event. Figure 9 displays real-time operational predictions of the NINO3 index made one month apart from January through April of that year. All predictions were suggesting warming later in the year; however, the amplitude of this shows a sudden jump between February and March with the later predictions being significantly more accurate. This behaviour suggests that something in the model initialization during March 1997 caused the predictions to respond rather dramatically and beneficially. The examination of the wind data for that month (the main source of coupled model initialization) shows that, indeed, there was a quite strong equatorial westerly wind event near the international date line in the first half of that month.

The case for the influence of the MJO on the ENSO development has also been made strongly from an observational perspective by Zhang & Gottschalck (2002) and Batstone & Hendon (2005). These authors show that a large fraction of equatorial oceanic Kelvin wave activity preceding ENSO events can be traced back to the MJO activity.
Finally, it is worth noting that several authors (Fedorov 2002; Flügel et al. 2004) have proposed, for differing reasons, that the stochastic influence on ENSO predictability may exhibit significant decadal variation.

6. Outlook

Perhaps the least defined aspect of the paradigm outlined here concerns the detailed nature of the stochastic forcing operating within the coupled system. In particular, while it is clear that the forcing most efficient at disrupting the system is large scale in character, its precise nature is clearly dependent on the detailed way in which the coupled dynamics are represented within a model. This dependency manifests itself in the differing stochastic optimals seen in different coupled models.

In a physically complete CGCM, such uncertainty derives from two important and only moderately well-understood physical parametrizations: atmospheric tropical convection and oceanic vertical mixing. Both these physical processes strongly influence the basic nature of the coupled system. It is to be expected that further deeper understanding of the stochastic forcing of ENSO will occur when understanding of these two physical processes increases.

Recent theoretical and modelling works (e.g. Perez et al. 2005; Jin et al. 2007) have suggested that the stochastic forcing operating within the coupled system should be dependent on the phase of ENSO. This amounts to an assumption of the multiplicative stochastic forcing as opposed to the simple additive forcing discussed in §4. Convective disturbances are confined generally to high SST regions. Such warm pool regions are a strong function of the ENSO phase and,
furthermore, there is evidence from atmospheric studies that the zonal extent of the warm pool influences the development of large-scale convective disturbances. These factors are all physical grounds for expecting multiplicative effects to be important. Clearly, this area deserves more investigation.

References


