ENSO theory

J. David Neelin, David S. Battisti, Anthony C. Hirst, Fei-Fei Jin, Yoshinobu Wakata, Toshio Yamagata, and Stephen E. Zebiak

Abstract. Beginning from the hypothesis by Bjerknes [1969] that ocean-atmosphere interaction was essential to the El Niño–Southern Oscillation (ENSO) phenomenon, the Tropical Ocean–Global Atmosphere (TOGA) decade has not only confirmed this but has supplied detailed theory for mechanisms setting the underlying period and possible mechanisms responsible for the irregularity of ENSO. Essentials of the theory of ocean dynamical adjustment are reviewed from an ENSO perspective. Approaches to simple atmospheric modeling greatly aided development of theory for ENSO atmospheric feedbacks but are critically reviewed for current stumbling blocks for applications beyond ENSO. ENSO theory has benefited from an unusually complete hierarchy of coupled models of various levels of complexity. Most of the progress during the ENSO decade came from models of intermediate complexity, which are sufficiently detailed to compare to observations and to use in prediction but are less complex than coupled general circulation models. ENSO theory in simple models lagged behind ENSO simulation in intermediate models but has provided a useful role in uniting seemingly diverse viewpoints. The process of boiling ENSO theory down to a single consensus model of all aspects of the phenomenon is still a rapidly progressing area, and theoretical limits to ENSO predictability are still in debate, but a thorough foundation for the discussion has been established in the TOGA decade.

1. Introduction

1.1. ENSO as Prototype

Gains in understanding and predicting the El Niño–Southern Oscillation (ENSO) phenomenon provided the goal and the major successes of the Tropical Ocean–Global Atmosphere (TOGA) program. ENSO, the largest interannual climate signal, is the first climate phenomenon shown to depend essentially upon coupled interactions of the dynamics of both ocean and atmosphere. As such, it has provided a prototype for laying the theoretical and modeling foundations of ocean-atmosphere interaction in the tropics more generally.

Studies of ocean-atmosphere interaction have advanced rapidly for the tropical regions because at large scales each medium is strongly controlled by the boundary conditions imposed by the other. The large-scale upper ocean circulation is largely determined by the past history of the wind stress, with internal variability occurring primarily at space scales and timescales well separated from the seasonal-to-interannual scales that have been the initial focus of study. Major features of the tropical atmospheric circulation, averaged over timescales longer than a month or two, are largely determined by sea surface temperature (SST). Although the tropical atmosphere does have significant internal variability, the decorrelation timescales of these are sufficiently short that a conceptual separation can often be made between atmospheric internal variability and that associated with slower ocean evolution communicated by SST. This contrasts to the midlatitude situation where internal variability of both atmosphere and
Plate 1. Chaotic versus frequency-locked behavior in an intermediate coupled model as a function of parameters. The coupling coefficient, $\mu$, affects the amplitude of the oscillation and thus the strength of nonlinearity. The surface-layer coefficient, $\delta$, affects the inherent ENSO period. Frequency ratio of the ENSO frequency to the annual frequency is shown by the color scale, with 0.25 corresponding to one El Niño every 4 years, etc. Chaotic regions are shown in black. The blank region at low coupling is below the primary bifurcation so the ENSO mode is stable. After Jin et al. [1996].
1.2. The Bjerknes Hypothesis

The reigning paradigm for ENSO dynamics, that it arises through ocean-atmosphere interaction in the tropical Pacific, dates back to a hypothesis of Bjerknes [1969]. The essence of Bjerknes’s postulate, as reinterpreted in light of our current knowledge, is that ENSO arises as a self-sustained cycle in which anomalies of SST in the Pacific cause the trade winds to strengthen or slacken and that this in turn drives the ocean circulation changes that produce anomalous SST. Bjerknes did not include mechanisms for how the system moves from a phase with warm SST anomalies, through phases with relatively little SST signature, to the subsequent cold phase, and the paradigm now includes the statement that the ocean, with its slower timescales of adjustment, provides the memory that carries the oscillation from phase to phase. Within this paradigm one may still distinguish a variety of mechanisms which potentially contribute to the maintenance, timescale, and spatial form of the cycle, and elaboration and winnowing of these was a centerpiece of TOGA.

As a background to understanding ENSO-related interannual variability, a brief description of the time-mean circulation and some essential features of ENSO is worthwhile. For a full description of the observations, see Wallace et al. [this issue]. Differential forcing of the atmosphere by the SST boundary condition drives thermodynamically direct circulation cells; convection tends to organize roughly over the warmest SST, producing the Intertropical Convergence Zones (ITCZs). The zonally symmetric component, i.e., the Hadley circulation, contributes an easterly (i.e., westward) component to tropical surface winds. This is strongly reinforced over the tropical Pacific by winds associated with the zonally asymmetric, i.e., the Walker circulation. Over the Pacific the Walker circulation is driven to a significant extent by the strong SST gradient across the basin, with warm waters in the west. The westward wind stress is balanced in the climatology largely by pressure gradients in the upper ocean associated with a sea level gradient of about 40 cm across the Pacific and a corresponding slope in the thermocline. Within the upper ocean the differential deposition of wind stress by vertical viscosity drives westward surface currents along the equator, and Ekman drift due to the Coriolis force to either side of the equator drives a narrow band of upwelling along the equator, especially under the regions of strong easterlies in the eastern/central Pacific. The combination of upwelling and shallow thermocline produces the “equatorial cold tongue” in the east, while the deep thermocline in the west is associated with warm SST, the western Pacific “warm pool”.

The important dependence of SST in the equatorial cold tongue region on wind-driven ocean dynamics (rather than just on air-sea heat exchange) and the Walker circulation response to anomalies in the SST pattern form the key elements of the Bjerknes hypothesis. Consider an initial positive SST anomaly in the eastern equatorial Pacific. This anomaly reduces the zonal SST gradient and hence the strength of the Walker circulation, resulting in weaker trade winds at the equator. This leads to a deeper thermocline and reduced currents and produces higher surface temperatures in the cold tongue region. This reduces the SST gradient still further in a positive feedback which modelers have argued can lead to instability of the climatological state via ocean-atmosphere interaction. The cyclic nature of the resulting mode depends on the timescales of response within the ocean, as will be discussed in section 4. Earlier or alternate views that did not subscribe to the theory of a cyclic behavior underlying ENSO variability will also be discussed. Supporting the modelers’ consensus of cyclic behavior is the existence of significant peaks in power spectra of observed ENSO time series. Although the time series, for instance, of eastern equatorial Pacific SST or atmospheric pressure differences across the basin are rather irregular, a low-frequency peak at around the 3-to-5 year period has been found to be significant in a number of studies [e.g., Rasmusson and Carpenter, 1982; Rasmusson et al., 1990; Jiang et al., 1995; Gu and Philander, 1995; Wang and Wang, 1996]. An additional, weaker peak around the 2-year period has been noted by several of the above studies. These are referred to as the quasi-quadrennial or low-frequency peak and the quasi-biennial peak, respectively. More discussion of these is given in section 6 in the context of the discussion of ENSO irregularity.

1.3. Model hierarchy

Hierarchical climate modeling employs a succession of models of various complexities, the most complex being the atmospheric, oceanic, and coupled general circulation models (AGCMs, OGCMs, and CGCMs). Tropical coupled modeling has benefited from arguably the most complete hierarchy of any climate field, since intermediate tiers of the hierarchy have been successfully filled. Coupling considerations tend to be similar to the procedures applied in GCMs, except that formulations in which anomalies are modeled about a specified climatology are common in the simpler models. Coupled GCMs are the subject of work by Delecluse et al. [this issue], so we here cover the theoretical results from other members of the hierarchy.

The foundations for the hierarchy of coupled models were laid through the study of the individual components. The dynamics of the equatorial ocean response to wind stress was examined in shallow water
models representing the dynamics of the upper ocean [e.g., Moore, 1968; Cane and Sarachik, 1977, 1981; McCreary, 1976], modified shallow water models [e.g., Cane, 1979a; Schopf and Cane, 1983], and ocean general circulation models [e.g., Philander and Pacanowski, 1980, Philander, 1981]. In the atmosphere, simple atmospheric models with steady, damped shallow water dynamics were shown to provide a reasonable approximation to the low-level tropical atmospheric response to SST anomalies [e.g., Gill, 1980; Webster, 1981; Zebiak, 1982; Gill and Rasmusson, 1983]. There is still disagreement as to the best formulation of these simple atmospheric models [Zebiak, 1986; Lindzen and Nigam, 1987; Neelin and Held, 1987; Neelin, 1989a; Allen and Davey, 1993; Wang and Li, 1993] but their simulation of anomalous wind stress from given SST has qualitative similarity to observations and to AGCM simulations [e.g., Lau, 1985; Palmer and Mansfield, 1986; Methoso et al., 1987; Shukla and Fennessey, 1988, and references therein].

The coupled model hierarchy constructed from these components often uses the following classifications: simple models, intermediate coupled models (ICMs), hybrid coupled models (HCMs) and coupled GCMs. The simple models include the simplest linear shallow water models [e.g., Lau, 1981; Philander et al., 1984; Gill, 1985; Hirst, 1986, 1988; Wakata and Sarachik, 1991a; Neelin, 1991; Wang and Weisberg, 1996], together with some useful models which condense the dynamics even further, usually at the cost of ad hoc approximations. The more complex and carefully parameterized of the modified shallow water models that include adequate nonlinearity to ensure that the system evolves in a bounded (and hopefully realistic) manner are deemed ICMs [e.g., Cane and Zebiak, 1985; Anderson and McCreary, 1985; Zebiak and Cane, 1987; Battisti, 1988; Battisti and Hirst, 1989; Schopf and Suarez, 1988; Yamagata and Masumoto, 1989; Graham and White, 1990; Xie et al., 1989; Jin and Neelin, 1993a, b; Yang and O'Brien, 1993; Chang et al., 1995]. The next most complex models are the HCMs, consisting of an ocean GCM coupled to a simpler atmospheric model [e.g., Neelin, 1989a, 1990; Latif and Villwock, 1990; Barnett et al., 1993; Waliser et al., 1994; Suy et al., 1995]. This design choice is made because the ocean contains both the memory and limiting nonlinearity of the system; the atmosphere is thus treated as the fast component of a stiff system. Divisions in the hierarchy are not sharp and some of the lowest-resolution coupled GCMs may not be that much more complex than the best ICMs. Many of these models produce interannual variability through coupled interactions which have significant parallels to ENSO dynamics.

2. Intermediate Coupled Models

Intermediate coupled models currently constitute the most important contributors to ENSO theory. Simulations and predictions with them can be quantitatively compared with observations, their parameter space can be inexpensively explored, and they can be used to produce associated versions that are linearized or reduced by asymptotic approximations to give insight into the model behavior. Most of the simple models can be described by comparison to the intermediate models, and reference to the ocean or atmospheric components is useful when describing the behavior of these subsystems. We thus present here a typical intermediate model for reference in the subsequent sections.

The intermediate coupled model of Cane and Zebiak [1985] and Zebiak and Cane [1987] (hereinafter, work by Cane and Zebiak [1985] and Zebiak and Cane [1987] is referred to as CZ), has proven influential in ENSO studies and has provided the first successful ENSO forecasts with a coupled model. Battisti [1988] and Battisti and Hirst [1989; hereinafter referred to as BH] use an independent version of the same model, and Jin and Neelin [1993a, b], Neelin and Jin [1993] (hereinafter Jin and Neelin [1993a, b], Neelin and Jin [1993] are referred to as JN), and Jin et al. [1994, 1996] use a related model. The details of the CZ model include some drawbacks that seem not to adversely influence the overall results and are immaterial to the presentation here but are noted below for completeness. Other intermediate models [Anderson and McCreary, 1985; Schopf and Suarez, 1988; Yamagata and Masumoto, 1989; Graham and White, 1990; Xie et al., 1989; Chang et al., 1995] share many of the same basic properties and are described by comparison at the end of the section.

The presentation is nondimensionalized following JN to bring out a few primary parameters from among the many lurking in the coupled system. These are: $\mu$, the relative coupling coefficient, which is the strength of the wind stress feedback from the atmosphere per unit SST anomaly, scaled to be order unity for the strongest realistic coupling; for $\mu = 0$ the model is uncoupled; $\delta$, the relative adjustment time coefficient, which measures the ratio of the timescale of oceanic adjustment by wave dynamics to the timescale of adjustment of SST by coupled feedback and damping processes; scaled to be order unity at standard values of dimensional coefficients; and $\delta_s$, the surface-layer coefficient, which governs strength of feedbacks due to vertical-shear currents and upwelling, $(u_s, v_s, w_s)$, created by viscous transfer between the surface-layer and the rest of the thermocline; as $\delta_s \to 0$, the effects of these become negligible.

A modified shallow water model with an embedded, fixed-depth mixed layer [Cane, 1979a; Schopf and Cane, 1983] provides the ocean-dynamics component

\[
(\delta \partial_t + \epsilon_m)u'_m - yv'_m + \partial_x h' = \tau'
\]  
(1a)

\[
yu'_m + \partial_x h' = 0
\]  
(1b)

\[
(\delta \partial_t + \epsilon_m)h' + \partial_x u'_m + \partial_y v'_m = 0
\]  
(1c)

\[
\epsilon_s u'_s - yv'_s = \delta_s \tau'
\]  
(2)

compared with observations, their parameter space can be inexpensively explored, and they can be used to produce associated versions that are linearized or reduced by asymptotic approximations to give insight into the model behavior. Most of the simple models can be described by comparison to the intermediate models, and reference to the ocean or atmospheric components is useful when describing the behavior of these subsystems. We thus present here a typical intermediate model for reference in the subsequent sections.

The intermediate coupled model of Cane and Zebiak [1985] and Zebiak and Cane [1987] (hereinafter, work by Cane and Zebiak [1985] and Zebiak and Cane [1987] is referred to as CZ), has proven influential in ENSO studies and has provided the first successful ENSO forecasts with a coupled model. Battisti [1988] and Battisti and Hirst [1989; hereinafter referred to as BH] use an independent version of the same model, and Jin and Neelin [1993a, b], Neelin and Jin [1993] (hereinafter Jin and Neelin [1993a, b], Neelin and Jin [1993] are referred to as JN), and Jin et al. [1994, 1996] use a related model. The details of the CZ model include some drawbacks that seem not to adversely influence the overall results and are immaterial to the presentation here but are noted below for completeness. Other intermediate models [Anderson and McCreary, 1985; Schopf and Suarez, 1988; Yamagata and Masumoto, 1989; Graham and White, 1990; Xie et al., 1989; Chang et al., 1995] share many of the same basic properties and are described by comparison at the end of the section.

The presentation is nondimensionalized following JN to bring out a few primary parameters from among the many lurking in the coupled system. These are: $\mu$, the relative coupling coefficient, which is the strength of the wind stress feedback from the atmosphere per unit SST anomaly, scaled to be order unity for the strongest realistic coupling; for $\mu = 0$ the model is uncoupled; $\delta$, the relative adjustment time coefficient, which measures the ratio of the timescale of oceanic adjustment by wave dynamics to the timescale of adjustment of SST by coupled feedback and damping processes; scaled to be order unity at standard values of dimensional coefficients; and $\delta_s$, the surface-layer coefficient, which governs strength of feedbacks due to vertical-shear currents and upwelling, $(u_s, v_s, w_s)$, created by viscous transfer between the surface-layer and the rest of the thermocline; as $\delta_s \to 0$, the effects of these become negligible.

A modified shallow water model with an embedded, fixed-depth mixed layer [Cane, 1979a; Schopf and Cane, 1983] provides the ocean-dynamics component

\[
(\delta \partial_t + \epsilon_m)u'_m - yv'_m + \partial_x h' = \tau'
\]  
(1a)

\[
yu'_m + \partial_x h' = 0
\]  
(1b)

\[
(\delta \partial_t + \epsilon_m)h' + \partial_x u'_m + \partial_y v'_m = 0
\]  
(1c)

\[
\epsilon_s u'_s - yv'_s = \delta_s \tau'
\]  
(2)
where latitude, $y$, appears due to the nondimensionalized
Coriolis force, and the equations are applied here to
departures, denoted $(\cdot)',$ from a specified climatology,
denoted $(\cdot)$. Anomalous vertical mean currents above
the thermocline, $(u'_m, v'_m)$, and thermocline depth, $h'$,
are governed by the shallow water component in the
long wave approximation, (1) (hereinafter (1) is used
to denote the system (1a)–(1e), and similarly for other
equations), with suitable boundary conditions at basin
boundaries [Gill and Clarke, 1974]. Vertical-shear currents,
$(u'_s, v'_s)$, are governed by local viscous equations,
(2). Both are driven by the zonal wind stress anomaly,
$\tau'$. The damping rates, $\epsilon_m$ and $\epsilon_s$, are not treated as
primary parameters because the former is small and the
latter can be largely absorbed into $\delta_s$. For a more formal
scaling, see JD; for the justification of several approxima-
tions, see Cane [1979a] and CZ. Vertical velocities
are given by the divergence of the horizontal velocities
and the values of surface currents and upwelling into
the surface-layer by the sum of anomalous mean and
shear contributions; $u = u'_m + u'_s, \quad w'_m + w'_s$.

Because SST serves as a key interfacial variable, care-
ful parameterization of processes which affect SST are
largely responsible for the success of the CZ model.
The direct effects of temperature variations in the surface-
layer on pressure gradients are neglected in (1), but a
prognostic equation for SST is carried separately, which
contains all the essential nonlinearity of the CZ model;

$$
\partial_t T + u \partial_x T + H(u)w(T - T_{\text{sub}})/H_{1.5} + v \partial_y T + \epsilon_T(T - T_0) = 0
$$

where $T$ is total SST. The heaviside function $H(w)$ is
positive when $w$ is upward, thereby cooling the mixed
layer, and zero when downward motion causes no change
in SST. The Newtonian cooling represents all physical
processes that bring SST toward a radiative-convective-
mixing equilibrium value, $T_0$. The subsurface tempera-
ture field, $T_{\text{sub}}$, characterizes values upwelled from a
depth $H(1.5)$ in the underlying shallow water layer and
is parameterized nonlinearly on the thermocline depth;
deeper thermocline results in warmer $T_{\text{sub}}$. The form of
(3) is almost identical when nondimensionalized; several
timescales arise, of which the simplest (and shortest) is
a typical value of $(w/H_{1.5})^{-1}$.

The simple atmospheric models which provide a zero-
order approximation to the wind stress response to SST
anomalies can be written

$$
\tau' = \mu A(T'; x, y)
$$

where $\mu$ is the coupling coefficient and $A(T'; x, y)$ is a
nonlocal function of $T'$ over the entire basin. A linear
atmosphere gives essential behavior, though some mod-
els include nonlinear terms, which are, unfortunately,
not always well justified since they are usually associated
with crude parameterizations of convection. For a
Gill [1980] model and with a specified meridional pro-
file of the forcing appropriate to the assumed SST $y$
dependence $A$ is a simple integral operator.

Coupling is carried out by flux correction, a method
of constructing a known climatological state that
approximates the observed in order to model anomalies
about it. The ocean model is run with observed clima-
tological wind stress to define the ocean climatological
state $(u, w, T, \ldots)$; SST anomalies, $T'$, with respect to
this are

$$
T' = T - \bar{T}
\tau = \bar{\tau} + \tau'
$$

with $\tau'$ from (4) and $T$ from (3).

For sufficiently small values of the coupling coeffi-
cient, $\mu$, the climatological state is unique and stable;
tenannual variability must arise by bifurcations from
this state as $\mu$ increases.

Anderson and McCreary [1985] used an ICM in which
the ocean component carries an equation for a vertically
constant temperature above a jump at the thermocline,
as well as equations for layer-mean currents and the
thermocline depth. Within the stated assumptions this
model formulation is hydrodynamically more consistent
than the CZ ocean approach, but the latter is char-
acterized by aptly chosen approximations. Xie et al.
[1989] showed that the Anderson and McCreary model
depends strongly on the entrainment parameterization
and that a judicious choice of this leads to ENSO sim-
ulations qualitatively similar to CZ. Wang et al. [1995]
introduced a version of the CZ approach with a variable-
depth, embedded mixed layer, and Chen et al. [1995]
used a model with two baroclinic modes. Both use a
parameterized entrainment temperature similar to $T_{\text{sub}}$
in (3). Like CZ, this assists simulation of SST but at
the cost of losing a consistent available potential en-
ergy equation. Schopf and Suarez [1988] use a two-layer
primitive equation model with temperature equations
in both variable-depth layers. For a more extensive re-
view of TOGA ocean models, including ocean GCMs,
see Stockdale et al. [this issue].

3. Dynamics of the Uncoupled Tropical Ocean

Wind stress is the primary feedback from the atmos-
phere to the ocean for ENSO variability, and the ocean
response to that wind stress is approximately determin-
istic. Since the ocean supplies the memory of the sys-
tem, the nonequilibrium response of the ocean is im-
portant. Theory for the adjustment of an uncoupled
shallow water ocean to time-varying winds provided a
prelude to TOGA (see work by Moore and Philander
[1977]; Cane and Sarachik [1983]; McCreary [1985] for
reviews) and has been extended during the TOGA pe-
riod. We review this theory with emphasis on relevance
for low-frequency motions such as occur in ENSO theo-
retical models. Part of the success of TOGA is arguably
due to the ocean response at interannual timescales being reasonably well captured by linear or weakly nonlinear approximations to the ocean dynamics [Kessler, 1991; Arnauld and Perigaud, 1992; Delcroix et al., 1994].

In the ICM above, the ocean dynamics is entirely linear and the limiting nonlinearity of the system resides entirely in the SST equation, so to the extent that its successes in simulating and predicting ENSO are valid, linear theory can be used to understand how the ocean provides memory to the system. This also implies that a separation can be made between short timescale motions and lower frequency motions. In a finite basin these have rather different properties. Recent TOGA theory suggests that it is useful to think about these differing timescales in slightly different ways.

3.1. Free Ocean Solutions

3.1.1. Kelvin and Rossby waves. Dearly beloved by equatorial oceanographers are the eigensolutions of an unforced stratified fluid linearized about a resting basic state, horizontally homogeneous, except for linear $y$ dependence of the Coriolis parameter, and with no horizontal boundaries. These eigensolutions [Matsuno, 1966; Gill, 1980] involve linear combinations of Parabolic cylinder functions in latitude scaled by the equatorial radius of deformation. Commonly denoted $\psi_n$, these form a complete basis for decomposing latitudinal structure of fields obeying reasonable conditions [Cane and Sarachik, 1977]. The Fourier-decomposed longitude and time dependence of these eigensolutions obey dispersion relations familiar to every physical oceanographer [see Gill, 1985; Philander, 1990], of which the most important for our applications are equatorial Kelvin and long Rossby waves. The long wave approximation, used in (1)-(3), filters all wave types except these and approximates them as nondispersive for zonal wavelengths long compared to the equatorial radius of deformation (about 300 km for the first baroclinic vertical mode). The Kelvin wave is trapped within a radius of deformation of the equator, with a Gaussian latitudinal structure, and propagates eastward (phase speed about 2.5 m/s for the first vertical mode), while the equatorial Rossby waves have successively more complex structure and propagate westward. Sudden wind stress events in one part of the basin excite free wave propagation to other parts of the basin, and from the early days of TOGA it was realized that this could be important to interannual variability [e.g., Busalacchi and O'Brien, 1981]. For ocean GCM examples, see Giese and Harrison [1991], and references therein. Wave fronts with Kelvin wave phase speeds and structures exist in observations of upper ocean motions [e.g., Miller et al., 1988; Busalacchi et al., 1994; Kindle and Phoebus, 1995; Kessler et al., 1995] and projection on Rossby wave structures gives reasonable fit to many events [e.g., du Penhoat et al., 1992; Delcroix et al., 1991]. These eigensolutions of the zonally unbounded problem can be matched at coasts to give reflection properties of a Kelvin wave into Rossby waves at an eastern boundary [Clarke, 1983, 1992], and of Rossby waves into a Kelvin wave at a western boundary [Clarke, 1991; du Penhoat and Cane, 1991, and references therein]. Examination of such reflections in observations has been carried out, for instance, by Clarke [1992]; Clarke and Van Gorder [1994]; Kessler [1991]; and Kessler and McPhaden [1995]. Kessler and McCreary [1993] and Kessler et al. [1995] note that eastern boundary reflections appear to produce less Rossby wave variance above the thermocline than would be expected from the simplest theory, possibly because the energy can be transmitted through the thermocline into the deep ocean [e.g., Rothstein et al., 1985]. Du Penhoat and Cane [1991] and Clarke [1991] argue that western boundary reflections are only modestly affected by gaps such as the Indonesian throughflow region, while Verschell et al. [1995] argue that realistic geometry can substantially affect the reflected Kelvin wave.

3.1.2. Ocean basin modes and scattering spectrum. For a zonally finite ocean basin bounded by eastern and western coasts the Kelvin and Rossby wave solutions of the unbounded problem are no longer the eigensolutions. Because such basin boundaries are important in ENSO theory and much of our theoretical understanding is based on eigenmodes of the coupled problem in a finite ocean basin, it is relevant to know properties of the eigenspectrum of an uncoupled ocean in a suitably defined finite tropical ocean basin. The most suitable simple configuration to consider is the case of east and west coasts unbounded in the poleward direction so no wave energy completes the circuit around northern or southern boundaries. The analogy to the southern Pacific is obvious, and this assumes that dissipation and complex boundaries prevent wall-trapped solutions from making it across the northern Pacific boundary. Because wave energy in the form of wall-trapped Kelvin waves leaks from the equatorial region poleward up the eastern coast, no discrete spectrum exists for this case [Moore, 1968] for the full shallow water equations. Rather, there is a continuum spectrum or "scattering spectrum"; wave energy that is input at any frequency creates complex patterns that can be thought of as multiple reflections of Rossby and Kelvin waves, but ultimately, this energy is lost up the leaky eastern boundary. Near-resonance occurs at certain frequencies, however, and in the longwave approximation these portions of the continuum become discrete modes; the ocean-basin modes of Cane and Moore [1981] in representations in terms of a finite number of Rossby waves the remainder of the scattering spectrum is also discretized as decaying eigenvalues [Neelin and Jin, 1993]. The lowest-frequency ocean basin mode has quite a short period, about 9 months for typical wave speeds, and may be too leaky to be significant in the real ocean [Kessler, 1991]. The important property for the coupled problem is that there is no well-distinguished
ocean mode at interannual frequencies for ENSO to be closely associated with [Neelin and Jin, 1993]. Any explanation for ENSO must thus come from fully coupled theory rather than as a weakly coupled perturbation to some ocean eigensolution.

3.2. Wind-Driven Ocean Response

3.2.1. Ocean spin-up. Extensive theory exists for the adjustment of the uncoupled shallow water ocean to time-varying winds (see work by Moore and Philander [1977], Cane and Sarachik [1983], and McCreary [1985] for reviews). Much of it is phrased in terms of adjustment to abrupt changes of a wind patch, which creates Kelvin and Rossby wave fronts that can easily be tracked over an interval small enough that multiple reflections from coasts do not complicate the solution. An example of such a spin up process is provided in Figure 1 [McCreary and Anderson, 1984], for the case of a wind patch in midbasin switched on suddenly over an ocean initially at rest. The wind patch tapers from a maximum at the equator to zero at 1500 km off the equator and extends longitudinally over the central half of the basin, tapering to zero from a midbasin maximum. At early times, Kelvin and Rossby wave packets carry information about the sudden switch-on eastward and westward, respectively (Figure 1a). After the packets reach the boundaries of the basin and reflections return into the interior, the system begins to approach its long-term steady response along the equator, although off-equatorial regions adjust more slowly because of the slower phase speed of Rossby waves at higher latitudes.

3.2.2. Steady ocean response to wind forcing. Along the equator the solution for the thermocline depth is dominated by the balance of pressure gradient with wind stress (neglecting the damping term in (1a))

$$\begin{align*}
\partial_th &= \tau
\end{align*}$$

(6)

To set the thermocline depth along the equator a boundary condition on (6) is needed. This is obtained by solving the problem with time dependent equatorial waves and damping present and taking the steady limit carefully. The full solution along the equator corresponding to (6) is then [Hao et al., 1993]

$$h = \int_0^1 \tau(x_0)x_0^\frac{1}{2}dx_0 - \int_x^1 \tau(x_0)dx_0$$

(6')

which includes the effects of equatorial wave adjustment on the final state.
Figure 2 shows the steady response of \( h \) to a steady wind patch, centered on the dateline and constant in latitude. A damping time of 250 days has been used. From (6) and Figure 2a, it may be noted that the mean response along the equator is not zero. Mass has been transferred from higher latitudes to the equator in response to the westerly wind stress forcing [Cane and Sarachik, 1981]. The region of deepened thermocline on the east side of the basin is greater than the shallowed region to the west of the wind stress. The off-equatorial minima in heat content in the west might be misidentified as free Rossby signals but are in steady balance with the stress. Likewise, the deepened region in the east does not have Kelvin wave structure.

### 3.2.3. Ocean response to periodic wind forcing.

Understanding of interannual coupled oscillations has benefitted greatly from examination of oscillatory interannual eigenmodes of coupled models. In light of this a much better prototype for the coupled system is to force an uncoupled ocean by low-frequency, time-periodic winds. Such solutions were developed by Cane and Sarachik [1981]; hereinafter referred to as CS) using near-analytic results in a shallow water ocean, and by Philander and Pacanowski [1981] in an ocean GCM.

Figure 3 shows a time-longitude plot of thermocline perturbations along the equator for such a case, where the wind has a 3-year oscillatory period. The western Pacific leads the eastern Pacific a little less than half this period. For frequencies lower than the gravest ocean basin mode (about 8–10 months) the west leads the east by between 90° and 180° in temporal phase.

For a case in which ocean adjustment proceeds quickly compared to the rate of change of the wind, so the ocean approaches equilibrium with the wind, there would be about a 180° phase difference between west and east. The lead of the western region gives the appearance of a slow eastward propagation. It should be emphasized that this is not a wave propagation in the sense of any individual free wave of the system but rather the sum total of the ocean response, which is not quite in equilibrium with the wind forcing. The slight departures from equilibrium, as measured by the difference from 180° lag, characterize the oceanic memory, which is so important to interannual variability.

Figure 3 is produced from an analytical solution for the thermocline response to periodic wind stress, on the basis of an extension of the CS results [Neelin and Jin, 1993]

\[
- \int_x \tau(x_0) \cos 2\phi(x-x_0) dx_0 \exp \left[ i \left( \frac{y^2}{2} \tan 2\phi(x-x_0) \right) \right] \tag{7a}
\]

with the value of \( h \) at the eastern boundary given by

\[
h_E = \int_0^1 \left( \frac{\sin 2\phi x_0}{\sin 2\phi} \right) \tau(x_0) dx_0 \tag{7b}
\]

where \( \phi = \omega - i\epsilon \), with \( \omega \) as the frequency. The evolution of this response off the equator may be seen in the latitude-longitude maps in Figure 4, for four phases of the forcing through half a cycle. The patterns look qualitatively like ENSO thermocline depth patterns [see Wallace et al., this issue; Latif et al., 1993b] and are very similar to the patterns obtained for periodic ENSO variability found in ICMs [Battisti, 1988; JN]. From the expression in (7) it may be seen that Kelvin wave structure, \( \exp[-y^2/2] \), does not appear anywhere in the solution; the latitudinal width of the pattern in the eastern basin with an equatorial maximum varies with longitude and depends on frequency and damping. This is because the solution has time to be influenced by all basin boundaries. Decomposing the relatively simple solution (7) into Kelvin and Rossby wave \( y \) structures and then heavily truncating the infinite sum has been the traditional approach.

At phases when wind is present (Figures 4a, 4b, and 4d), the patterns tend to resemble the steady response pattern in Figure 2, multiplied by the amplitude of the wind at that phase. Thus the structures off the equator to the west of the wind are partly free Rossby waves that are still in the process of adjusting to the wind but partly are just forced response in balance with the wind stress. If we consider this pattern as the ocean half of a periodic coupled mode, only that part of the response that is not in balance with the wind, the "ocean memory," can affect the later evolution of the mode. One way to define this ocean memory contribution that is
appropriate for low-frequency behavior, where the solution has time for wave dynamics to communicate the effects of the boundaries into the interior, is to subtract the solution that would be in steady balance with the wind stress at each time. This allows departures from balance to be more clearly seen. Figure 5 shows maps of this ocean memory contribution for the same phases as Figure 4. The region west of the wind stress forcing, off the equator is the region with the most significant memory [BH; Battisti, 1989; Graham and White, 1991]. It acts as a reservoir of nonadjusted heat content that is fed down the western boundary to an equatorial boundary layer. At these timescales the “Kelvin wave” solution is essentially acting like an invisible tube, feeding mass across to the eastern basin, where it spreads off the equator due to the influence of the eastern boundary. These ocean patterns will be discussed again in a coupled context in section 5.

4. Atmospheric Response to SST

ENSO theory has benefitted greatly from the availability of simple atmospheric models that, despite gross simplifications in representing tropical moist dynam-
ics, can be tuned to give an adequate representation of the low-level wind field. The success of these models is partly because, to a first approximation, what is required of them in ENSO theory is modest, a rough representation of the anomalous wind stress response to an SST anomaly. Because the ocean response tends to integrate in space over the wind stress (see (6) or (7)), it turns out that the coupled response can, in some circumstances, be fairly forgiving of atmospheric errors in position or extent of the wind stress anomaly [see, e.g., Cane et al., 1990]. The main requirement, in hindsight, is that the response to a positive equatorial SST anomaly be a region of equatorial westerlies shifted toward the west of a positive SST anomaly. Nonetheless, it was a piece of good fortune for the TOGA program that the complexity of tropical atmospheric dynamics could be even roughly captured in simple models. For instance, the current lack of such models for the midlatitude atmosphere is a primary hurdle to advancements in midlatitude coupled theory. Detailed review of atmospheric modeling and observations in TOGA may be found in work by Trenberth et al. [this issue]; here we review only aspects of simple models most relevant for ENSO theory.

These simple atmospheric models mostly use variations on damped shallow water dynamics, typically of the form

\[ \varepsilon_a u_a - f v_a + \partial_z p = 0 \]  
\[ \varepsilon_a v_a + f u_a + \partial_y p = 0 \]  
\[ \varepsilon_a p + c_s^2 (\partial_z u_a + \partial_y v_a) = -Q \]  

where \( \varepsilon_a \) is an inverse damping time, \( u_a \) and \( v_a \) are atmospheric low-level velocities, and \( p \) is a pressure perturbation scaled by mean density, which is sometimes taken proportional to a layer mean temperature perturbation \( T_a \). The forcing, \( Q \), and the effective phase speed, \( c_s \), have different interpretations in different approaches.

Matsuno [1966] introduced these equations; they have become known as “Gill models” since Gill [1980] used one to examine atmospheric response to prescribed diabatic heating. Webster [1981] and Zebiak [1982] took the first steps toward relating the atmospheric forcing to SST. These models have been used in two types of problem: (1) to obtain low-level winds or surface stress from the forcing \( Q \) taken to be a specified convective heating; (2) to obtain winds and convective heating from the forcing \( Q \) specified from SST. The first problem is much easier, both in justifying the model and in obtaining accurate simulation. An influential example of this heating-forced problem, from the beginning of the TOGA decade, is given in Figure 6. The main region of low-level westerlies along the equator near the date line is qualitatively represented, although some of the other features are less satisfactory. The model uses a strong linear Raleigh friction, with a damping timescale of about 2 days. This may be justified by vertically integrating over the atmospheric lower layer and taking surface stress to be in the direction of the average wind in the layer [Lindzen and Nigam, 1987; Neelin, 1988]. This gives a timescale \( H/(C_D V) \), where \( H \) is lower layer depth, \( C_D \) is the drag coefficient, and \( V \) is the wind speed. The neglect of turning of the wind with height causes the surface stress simulation in these models to be overly zonal [Neelin, 1988; Wang and Li, 1993]. Specifying the heating roughly determines boundary layer divergence, and so much of the model amounts to obtaining the rotational part of the wind from the heavily damped, linear vorticity equation. Clarke [1994] discusses the balances that lead to anomalous westerlies occurring under anomalous convection.

For the harder problem, that of going from SST boundary conditions, the crudest solution is just to notice empirically that convection tends to occur preferentially over warm SST. However, this relation is far from perfect, and there have been several attempts to justify shallow water type dynamics by a variety of physical mechanisms. One approach has been to parameterize convective heating as proportional to convergence, with an arbitrarily specified “convergence feedback” parameter [Webster, 1981; Zebiak, 1986]. Another has been to specify moisture in convective regions (typically as a given fraction of saturation) and use a lower level moisture budget to give convective heating [e.g., Neelin and Held, 1987; Davey and Gill, 1988; Hendon, 1988; Kleeman, 1991; Wang and Li, 1993; Hendon and Salby, 1994]. In a two-level model this may be written schematically as

Figure 6. Simulation of low-level winds from by a simple, damped, shallow water atmospheric model in response to prescribed heating for September 1982 during the early part of the 1982-1983 El Niño. (a) Simulated low-level winds (arrows) forced by heat source anomalies (contours) specified as an approximation to observed convective anomalies. (b) Observed 850 mbar wind anomalies. Contours indicate wind speed (interval of 5 m/s). After Gill and Rasmusson [1983], reprinted by permission from Nature 306: 229-234, copyright (1983) Macmillan Magazines Ltd.
\[
\partial_t T_a - \hat{S} \nabla \cdot \mathbf{v} = Q_c + Q_R + Q_{SH} + q_b \nabla \cdot \mathbf{v} = -Q_c + E \tag{9}
\]

where \(T_a\) is column average or midlevel temperature and \(q_b\) is the low-level moisture, both in energy units; \(\nabla \cdot \mathbf{v}\) is low-level convergence; \(\hat{S}\) is the mean dry static stability; \(Q_c\), \(Q_R\), and \(Q_{SH}\) are the convective, longwave radiative, and sensible heat; and \(E\) is the evaporation, all normalized by column mass. Adding the two equations and linearizing gives the moist static energy equation

\[
\partial_t T_a - \nabla \cdot \mathbf{v} (\hat{S} - q_b) = F_{\text{net}} \approx E + \epsilon_1 T_a - \epsilon_2 T_a \tag{10}
\]

where \(F_{\text{net}}\) is the net energy flux into the column per unit mass. When linearized, the sensible heating and radiative fluxes tend to give terms proportional to surface temperature, \(T_s\), and column temperature with inverse damping times, \(\epsilon_1\), \(\epsilon_2\). Additional dependence on \(T_s\) arises in \(E\). The latent heating, \(Q_c\), does not appear in (10) since moisture sink and latent heat source necessarily cancel. This is a property of the vertically integrated moist static energy equation for the primitive equations, and consistent simpler models should preserve it. The quantity \((\hat{S} - q_b)\) acts as an effective static stability for the moist motions, termed the “gross moist stability.” In convective regions of a two-layer model it determines the phase speed of organized convecting motions, such as the Madden-Julian wave [e.g., Neelin and Held, 1987; Hendon, 1988; Wang and Li, 1993]. It must be positive for steady solutions, such as are used in many ENSO models, to be meaningful. An analogous quantity can be derived from the vertically continuous primitive equations using moist convective adjustment [Neelin and Yu, 1994; Yu and Neelin, 1997] that indeed is positive. Simple models that use a convergence feedback parameterization for condensation [e.g., Zebiak, 1986] can be viewed as approximating this equation (although without a moisture equation care must be taken to avoid inconsistencies in the atmospheric energetics). Taking \(p \propto -T_a\), the steady state version of (10) replaces (8c) to give a model that is, to a first approximation, similar to a Gill model (8) with forcing \(Q\) proportional to SST [Neelin, 1989a]. As a result, it becomes difficult to distinguish the Lindzen-Nigam model from other models based on (8) or on (8a) and (8b) with (10). Fitting of models of this form, under one interpretation or another, to observations has been carried out by Zebiak [1990] and Allen and Davey [1993]. Comparison to results of a model with a similar approach to (8) but with an additional surface-layer is found in work by Wang and Li [1993, 1994]. For discussion of an alternate approach based on approximations to the Betts and Miller [1986] deep convective scheme, see Yu and Neelin [1997] and Neelin [1997].

For TOGA studies this disagreement on mechanisms driving tropical flow was not a major impediment since mainly the wind stress feedback was required. For modeling in which heat fluxes become more important, notably over land surfaces, or in which details of wind stress matter, an improved understanding of tropical atmospheric flow appears essential.

5. Basic ENSO Mechanism

In describing our understanding of the basic mechanism that lies at the heart of the ENSO phenomenon we follow a loosely chronological outline. The approaches and models used fall naturally into an exploratory “early TOGA” phase, a “mid-TOGA” phase, characterized by articulation of dominant ideas, and an “end of TOGA” phase, which brought together some of the apparent contradictions of the mid-TOGA phase and opened new questions that will carry over into the post-TOGA era. The discussion deviates from strict chronology where ideas raised in one phase were more clearly understood in a subsequent phase.

5.1. Early TOGA

Early theoretical work on the coupled system included an exploration of some potential oscillation mechanisms in simple low-order systems [McWilliams and Gent, 1978] and coupled nonrotating shallow water models with coupling proportional to thermocline depth [Lau, 1981]. McCreary [1983] and McCreary and Anderson [1984] explored shallow water ocean dynamics coupled to wind stress patterns that changed by a discontinuous switch depending on thermocline depth. These
papers have often been omitted from recent citation because of the unrealistic switch condition in the atmosphere, but their discussion of basin adjustment processes influenced subsequent work. The conjecture of multiple equilibria for ENSO was introduced (warm and cold stationary states, with some additional process such as noise causing transitions between them) and it was not until the very end of TOGA that this possibility was eliminated.

A notable observation during pre-TOGA and early TOGA was the increase in western Pacific sea level prior to ENSO warm phases [Wyrtki, 1975, 1985]. Assimilation of this important observational fact into ENSO theory was complicated by the interpretation associated with it at the time, a postulated “energy relaxation,” with emphasis on deterministic ocean, purely stochastic atmosphere, and sudden discharge. The movement of ocean heat content on and off the equator in accordance with smoothly posed ocean dynamics and with atmospheric coupling has since become central to a more consistent theoretical interpretation of the ENSO cycle, summarized as the subsurface memory paradigm below.

The first linear stability study in a coupled shallow water system was presented by Philander et al. [1984]. With SST proportional to thermocline depth, with a proportionality constant independent of space, and a Gill atmospheric model, an initial SST anomaly was found to grow and extend along the equator. Modes that grow under such assumptions about SST propagate eastward, as shown by Hirst [1986]. Similar eastward propagation may be found in work by Yamagata [1985]. A mechanism for westward propagation involving zonal advection of an east-west SST gradient with warmer water in the west was noted by Gill [1985]. Hirst [1986, 1988] analyzed the effects of the various terms of an idealized SST equation in a linear modified shallow water model. He obtained eastward and westward propagating modes in both a zonally periodic case and cases with a basin of finite zonal extent. The propagation characteristics depended on the terms retained in the SST equation, with zonal advection tending to create an unstable westward propagating mode and the thermocline feedback tending to create an unstable eastward propagating mode.

At approximately the same time, Cane and Zebiak [1985] and Zebiak and Cane [1987] obtained sustained oscillations in an ICM of the form (1)-(4). The spatial form of these oscillations appeared reasonably close to the observed ENSO, and the amplitude of the oscillation was limited by nonlinearity. For some model parameters the amplitude of the oscillation behaved irregularly, in a manner reminiscent of observed ENSO time series. Figure 7 shows SST indices from this model in the regime CZ chose as their standard. By the end of TOGA, theoretical reasons for much of this behavior had been unraveled, as discussed in the following sections. Parallel work by Battisti [1988] showed the role of various feedback mechanisms in supporting the growth of SST anomalies during the development stages of a warming. Notably, several of the terms in the SST equation (3) cooperate to produce warming: zonal advection of warmer SST from the west, upwelling of anomalously warm subsurface temperatures associated with deeper thermocline in the eastern part of the basin, and reduction of upwelling due to reduced Ekman pumping as the easterlies weaken. Battisti’s model was very similar to the model by CZ, with slight parameter differences and improvement of some numerics, and yet tended to give much more regular oscillations. For instance, Battisti’s [1988] Figure 2 (not reproduced here) corresponding to Figure 7, has a curious repeating sequence of four El Niños every 13 years, with every fourth event being larger than the others. This difference between these two extremely similar models was not resolved until almost the end of TOGA, when it turned out to be symptomatic of a scenario for ENSO chaos.

The other major development of this period came when Cane et al. [1986] issued the first ENSO forecasts made with a physically based coupled ocean atmosphere model. At the time, this was considered controversial by some, since the mechanisms producing the oscillation in the CZ model were not yet understood, much less agreed upon as being the processes relevant to observations. Subsequent developments have seen evidence and consensus grow that such forecasts of seasonal-to-interannual variability have practical utility and a sound physical basis, as reviewed in Latif et al. [this issue]. Thus, while the theoretical basis for building ENSO models was used in early TOGA, most of our theoretical understanding of the resulting coupled oscillations, developed during later TOGA, actually followed the beginning of experimental ENSO predictions.
5.2. Mid-TOGA

5.2.1. SSBH delayed oscillator model. One of the most influential theoretical developments of the mid-TOGA period was a simple model developed by Schopf and Suarez [1988], Suarez and Schopf [1988] and Battisti and Hirst [1989]. We refer to this as the SSBH delayed oscillator model, since both groups presented similar derivations at nearly the same time. From an observational standpoint, Graham and White [1988] presented evidence that the temporal relations among western Pacific island sea levels, central Pacific zonal wind, and eastern Pacific SSTs are consistent with a coupled oscillator, with a similar conceptual picture.

Figure 8 shows results from the Schopf and Suarez [1988] intermediate model, which they used to argue heuristically for a mechanism that might provide oscillation. SST anomalies develop largely in place in the eastern Pacific, and zonal wind stress anomalies in the central basin are essentially slaved to the SST. The memory of the oscillation between phases of warm and cold SST must be provided by ocean subsurface adjustment processes. In work by Schopf and Suarez [1988], these were idealized as being due to individual free Kelvin and Rossby waves. In retrospect we can recognize that the ocean surface height, projected on Kelvin wave meridional structure in Figure 8b, shows relatively little sign of the free Kelvin wave phase speed and that the apparent eastward propagation is better explained by the west-leads-east structure of the thermocline response discussed in section 3. The sum over Rossby wave meridional structure contributions to off-equatorial surface height in Figure 8d does show some indication of westward propagation characteristic of evolving Rossby wave packets to the west of the wind stress. BH described a continuous cyclic forcing of ocean signals in the western Pacific, with free Kelvin and Rossby signals in the western basin, much like that displayed in Figure 8. The essence of the SSBH postulate is that the delay for propagation of Rossby waves from the wind stress region to the western boundary and the return of this signal as a reflected Kelvin wave to the eastern basin provide the memory for the oscillation.

SSBH used versions of the following differential delay equation to represent local coupled feedback processes in the eastern basin, being influenced by the return effects of a single Rossby wave reflected as a Kelvin wave from the western boundary

\[
\frac{dT'}{dt} + \left[ T' - T_{\text{sub}}'(h') \right] = 0 \quad (12a)
\]

\[
h'(t) = \mu [b_0 T'(t) - b_1 T'(t - 4\delta)] \quad (12b)
\]

where the subsurface temperature \( T_{\text{sub}}' \) may depend nonlinearly on \( h' \) as in the ICM; \( T_{\text{sub}}' = \gamma h' \) when linearized; \( \mu \) is the coupling coefficient; time has been

![Figure 8](image-url)

**Figure 8.** Time-longitude plots of anomalies from the Schopf and Suarez [1988] intermediate coupled model over 15 years of simulation. Negative anomalies are hachured and contour intervals are given in brackets; (a) equatorial SST (0.5°C); (b) ocean surface height projected on Kelvin wave meridional structure (1 cm); (c) equatorial low-level zonal wind (0.2 m/s); and (d) ocean surface height associated with Rossby wave meridional structures averaged between 5° and 7°N/S (1 cm). Ocean surface height is approximately proportional to thermocline depth. After Schopf and Suarez [1988].
nondimensionalized by the mean upwelling timescale for SST, \((\langle w / H_1 \rangle)\); and 4δ is the crossing time between the western boundary and the wind/SST point (with the westward Rossby wave taking 3 times as long as the eastward Kelvin wave). The coefficients \(b_0\) and \(b_1\) are the projections of the meridional shape of the wind stress forcing on the Kelvin wave and the negative of the projection on the first Rossby wave meridional structure, respectively (both positive). The ratio \(b_1 / b_0 = 0.5 \exp(-4\epsilon)\). For an inverse damping time, \(\epsilon \approx 250\) days, a coupling point mid-Pacific and wind constant in \(y\), this gives \(b_1 / b_0 \approx 0.3\). BH use a value of \(0.6\) on the basis of fitting the simple model to an intermediate model.

The form and derivation used here is in essence the same as BH but follows the Neelin et al. [1994] review in deriving the SSBH model from the ICM (1)-(4) by way of a more complete “point-coupling” model due to Münich et al. [1991]. The derivation goes as follows: (1) assume that SST anomalies are highly localized near a single point on the equator in the eastern basin upwelling region; (2) assume that the wind stress is localized near a single point slightly to the west of the SST point (by a distance small enough that the Kelvin transit time between them is negligible); (3) remove the eastern boundary of the basin (this does irreparable damage to the uncoupled ocean dynamics but less to the strongly coupled system); (4) truncate the meridional structure to include projection only on a single Rossby wave structure. If the last two assumptions are not made, (12) becomes the Münich et al. model, and (12b) has an infinite sum over higher Rossby wave terms, \(b_n\), with longer delay times, and an additional infinite sum due to reflection from the eastern bound-

One of the questions that the SSBH model at first appears to answer and then poses as a deeper mystery is “what sets the timescale for ENSO?” In the delayed oscillator model, there are two clear time scales: one for adjustment of SST and another for wave transit time. The period from the model, however, is not closely related to either of these, ranging from order of 2 years.
to infinity within a reasonable range of parameters in Figure 9. Another curiosity is that, because infinite sums of Rossby waves add up to something sensible, Cane et al. [1990] were able to obtain a simpler dispersion relation from a fuller point-coupling model. Schopf and Suarez [1990] also discuss a more detailed point-coupling model.

5.2.2. Linear versus nonlinear period. An important contribution of BH, independent of the delayed oscillator equation, was to establish that the period of the nonlinear oscillation in the CZ ICM is close to that of the leading eigenmode of the linear problem (aside from technical details about discontinuous derivatives in the CZ version). At the same time, Neelin [1989b, 1990] noted that the ENSO oscillation in an HCM arose as a supercritical Hopf bifurcation [see Guckenheimer and Holmes, 1983]. This occurs commonly in nonlinear systems when an oscillatory eigenmode (of the system linearized about a steady solution) becomes unstable as a parameter (here the coupling coefficient $\mu$) is changed. The bifurcation occurs at the value of the coupling where the mode becomes unstable, and the system changes from having a stable steady solution to having a nonlinear limit cycle about the unstable steady solution. The spatial structure and period of the nonlinear oscillation is approximately determined by the linear oscillation at the bifurcation. This is useful because the mechanisms determining oscillation and spatial pattern can thus be studied in the linear system. For moderately supercritical values of coupling, the nonlinearity simply balances the instability to create a finite amplitude cycle. The nonlinear period tends to remain closer to the linear period at the bifurcation than to the linear period at the same parameter values. In later TOGA, JN showed the same for a CZ-like model, and Jin et al. [1996] showed that the linear spatial form and period still dominate even in chaotic regimes.

The Suarez and Schopf [1988] version of the delayed oscillator had been in a regime where a purely growing unstable mode leads to multiple equilibria, and the oscillation appears more complex (see Dijkstra and Neelin [1995] for an ICM version; Neelin and Dijkstra [1995] later showed that the multiple equilibria are a model artifact). BH were able to argue that the realistic range for both the ICM and the SSBH model was in the oscillatory regime. Their work makes a clear case for the role of subsurface adjustment in providing a memory for ENSO, through a complex interaction with coupled processes. As modified by later work, we refer to this general concept as the "subsurface memory paradigm."

Figure 10 shows the SST and thermocline depth anomalies over one period of the simulated ENSO cycle from the linearized version of the CZ ICM used by BH to examine the essential dynamics. The thermocline depth may be compared with the ocean model forced by periodic wind stress in Figure 3. Details of the transition between west and east differ from Figure 3 and from observed because the BH simulated winds are shifted eastward relative to observed, but the cycle is not strongly sensitive to this. The typical stationary oscillation in SST may be seen, with the lead of the western basin thermocline-depth anomaly rela-
tive to the eastern basin characterizing the subsurface memory. During the phase when wind stress passes through zero the thermocline depth signal is small on the equator, since the model is close to steady balance along the equator. The memory of the system is in the thermocline depth off the equator to the west of the wind, as seen in Figure 5. As these off-equatorial anomalies are gradually fed onto the equator along the western boundary, they provide a sustained tendency of decreasing thermocline depth on the equator during a transition from warm phase to cold phase and vice versa.

5.2.3. SST modes and propagating variability. Although observed ENSO SST anomalies tend to appear in the cold tongue region with relatively little signature of eastward or westward propagation along the equator, many models continued to exhibit ENSO-like variability with marked propagation. This includes the ICM of Yamagata and Masunogo [1989], GCMs by Meehi [1989, 1990], and Lau et al. [1992], and some of the other models collected in the intercomparison of Neelin et al. [1992]. Neelin [1991] presented a theoretical case in which the interannual variability arose as modes associated with the time derivative of the SST equation, referred to as “SST modes.” To obtain simplified results, the limit in which wave dynamics is assumed to bring the ocean quickly to adjustment on the timescale of SST change was used, referred to as the “fast-wave limit,” $\delta \to 0$ in (1). Otherwise the model is roughly the same as the ICM (1)-(4), with an equatorial strip approximation in the SST equation (3). The thermocline solution to (1) along the equator is given by ($6$'). These SST modes in the fast-wave limit were intended to provide a case where timescales of subsurface ocean adjustment are explicitly not important, even though the thermocline feedback is included. Indeed, some of the resulting interannual variability looks surprisingly realistic in a nonlinear fast-wave limit model in a finite basin [Hao et al., 1993]. Later it was shown that the SST modes and the modes represented in the SSBH model are in fact closely related, as discussed below. SST modes provide a handy prototype for examining variability, whose period is dominated by propagation along the equator, and thus do not depend crucially on delays due to subsurface adjustment.

Typical mechanisms associated with propagating variability include competition between surface-layer feedbacks and the thermocline feedback. From the atmospheric model, westerly wind anomalies tend to lie over and to the west of an SST anomaly. From this the various feedbacks are associated with terms in the linearization of the SST equation (3). In the surface-layer feedback the surface-layer eastward current and downwelling anomalies occur under the westerlies and thus tend to reinforce the original anomaly and shift it to the west by SST terms $u'_w \partial_z T$ and $u'_w (T - T_{sub})/H_{1.5}$. To the east of the original warm anomaly, easterly winds tend to create cold anomalies by this mechanism, potentially resulting in a westward propagating succession of warm and cold anomalies. In the thermocline feedback, on the other hand, the thermocline slope tends to balance the wind stress, resulting in deeper thermocline that is under and to the east of the original SST anomaly. This creates warm subsurface temperature anomalies that are carried to the surface by mean upwelling, $\bar{w} T_{sub}/H_{1.5}$, reinforcing the original anomalies and tending to move them toward the east. This can tend to create an eastward propagating succession of warm and cold anomalies.

Tendencies by both feedbacks are indicated diagrammatically in Figure 11, which is drawn for the more ENSO-like case, where the mode does not propagate but undergoes stationary growth. This occurs in a finite basin if the atmospheric length scales are long enough and if the surface-layer feedbacks are not too large, as elaborated in later TOGA by Hao et al. [1993] and JN. The thermocline feedback can create nonpropagating modes because of the effects of basin boundaries in both atmosphere and ocean. In Figure 11, the easterly winds to the east of the warm anomaly lie over land and thus do not produce the cold anomalies that would be required for a propagating mode. The westerly wind anomalies produce an average deepening of the thermocline on the equator, as seen from ($6'$) in section 3.2. This favors the growth of the warm SST anomaly in the eastern basin that in turn produces the westerlies in the central basin. This internal determination of the spatial structure favors eastern basin anomalies even without taking into account the effects of the climatology that further favor the eastern basin. The mechanisms in Figure 11 are the same as noted by Battisti [1988] for the maintenance of warm anomalies in an ENSO warm phase. This is, of course, no coincidence, as was found when subsurface memory and SST mode trains of thought were brought together in later TOGA.

![Figure 11. Schematic of warming mechanisms that amplify and maintain SST anomalies during an ENSO warm phase. The converse applies during cold phases. The mechanisms are drawn here for a balance of processes and an internally determined spatial pattern that would give a purely growing SST mode if ocean dynamics did not supply the memory needed for oscillation. A change in balance of mechanisms could produce eastward or westward propagating modes.](image-url)
5.3. End of TOGA and Beyond

5.3.1. Mixed SST–ocean-dynamics modes. Because the response to a periodic wind stress forcing involves a complex sum of Rossby and Kelvin waves as discussed in section 3.2, Chao and Philander [1992], argued that the delayed oscillator model required generalization. Wakata and Sarachik [1991a] pointed out that the eastward propagation of modes with large thermocline feedbacks in the SST equation is sensitive to the meridional extent of the upwelling. They noted that models that replace equatorial upwelling by a coefficient independent of latitude, such as Hirst [1988], favor eastward propagating modes, whereas for narrow equatorial upwelling these modes can transition to a regime where SST has a standing oscillation and subsurface memory is thus crucial to the period.

JN examined the relations among coupled modes in different parameter regimes in an ICM similar to (1)-(4). The most realistic ENSO oscillation regime follows CZ and Battisti [1988] in having a standing oscillation in SST with subsurface memory carrying the oscillation between phases. We refer to this for brevity as the standing SST–oscillatory (SSO) regime. By exploring the relation of the SSO regime to regimes with slightly different behavior or to regimes in which the behavior is simpler to understand, they argued that various views of ENSO could be unified. In particular, the qualitative relation of the SSBH delayed oscillator model to the ICM could be understood, establishing the SSBH model as a good metaphor for aspects of the behavior in the most realistic SSO regime. The main parameters used to cut quickly across regimes of behavior are those presented in section 2: coupling coefficient $\mu$, relative timescale coefficient $\delta$, and surface-layer coefficient $\delta_s$. The main results can be summarized as the following:

1. Weak coupling is not a useful limit for understanding realistic regimes. The transition from uncoupled ocean dynamics modes (section 3.2) and uncoupled SST modes to behavior at realistic coupling involves many complex mergers between modes that radically change the behavior.

2. The case where wave adjustment times are relatively short compared to overall coupled timescales (the fast-wave limit case) loses the important source of oscillation due to subsurface memory but is useful for studying growth and mechanisms governing spatial structure of the coupled modes. A stationary (purely growing) SST mode with the same spatial structure connects to the oscillating mode with standing oscillation in SST in the most realistic regime.

3. At strong coupling, local growth mechanisms tend to dominate over basin adjustment processes, so the mode of the SSO regime connects to a purely growing mode with similar spatial structure. This mode is essentially a stationary SST mode. The transition curve marked A in the dispersion diagram for the SSBH delayed oscillator (Figure 9) corresponds to this transition in the ICM.

4. The SSBH model can be viewed as taking an SST mode in the fast-wave limit and perturbing it by adding a simple representation of ocean dynamical memory to capture oscillations in the SSO regime.

5. Modest changes in parameters can create eastward or westward propagation tendencies within the same mode that gives the ENSO oscillation in the SSO regime. It is thus not surprising that some observed ENSO phases exhibit slight propagation while others do not. It also implies that models that are dominated by propagation may not be entirely off the mark but may simply have a slightly unrealistic balance of mechanisms.

5.3.2. Demise of multiple equilibria. Many ENSO anomaly models exhibit multiple equilibria if the coupling is strong enough [Suarez and Schopf, 1988; McCreary and Anderson, 1991; Wu et al., 1993; Wakata and Sarachik, 1994; Dijkstra and Neelin, 1995]. These multiple equilibria at first seemed to arise from physical mechanisms associated with purely growing modes. A warm SST anomaly produces westerly winds that reduce upwelling and deepen the thermocline in the east, producing an amplifying instability that equilibrates through nonlinear terms to a warm stationary state and vice versa for a cold state. When these multiple equilibria are mapped out in phase space, they turn out to be associated with transcritical bifurcations from the constructed climatological state in flux-corrected models (see (5) in section 2). Such bifurcations are not robust to relaxation of the flux correction and disappear when the climatology is simulated rather than specified [Neelin and Dijkstra, 1995]. This is because the warm state in flux-corrected models is created by an opposition of westerly winds, $\tau'$, from the anomaly atmosphere, canceling easterly winds, $\bar{\tau}$, introduced artificially by flux correction. When there is no flux correction, the physical feedbacks described above act to make the cold tongue stronger but not to produce multiple equilibria.

5.3.3. Subsurface memory paradigm. Evidence that subsurface memory is the dominant source of oscillation in the observed ENSO accrued during this period. Latif and Graham [1992] expanded on findings by White et al. [1987] that considerable predictability is associated with subsurface thermal structure. Kleeman [1993] found that the prediction skill of a model with parameters in the SSO regime was considerably greater than that for the same model with different parameters (i.e., with larger $\delta_s$) that shift the interannual mode to a propagating regime. Barnett et al. [1993] found that the simulation and prediction skill in an HCM was consistent with the role of ocean heat content. Ji et al. [1994] found that assimilation of subsurface ocean data improved forecast skill of a coupled GCM. Li and Clarke [1994] raised an apparent conundrum by noting that western boundary sea level (using northwestern Australia sea level data) is not strongly positively correlated with later wind changes, as it should be if subsurface memory associated with western boundary
reflections causes later changes in SST and thus in wind. Mantua and Battisti [1994] partly resolved this by showing that such lag correlations are not large in ocean models forced by observed winds or in ENSO models with irregular time behavior, even when the latter conform to the subsurface memory mechanisms. White et al. [1989], Kessler [1991], and Wakata and Sarachik [1991b] argued for western boundary reflections associated with subsequent changes in ENSO phase from data and ocean model diagnostics, respectively. Graham and White [1991] and Battisti [1989] argued that Rossby waves poleward of 6° could or could not play an important role in ENSO, respectively, but either conclusion would fit within the subsurface memory paradigm.

Figure 12 shows anomaly patterns from coupled GCM results [Latif et al., 1993b]. Principal oscillation pattern (POP) analysis [Hasselmann, 1988] has been used to estimate the dominant coupled mode, choosing crucial fields according to the subsurface memory theory. Oscillations are represented by a cycle of patterns in temporal quadrature. Figures 12b, 12d, and 12f show conditions during the warm phase of the ENSO cycle (cold conditions have reversed sign); Figures 12a, 12c, and 12e apply 90° of temporal phase prior to the warm phase. The mode has a period of just under 3 years in the GCM. During the warm phase, warm subsurface temperatures (Figure 12f), lead to warm SST in the upwelling region (Figure 12d), which produces westerly winds along the equator in the atmosphere (Figure 12b). The winds in the GCM are shifted eastward compared to observations, which are centered at the date line, but allowing for this, the ocean surface height patterns resemble those shown in Figure 4 for a shallow water model forced by periodic winds. Latif et al. [1993a] used the wind pattern of Figure 12b with a specified period to force the ocean model and obtained good agreement in spatial pattern of ocean height anomaly (not shown) with the GCM. We can thus use the breakdown of Figure 4 into instantaneous response (Figure 2) and ocean memory (Figure 5) to qualitatively discuss the GCM results. During the warm phase in Figure 12f the ocean height would be substantially in balance with the wind stress, especially along the equator, but with slight departures associated with the ocean memory as in Figure 5d. The seeds of the next cold phase are thus already sown. During the transition phase in Figure 12e the pattern is dominated by ocean memory from the previous cold phase, with off-equatorial heat content being fed slowly back onto the equator to carry the oscillation into the subsequent warming. Similar mechanisms were diagnosed using different techniques in coupled GCMs by Philander et al. [1992], Nagai et al. [1992], and others [see Delecluse et al., this issue].

Figure 12. Spatial patterns of the dominant mode of ENSO variability from the coupled GCM of Latif et al. [1993b], as represented by the leading principal oscillation pattern (see text). The oscillation is represented by two time phases in quadrature during the cycle. Signs for the extreme phase are for ENSO warm conditions; signs for the transition phase are for conditions preceding the warm phase. (a, b) Wind stress anomaly (c, d) sea surface temperature anomaly, and (e, f) ocean height anomaly (approximately proportional to upper ocean heat content anomaly, related to thermocline depth). From Latif et al. [1993b].
tually be possible to verify directly the role of subsurface dynamics in producing the dominant ENSO spectral peak at around 3-5 years period. With the current shorter time series of subsurface ocean surface height data, examining Rossby and Kelvin waves at shorter timescales is possible [e.g., Fu et al., 1991; Picaut and Delcroiz, 1995; Delcroiz et al., 1991, 1994; Busalacchi et al., 1994]. This provides a way of checking the dynamics of ocean models and the balances of particular events, but it remains a considerable challenge to scale up to the longer timescales.

5.3.4. End of TOGA models. The view of ENSO-related variability as mixed SST-ocean dynamics modes helps unify our understanding but is too complex for teaching purposes or for communication to nonspecialists. The SSBH delayed oscillator model is useful for this, but the length of caveats is lengthy, and the dispersion relation is not easy to work with. Since we know that ENSO arises as a Hopf bifurcation in models, this implies that a simple model for ENSO can be constructed that has only two ordinary differential equations (ODEs). Formally, this could be done by the method of normal forms [Guckenheimer and Holmes, 1983], though this might not clarify the underlying physics. The route to chaos discussed in the next section also strongly suggests that ENSO chaos currently found in ICMs can be reproduced in a simple model with just two ODEs and seasonally varying coefficients. There are several current attempts to produce such models. The challenge is to create the model by derivation from an ICM rather than by ad hoc assumptions.

A recent simple model by Jin [1997a, b] may be written

\[
\frac{dT'}{dt} = RT' + \mu \gamma h'_w \tag{13a}
\]

\[
\delta \frac{dh'_w}{dt} = -\tau h'_w + \mu \alpha T' \tag{13b}
\]

where \(T\) is an average SST anomaly over the eastern equatorial basin; \(R\) is an average over surface flux, upwelling, and surface-layer feedback terms proportional to SST in this region in the linearization of the SST equation (3); and \(\gamma = \frac{\delta T_{sub}}{dh}\). The equation for thermocline evolution (13b) is written in terms of the thermocline value at the western boundary and approximates the first-order departures from Svendrup balance (6') due to ocean adjustment, with the atmospheric model simply a fixed wind stress anomaly pattern proportional to \(T\); \(\alpha\) and \(\tau\) depend on the wind stress pattern and the ocean dynamics. (The relative adjustment time coefficient, \(\delta\), is as in (1)). This model gives interannual periods of 3-5 years for reasonable parameter estimates. A model by Wang and Fang [1996] shares some characteristics, although the derivation differs. Both the Jin and the Wang-Fang models are clearly for mixed SST-ocean dynamics modes; one time derivative comes from SST, the other from ocean dynamics. This implies that they differ fundamentally from the SSBH delayed oscillator, which can oscillate due to ocean memory even when the SST adjustment time is fast. Specifically, the SSBH model has oscillations even for large \(\delta\) in Figure 9, whereas (13) would have pure growth or decay modes.

Schneider et al. [1995] examined departures of an ocean GCM from equilibrium with ENSO wind stress forcing in a clever experiment. They specify observed wind stress time series in one integration and compare it to a second integration in which the wind stress series is reversed in time. The part of the ocean response that is essentially in steady balance with the wind stress is the same in both runs, and this tends to make the main warm and cold phases similar in both runs. The difference between the runs gives a measure of the ocean memory, in the same spirit as illustrated in Figure 5, by removing the instantaneous balance contribution. Schneider et al. focused on the role of the zonal average heat content along the equator in the ocean memory (as originally discussed by CZ, and Zebiak [1989]). Considering this in terms of the shallow water solution in Figure 5, the ocean memory contribution to the thermocline depth varies strongly in \(z\) off the equator, but tends to be more constant in \(z\) on the equator since Kelvin adjustment times across the basin are fast. Schneider et al. assumed a delay relation for thermocline depth at the western boundary, \(h_w\), of the form

\[
h_w = -A \int_0^1 \tau(x_0, t - s)dx_0 \tag{14}
\]

to provide a boundary condition to (6), where the delay time \(s\) and the parameter \(A\) were obtained by fitting to OGCM experiments. They noted that while (14) seemed to give a reasonable fit to the uncoupled OGCM, in a simple coupled model a low-ENSO frequency could only be obtained if \(A\) depended on frequency. We also note that (14) is not consistent with the shallow water equation solution for periodic winds (7) nor its steady limit (6'). Despite such technical difficulties in finding an optimal simple model, the Schneider et al. [1995] OGCM experiments provide corroboration of the subsurface memory paradigm. They appear to disagree with the Jin [1997a] model and Wang and Fang [1996] models of the form (13).

Delcroiz et al. [1994] and Boulanger and Menkes [1995] argued on the basis of altimeter data that local wind forcing rather than western boundary reflection appeared to be responsible for most of the Kelvin wave projected sea level signal during 1986-1989 and 1992-1993 respectively. Kessler and McPhaden [1995], using data from the TOGA-Tropical Atmosphere-Ocean (TAO) buoy array and expendable bathythermographs over 1988-1993 and comparing to simplified, wind-forced Kelvin and Rossby models, argued that the ENSO cycle must be more complex than the idealization presented in the SSBH delayed oscillation model. They argue that while signatures associated with upwelling/downwelling Rossby waves arriving at the western boundary appear
to play a role in the termination of the 1991-1992 El Niño, the onset of the 1991-1992 and 1993 warm phases appear not to be consistent with simple wave-reflection mechanisms. Springer et al. [1990] note an opposition between Ekman pumping and geostatic contributions to heat content changes on and off the equator in the western Pacific. Weisberg and Wang [1997] postulate that off-equatorial SST anomalies, initiating winds in the far western Pacific of opposite sign to winds in the central Pacific, might play an overlooked role in the ENSO cycle. McPhaden and Picaut [1990], using velocity measurements, Picaut and Delcroix [1995], using Geosat-derived currents, and Picaut et al. [1996], using several current data sets and ocean models, demonstrated the importance of zonal advection in the 1986-1994 succession of warm and cold phases, in addition to direct effects of vertical displacements of the thermocline. The mixed SST-ocean dynamics mode view of ENSO variability suggests that such apparent deviations from delayed oscillator idealizations are likely to be resolvable since multiple mechanisms can contribute cooperatively to the ENSO mode. The challenge to the theoretical community is to produce a consensus simple model that retains sufficiently detailed physics and spatial structure to directly confront such observations.

6. ENSO Irregularity and Interaction

With the Seasonal Cycle

While the oscillatory tendency of ENSO is now reasonably well understood, the origin of its irregularity is currently a major question. Spectra of observed ENSO time series have power at all frequencies, associated with the irregularity, but with preferred timescales giving rise to broad spectral peaks at roughly 3-5 years and, arguably, around 2 years. Rasmussen et al. [1990] found evidence for a quasi-biennial and a lower frequency 3-6-year peak. Jiang et al. [1995], using multichannel singular spectrum analysis, corroborate this split of the ENSO variability and refine the low-frequency peak to be quasi-quadrennial. Interaction of quasi-quadrennial and quasi-biennial bands has been examined by Barnett [1991]. Also at issue is the interaction of ENSO with the annual cycle. The tendency of ENSO to phase lock to the seasonal cycle has long been known, although aspects of the seasonal locking may vary from one decade to another [Mitchell and Wallace, 1996]. Predictability of El Niño can exhibit seasonal dependence [Cane et al., 1986; Blumenthal, 1991; Webster, 1994]. Most intermediate coupled models have ENSO cycles with an internally determined interannual period when the annual cycle is suppressed. Often, this period becomes frequency locked to some rational multiple of a year when the annual cycle is included [Battisti, 1988; Barnett et al., 1993; Syu et al., 1995].

The two major contenders as sources for ENSO irregularity are (1) deterministic chaos within the nonlinear dynamics of the "slow" components of the coupled system [Münich et al., 1991; Jin et al., 1994; Tziperman et al., 1994; Chang et al., 1994] and (2) uncoupled atmospheric "weather noise." By the latter we mean atmospheric variability with relatively short decorrelation times, on the order of a month or less, associated with synoptic-scale systems, convection, and so on [Hasselmann, 1976]. The atmosphere's statistically steady response to changes in oceanic boundary conditions is considered part of the slow coupled dynamics. Fortunately, in the tropics it is possible to model this steady response directly in intermediate coupled models, and thus these effects can be studied separately. The main deterministic chaos hypothesis for ENSO irregularity is intimately associated with the nonlinear interaction of the ENSO mode with the seasonal cycle. We therefore discuss first some basic principles of the ENSO-seasonal cycle interaction then the routes to chaos. The weather noise hypothesis does not depend essentially on the seasonal cycle interaction, but is affected by it, and is discussed last.

6.1. ENSO Interaction With the Annual Cycle

It has long been known that the annual cycle is strongly involved in the evolution of ENSO [Rasmussen and Carpenter, 1982; Philander, 1990], but much of the progress in understanding ENSO mechanisms over the past decade has been made in models with no annual cycle present [e.g., BH, Schoop and Suarez, 1988; Yamagata and Matsumoto, 1989; JN; Philander et al., 1984; Hirst, 1988; Wakata and Sarachik, 1991a; Neelin, 1991; Suarez and Schoop, 1988]. There has been considerable confusion in the ENSO literature regarding how to relate stability analysis of time independent states to linear and nonlinear models that do include the annual cycle. One approach has been to perform linear stability analysis of a time independent state similar to the climatology of a particular month of the seasonal cycle, for instance, perpetual October [Zebiak and Cane, 1987; Battisti, 1988; BH; Tziperman et al., 1995, 1997]. This approach usually typifies one season as having the most unstable ENSO mode and another season as having a similar ENSO mode most strongly decaying. Useful insights into the system have been obtained from this approach, although such analysis can be rigorously justified only if the evolution timescale for the eigenmode is much faster than the changes in the seasonal cycle; this assumption is violated for the ENSO period, which is longer than a year. However, basic mechanisms by which the annual cycle affects SST are known: a main factor is that in spring, stratification tends to be large and upwelling small, while in fall, the converse holds [Battisti, 1988; Webster, 1994]. Tziperman et al. [1997]; Xie [1995] emphasize the importance in the CZ model of the climatological atmospheric convergence zones, which tend to promote destabilizing feedbacks in spring when the CZ model has specified climatological convergence over the equator in the eastern
Floquet theory [Hartman, 1982; Iooss and Joseph, 1990; Strong et al., 1995] provides a consistent approach to the analysis of eigenmodes about a time-periodic state. Floquet analysis is not yet routinely used in climate sciences, so a brief description is included here (for details, see Iooss and Joseph [1990]). The Jacobian matrix evaluated about the climatological annual cycle orbit, $M(x, t)$, depends on space $x$ and is periodic in time period $P = 1$ year. In Floquet analysis one simply integrates

$$\frac{dQ(t)}{dt} = M(t)Q(t)$$

through one period and calculates the eigenvalues of $Q(P)$, termed the monodromy matrix. The logarithm of these eigenvalues gives the Floquet exponents, $\sigma_j$, which, in the ENSO application, determine the interannual period of the coupled modes, as well as the growth rate. The full space and time dependence of the modes is given by

$$V_j(x, t) e^{\sigma_j t}$$

where $V_j(x, t)$ is the $j$th eigenvector. As in standard eigenanalysis, the eigenvectors determine the spatial dependence of the modes, but in Floquet analysis they also have a periodic time dependence. They are obtained by integrating through one cycle

$$\frac{dV_j(t)}{dt} = (M(t) - \sigma_j)V_j(t)$$

starting from the eigenvector of the monodromy matrix associated with $\sigma_j$.

When Floquet analysis is used to examine the eigenmodes of the tropical ocean-atmosphere system linearized about a climatological state that includes the seasonal cycle [Jin et al., 1996], the results are gratifyingly simple. The eigenmodes are quite similar to the modes of the annual-average case: in interannual period and in the combined spatial and temporal evolution given by the eigenstructure. The annual cycle in the basic state modulates the interannual modes and slightly increases their space-time complexity. The Floquet analysis thus suggests that linear theory based on the annual-average basic state does capture the fundamental coupled dynamics of interannual ENSO modes. 

This puts a decade of ENSO theory based on linearization about such a state on firmer ground. At the same time, it gives a framework for understanding linear effects associated with the annual cycle. From (16) the amplitude of the SST component in the eigenvector has seasonal dependence; in seasons where this is larger the interannual evolution given by $e^{\sigma_j t}$ can produce larger extrema, thus giving a preferred season for large warm or cold phases. By (17) this is linked to an integration through seasons where the SST equation terms in $M(t)$ favor increases. Quantitative examination of these effects may be useful in the future.

The instability of the annual cycle periodic orbit gives rise to the interannual ENSO mode via a bifurcation that is closely related to the Hopf bifurcation found in the annual average case. The basic spatial structure and mechanisms in the linear results carry over to the nonlinear case in some neighborhood of the bifurcation [Nayfeh and Balachandran, 1995]. In practice the linear modes largely determine the structure and dominant interannual timescale, even in cases with realistically strong nonlinearity [Jin et al., 1996]. However, the nonlinear interactions with the annual cycle do create important changes in exactly how the ENSO mode evolves in time. Over wide regions of parameter space, nonlinear interaction with the annual cycle modifies the frequency of the ENSO mode to a rational fraction of the annual frequency. Such frequency locking is a very common phenomenon in nonlinear systems [e.g., Iooss and Joseph, 1990], and since only a modest change to the ENSO frequency is necessary, locking can occur even for weak nonlinearity. Frequency locking was noted in many ENSO models [e.g., Battisti, 1988; Barnett et al., 1993; Syu et al., 1995] long before its significance to ENSO chaos, discussed in section 6.2, was understood.

### 6.2. ENSO Chaos

Zebiak and Cane [1987] noted that irregular behavior could be found in their model through deterministic coupled dynamics alone. Concerns that numerically induced noise contributed to this have since been obviated by reproduction of chaos in other models. Bifurcations toward more complex behavior were noted in an HCM [Neelin, 1990], but the first clear demonstration of a bifurcation sequence into chaotic ENSO behavior was given by Münich et al. [1991] in a point-coupling model. Earlier consideration of chaotic behavior in an ad hoc model [Vallis, 1986] turned out not to be physically based.

Jin et al. [1994] and Tziperman et al. [1994] found independently that nonlinear interaction of the annual cycle and the coupled ENSO mode leads to ENSO chaos in an intermediate and a simple model, respectively. Both noted a transition by the quasi-periodicity route to chaos, found in periodically forced nonlinear systems [Jensen et al., 1984; Bak, 1986, and references therein]. Chang et al. [1994] examined the transition to ENSO chaos in a slightly different intermediate coupled model via a parameter that controlled the amplitude of the annual cycle and described the scenario as period doubling. More complete results [Chang et al., 1995] suggest that when the annual cycle is near realistic amplitude, the transition to chaos in that model is consistent with the quasi-periodicity route. Period-doubling sequences of phase-locked oscillations as a subcase of the quasi-periodic route have been noted in a number of systems [see Nayfeh and Balachandran, 1995]. There is discussion about whether chaotic behavior in the CZ model follows the quasi-periodicity scenario [Tziperman,
et al., 1995] or arises instead by interaction between two distinct modes [Mantua and Battisti, 1995]. Tziperman et al. [1994] and Chang et al. [1996] use dimension estimation methods to confirm low-order chaos in ICMs.

The quasi-periodicity route to chaos requires two parameters to understand: one that affects the inherent frequency of the ENSO oscillation relative to the annual cycle and one that affects the strength of nonlinearity. One choice for the first in an ICM is the surface-layer coefficient $\delta_s$ (see section 2), which tends to smoothly increase the ENSO frequency when the annual cycle is absent. The coupling coefficient, $\mu$, increases the amplitude of the oscillation and hence the nonlinearity. As nonlinearity increases, so does the tendency of the ENSO cycle to frequency lock to rational fractions of the annual frequency. This gives wider bands of frequency-locked behavior as $\delta_s$ is varied; when these bands overlap, chaos can ensue, as the system jumps between the various subharmonic resonances.

In ICMs the transition to chaos occurs at very low amplitudes of the ENSO oscillation, and the question becomes: which behavior is more typical, chaos or frequency locking? Plate 1 shows a mapping of behavior regimes from more than 3000 runs of 500 years each from the JN ICM. Frequency-locked regimes with frequency ratios $1/n$ corresponding to one ENSO cycle every 5, 4, 3 or 2 years dominate the plot. Very small regions with frequency ratios $m/n$, e.g., three ENSO cycles every 10 years, are found at lower coupling values, just above the primary bifurcation. Chaotic regions are confined to relatively narrow slivers between the very stable $1/n$ subharmonic regimes. Parameters are uncertain within a subregion of the plot that gives dominant periods in the 3- to 5-year range. Since frequency-locked behavior covers more area than chaotic regimes, this model suggests a greater likelihood that the real system would fall in a regime that is frequency locked than in a regime with chaos in the slow components of the system. In frequency-locked regimes, noise becomes the default explanation for ENSO irregularity. However, the behavior in Plate 1 can be model dependent, so this remains an open question. Furthermore, the extent to which parameters that are fixed in this model change from decade to decade in observations is poorly known.

6.3. Stochastic Forcing by Weather Noise

It is clear that there exists significant variability with short decorrelation times in the atmosphere. We remind the reader that at timescales sufficiently longer than the typical decorrelation time the spectral signature of such processes appears as an approximately white noise, which can then act on the slower components of the climate system. This can occur via linear mechanisms for directions in phase space that are stable, in which case, standard linear filtering theory can provide some insights. For nonlinear stochastic differential equations, see, e.g., Gardiner [1985].

Noise forcing has been considered at various times during TOGA, but recently, there have been attempts to quantify its effect. Zebiak [1989a] found that random forcing similar to westerly wind bursts or the atmospheric 30-60-day oscillation had only modest effects in the CZ model. Penland and Sarineshuk [1995] argued using an empirical model based solely on SST data [Penland and Magorian, 1993; Penland and Matrosova, 1994] that ENSO variability is due to several decaying modes maintained by atmospheric stochastic forcing. Kleeman and Power [1994] suggested significant error growth within only 4 months due to stochastic forcing in an ICM predictability study. Recently, at least six studies have independently underlined the importance of weather noise in ENSO irregularity. Blanke et al. [1997] and Eckert and Latif [1997] both used HCMs to evaluate effects of realistic stochastic forcing based on Florida State University (FSU) wind stress observations [Legler and O'Brien, 1984]. Kleeman and Moore [1997], Jin et al. [1996], Chang et al. [1996], and Flügel and Chang [1996] used ICMs with various estimates of stochastic forcing. Eckert and Latif, like Kleeman and Power [1994], estimated their stochastic forcing by high-pass filtering observed time series. This was intended to remove atmospheric variance associated with SST forcing, although this filtering omits some variance that should be associated with the noise process. Blanke et al. took an alternate approach to estimating the noise forcing by removing variance associated with SST from the observed wind stress record using a linear empirical atmospheric model and then using random picks among the maps from the remainder time series to create a white noise product preserving spatial correlations. Kleeman and Moore compared both high-pass-filtered and remainder time series methods using daily datasets from European Centre for Medium-Range Weather Forecasting (ECMWF) analyses, creating a noise product from the empirical orthogonal functions (EOFs) of each series. All these methods depend on the accuracy of the observed data sets, since they cannot distinguish observational error from atmospheric internal variability. Chang et al. [1996] and Jin et al. [1996] used more idealized stochastic products, where the scaling of the noise amplitude was estimated more roughly Chang et al. used higher EOFs of observed wind with the variance doubled, and Jin et al. used idealized spatial patterns). However, similar results are obtained in all cases.

The effects of realistic noise applied to an HCM or ICM in a regime that would otherwise be periodic are sufficient to produce irregularity generally consistent with observed ENSO signals. In power spectra the main spectral peak is broadened and rises modestly above a noise background. Thus weather noise appears to be a very viable explanation for ENSO irregularity. The spectral signature in the presence of noise typically differs from models with irregularity induced purely by chaos, which tend to retain fairly sharp dominant spectral peaks.
Blanke et al. [1997], Chang et al. [1996] and Jin et al. [1996] further examined cases where the model ENSO cycle was stable in absence of noise. All found that the stochastic forcing was able to produce a spectral peak with period spatial patterns consistent with the ENSO mode that goes unstable at higher coupling, as expected for a weakly damped oscillation in presence of noise. Blanke et al. found that for realistic wind noise the expected peak and variance were approximately consistent with that observed and thus that the hypothesis of a stable ENSO cycle maintained by noise is plausible. Chang et al. used more sophisticated time series techniques to draw similar conclusions from an ICM. Jin et al. noted that the stable, stochastically maintained case does not possess a secondary quasi-biennial spectral peak and argued that this appears to favor a self-maintaining ENSO cycle interacting nonlinearly with the seasonal cycle by mechanisms similar to the chaotic case but with noise producing the irregularity. Grieser and Latif [1994], fitting a low-order model to ENSO time series, likewise favored a finite amplitude cycle.

If this role of weather noise is correct (promoting irregularity of the cycle, but not essential to its existence) then it resolves a longstanding divide between the coupled modeling community and the group advocating the importance of westerly wind bursts [see, e.g., McPhaden et al., 1992; Kindle and Phoebus, 1995, and references therein]. The westerly wind bursts would become one contributor, perhaps an important one, to the atmospheric noise. Kleeman and Moore [1997] discuss this possibility in terms of singular vectors and optimal perturbations.

6.4. ENSO Predictability

Prediction and predictability is reviewed at length in work by Latif et al. [this issue] and previous reviews on ENSO prediction may be found in work by Barnett et al. [1988], Latif et al. [1994], and Battisti and Sarachik [1995]. We thus confine our remarks here to the implications of theory for the limits of ENSO predictability, a subject of ongoing debate. Both of the hypotheses for ENSO irregularity discussed in the previous section (chaotic ENSO dynamics and weather noise) imply fundamental limits to predictability. The timescale of these limits is as yet poorly determined, although the estimated effects of noise in work by Kleeman and Power [1994], Flügel and Chang [1996], Latif and Eckert [1997], Blanke et al. [1997], and Kleeman and Moore [1997] suggest that weather noise has a substantial effect within a year, certainly within the first half cycle. Since better initialization of the system cannot improve the forecast against effects of such short decorrelation time noise, this affects our estimate of what we should expect from improved data assimilation systems. Flügel and Chang found that even growth associated with chaos was too slow to be important compared to noise effects in an ICM.

Another theoretical approach examines error growth due to algebraic growth of nonnormal modes, introduced in an ENSO context by Blumenthal [1991]. In a non-self-adjoint stable linear system an initial error in phase space, an “error ball,” contracts with time. Rapid contraction of some directions can permit other directions to grow for a finite time before eventually decaying. This finite time growth of an initial error can potentially contribute more to forecast error than slow exponential growth of an unstable mode. This may explain the faster of the two timescales of error growth noted by Goswami and Shukla [1991] in model-model experiments with the Cane and Zebiak [1985] model (although estimating predictability by these linear methods depends on initial error being small). Xue et al. [1994], Chen et al. [1997], and Moore and Kleeman [1996, 1997a, b] elaborated on similar mechanisms by singular vector analysis. The non-self-adjointness of the system is associated with the mergers of modes seen in mixed SST–ocean dynamics modes. Error growth associated with nonnormal mode growth can be improved by better data assimilation and more accurate observations.

7. Summary
7.1. What is Understood

Since early in the TOGA decade, modeling and theoretical evidence have been amassing that ENSO variability can arise through ocean-atmosphere interaction within the tropical Pacific basin. This revived and gave flesh to the older hypothesis by Bjerknes [1969] that ENSO arises by interaction between the trade winds and the ocean dynamics maintaining the cold tongue. The turnabout between warm and cold phases of the ENSO cycle is now thought to occur by subsurface ocean adjustment, which we refer to as the subsurface memory paradigm. A simple model articulated during mid-TOGA, the SSBH delayed oscillator model, provided an example of how ocean adjustment processes of a single Rossby wave, reflected at the western boundary as a Kelvin wave, coupled to a localized point wind stress anomaly responding to a point SST anomaly, could produce cycle behavior. Evidence that ENSO is inherently a cyclic phenomenon comes observationally from a broad spectral peak in ENSO time series and from a consensus of underlying cyclic behavior in ENSO models. In most models the cycle arises by instability of a cyclic mode which equilibrates to a nonlinear cycle. Multiple equilibria that give warm states and cold states in ENSO models by shutting down or increasing the cold tongue have been shown to be spurious by-products of flux correction. The nonlinear cycle in ENSO models is only weakly nonlinear, in the sense that the spatial structure of the variability and the dominant timescale are determined by the leading coupled mode of the system linearized about the climatology.
Current success in modeling and predicting ENSO with intermediate coupled models provides strong evidence of this weak nonlinearity. This, no doubt, contributed to the success of the TOGA program.

Although the inherent ENSO period can be understood from the linear problem, the processes setting the timescale are not simple. The ocean dynamics in the low-frequency ENSO coupled mode are not closely related to any mode of the uncoupled ocean. Because the frequency is much lower than the basin crossing time for Kelvin waves, the best prototype for ENSO mode behavior from uncoupled oceanic dynamics appears to be the periodically wind forced case first examined by Cane and Sarachik [1981] and Philander and Pacanowski [1981]. At these frequencies the ocean is nearly in balance with the wind stress along the equator, but the region off the equator, especially west of the wind stress region, has slower adjustment times and so provides a source of memory to the system, creating a small but persistent tendency in equatorial thermocline depth that can carry the oscillation between phases. The period reflects a competition between this memory term and feedbacks involving the part of the thermocline that is in steady balance with the wind stress. The period in ENSO models thus tends to depend considerably on parameters.

Some ENSO models possess chaotic regimes arising through interaction of the slow components of the ocean-atmosphere system with the seasonal cycle. Other scenarios for ENSO chaos have also been noted. Chaotic regimes do not occur everywhere in the realistic range of parameter space. Stochastic forcing by uncoupled atmospheric variability (weather noise) disrupting an ENSO cycle that would otherwise be frequency locked to the annual cycle appears to be an equally viable explanation for ENSO irregularity. The two explanations share some features and are not mutually exclusive.

7.2. What Lies Ahead

Although these areas are still subject to debate and investigation, here are some guesses on the basis of what is known at the end of TOGA. It is hoped that a simple consensus model will soon emerge that provides quantitative insight into the dominant ENSO period and that excludes the spurious nonoscillatory instabilities associated with flux correction found in current models. Such a model must capture the importance of subsurface memory, as did the simple model that so aided progress in mid-TOGA, the delayed oscillator model, hopefully in a simpler, algebraic system. It would be useful to retain in a simple model more accurate spatial structure than in point-coupling models, allowing it to meet some of the observational challenges such as those posed by Picaut et al. [1996]. A more quantitative explanation of ENSO period that includes detailed balance between the several mechanisms that interplay in ENSO, including the recent evidence for importance of zonal advection, would be beneficial. The question of whether the ENSO mode is unstable, leading to a self-sustaining cycle, or stable, sustained by noise, should be better substantiated. In the former case the modeling community must elucidate the energy source, currently a glaring lack in ENSO theory.

The role of weather noise should be refined, building bridges between the observational community that works in terms of specific events and the theoretical community that thinks in terms of ensemble properties. The mechanisms of seasonal phase locking of ENSO (or lack thereof) are in the process of being clarified by current debates. We anticipate considerable progress in bringing together theory for ENSO variability and ocean-atmosphere interaction within the tropical climate.

In atmospheric models, gaps that are in the process of being filled include: building intermediate atmospheric models that treat tropical moist convection, radiation, and clouds better than the simple models, while remaining less complex than AGCMs, and attempting to resolve disagreement among simple atmospheric models [e.g., Seager and Zebiak, 1994, 1995; Yu and Neelin, 1997].

Areas of tropical ocean-atmosphere interaction that have been initiated during TOGA but as yet lack a clear theoretical perspective include the following: interdecadal variability of ENSO [e.g., Wang and Rogelewski, 1995; Gu and Philander, 1995; Wang, 1995; Wang and Wang, 1996; Brassington, 1997]; ENSO interaction with the monsoon system [see Webster et al., this issue and references therein], interannual variability in other basins [e.g., Lamb et al., 1986; Philander, 1986; Carton and Huang, 1994; Chang et al., 1997; Zebiak, 1993], basin-basin interaction [e.g., Delecluse et al., 1994; Latif and Barnett, 1995], ocean-atmosphere interaction in the tropical climatology [e.g., Xie and Philander, 1994; Xie, 1994a, b; Dijkstra and Neelin, 1996; Xie, 1996; Sun and Liu, 1996; Jin, 1996; Liu, 1997; Li, 1997], and seasonal cycle [e.g., Chang and Philander, 1994; Liu and Xie, 1994; Xie, 1994a; Sun et al., 1995; Nigam and Chao, 1996; Chang, 1996], ocean-atmosphere interaction in tropical-subtropical interactions [e.g., Gu and Philander, 1997], and ocean-atmosphere-land interaction. If the next 10 years see as much progress in these areas as has occurred in ENSO theory over the TOGA decade, they will be fruitful ones indeed.

Acknowledgments. Preparation of this review was supported in part by National Science Foundation grant ATM-9521389 and National Oceanic and Atmospheric Administration grant NA46GP0244. The lead author thanks William Weibel and Johnny Lin for computations and graphics (Figures 2-5, and Plate 1; and Figure 9 respectively). This is UCLA-IGPP Contribution Number 5068.

References


Delcroix, T., J.-P. Boulanger, F. Masia, and C. Menkes, Geosat-derived sea level and surface current anomalies in


Kindle, J. C., and P. A. Phoebus, The ocean response to


Moore, D. W., and S. G. H. Philander, Modeling of the tropical oceanic circulation, in Marine Modelling, The Sea,


D. Battisti, Department of Atmospheric Sciences, AK-40 University of Washington, Seattle, WA 98195. (e-mail: david@atmos.washington.edu)
A. C. Hirst, CSIRO Division of Atmospheric Research, Private Bag 1, Mordialloc, Victoria 3195, Australia. (e-mail: ach@atmos.dar.CSIRO.au)

F.-F. Jin, Department of Meteorology, University of Hawaii at Manoa, 2525 Correa Road, Honolulu, HA 96822. (e-mail: jff@soest.hawaii.edu)
J. D. Neelin, Department of Atmospheric Sciences, Post Office Box 951565, University of California, Los Angeles Los Angeles, CA 90095-1565. (e-mail: neelin@atmos.ucla.edu)
Y. Wakata, Department of Earth and Planetary Physics, University of Tokyo, 2-11-16 Yayoi, Bunkyo-ky, Tokyo 113 Japan. (e-mail: wakata@geoph.s.u.tokyo.ac.jp)
T. Yamagata, Department of Earth and Planetary Physics, University of Tokyo, 2-11-16 Yayoi, Bunkyo-ky, Tokyo 113 Japan. (e-mail: yamagata@geoph.s.u.tokyo.ac.jp)
S. E. Zebiak, Lamont Doherty Earth Observatory Columbia University, Palisades, NY 10964. (e-mail: steve@ldeo.columbia.edu)

(Received July 29, 1996; revised November 17, 1997; accepted November 21, 1997.)