recent assessment is published as Report No. 44 in the Global Ozone Research and Monitoring series of WMO, Geneva, Switzerland (1998).

32. See (*15*) and references (*11*, *12*, *14*, *15*, *17*, *19*) in that review for the thermohaline circulation hysteresis.
33. E. Roeckner, L. Bengtsson, H. Feichter, J. Lelieveld, R. Rodhe, *J. Clim.*, in press.
34. The original publication in German [H. Grassl, *Strahlung in getrübten Atmosphären und in Wolken, Hamburger Geophysikalische Einzelschriften, Serie B*, No. 74 (1978)] has in part been repeated in *The Changing Atmosphere*, F. S. Rowland and I. S. A. Isaksen, Eds. (Chichester, London, 1988), chap. 11.
35. WCRP, the International Geosphere/Biosphere Programme (IGBP), and the International Human Dimensions Programme (IHDP) on Global Environmental Change already jointly sponsor the Global Change System for Analysis, Research and Training, which establishes research networks in large regions (such as Southeast Asia, northern Africa, and southern Africa) and promotes projects such as Climate Predictions for Agriculture. However, the infrastructure (an international secretariat and international project offices) is weak for IHDP, and IGBP is also weak in terms of project offices. A fourth program, called DIVERSITAS and devoted to biodiversity, does not yet have what one can call a functioning infrastructure.
36. Global satellite data sets are especially useful for establishing parameters for which no in situ equivalent exists, for example, monthly mean precipitation (also over oceans) as derived by the global Precipitation Climatology Project of the Global Energy and Water Cycle Experiment [G. Huffman *et al.*, *Bull. Am. Meteorol. Soc.* **78**, 5 (1997)]. Ocean surface heat fluxes also belong to this category.
37. The theoretical framework for the optimal fingerprint method was laid by K. Hasselmann, *J. Clim.* **6**, 1957 (1993); K. Hasselmann, *Clim. Dyn.* **13**, 601 (1997); and K. Hasselmann, *Quart. J. R. Meteorol. Soc.* **124**, 2541 (1998).
38. L. Bengtsson, E. Roeckner, M. Stendel, *J. Geophys. Res.* **104**, 3864 (1999).

REVIEW

Is El Niño Changing?

Alexey V. Fedorov and S. George Philander

Recent advances in observational and theoretical studies of El Niño have shed light on controversies concerning the possible effect of global warming on this phenomenon over the past few decades and in the future. El Niño is now understood to be one phase of a natural mode of oscillation—La Niña is the complementary phase—that results from unstable interactions between the tropical Pacific Ocean and the atmosphere. Random disturbances maintain this neutrally stable mode, whose properties depend on the background (time-averaged) climate state. Apparent changes in the properties of El Niño could reflect the importance of random disturbances, but they could also be a consequence of decadal variations of the background state. The possibility that global warming is affecting those variations cannot be excluded.

The two most intense El Niño episodes in more than a century occurred during the past two decades, in 1982 and 1997. Whether these exceptional warmings of the eastern tropical Pacific Ocean and the associated changes in global weather patterns were manifestations of global warming and how the continual rise in the atmospheric concentration of greenhouse gases will affect El Niño in the future are issues currently being debated. The disagreements mainly concern the causes of the irregularities in the continual climate fluctuation, the Southern Oscillation, between complementary El Niño and La Niña states. This natural mode of oscillation, attributable to ocean-atmosphere interactions in which the winds create sea surface temperature gradients that in turn reinforce the winds, plus negative feedbacks involving the dynamical response of the oceans to changes in the winds, is neutrally stable, so that random disturbances contribute to its irregularities. Other causes for variations in the properties of this mode (and more generally of a spectrum of possible modes including those involved in the seasonal cycle) include changes in the background climate state, which is described by factors such as the intensity of the time-averaged trade winds, τ, and the spatially averaged depth of the thermocline, H. Changes in that state can explain why El Niño has different properties in paleorecords from different times and why it appears to be changing gradually in response to the decadal fluctuation that modulates the background state in records for the past century. That fluctuation, which brought relatively weak trade winds and unusually warm surface waters to the eastern equatorial Pacific in the late 1970s, is of uncertain origin, but it could be under the influence of global warming. Different climate models differ in their assessment of how that warming will affect El Niño because they reproduce different background states for the future.

Atmospheric Aspects

The Southern Oscillation, the dominant signal in Fig. 1A, shows sea surface temperature variations as measured on the equator to the west of the Galapagos Islands. (The seasonal cycle and higher frequency variations are filtered out.) Figure 1, B and C, depicts conditions at the peaks of particularly intense El Niño and La Niña episodes. Such changes in sea surface temperature have a profound effect on climate throughout the tropics because, in low latitudes, the correlation between sea surface temperature and rainfall patterns is almost perfect: Moist air rises spontaneously into cumulus towers over the warmest regions, which therefore have plentiful rainfall; aloft, the air that has been drained of its moisture diverges from these regions and subsides over the colder regions that get little precipitation. Surface winds, the trades in the case of the Pacific, restore moisture to the air by means of evaporation while returning it to the warmest regions. These direct thermal circulations are controlled by surface temperatures, so changes such as those in Fig. 1 substantially alter rainfall, winds, and other atmospheric variables.

During La Niña, the trade winds are intense, and heavy rains fall mainly over the far western tropical Pacific; during El Niño, the winds relax and the heavy rains move eastward, so that the coastal zones of Ecuador and Peru have severe floods, whereas New Guinea and Indonesia experience relatively dry conditions. The expanse of warm waters in the Pacific during El Niño is so vast and causes such a huge increase in evaporation from the ocean (and hence in the release of latent heat in the atmosphere when the water vapor condenses to form clouds) that weather patterns are affected globally. Numerical models of the atmosphere that are used to predict the weather (those with forecast skills that are limited to a few days at most) are capable of reproducing realistically and deterministically the atmospheric aspects of the Southern Oscillation over extended periods of several decades, provided that the observed sea surface temperature variations of the tropical Pacific are specified as boundary conditions. (Calculations in which the boundary conditions correspond to the climatological seasonal cycle fail to simulate the Southern Oscillation.) This means that it is possible to predict certain time-averaged atmospheric conditions indefinitely into the future, provided that we know how sea surface temperatures will vary (*1*, *2*). From an atmospheric perspective, the problem appears to be oceanographic.

Department of Geosciences, Princeton University, Princeton, NJ 08544, USA.

Oceanic Aspects

Heat fluxes across the ocean surface determine sea surface temperatures over much of the globe. The tropics, however, are different because there the dynamical response of the oceans to the winds is of primary importance. This is because the warm surface waters of the tropics constitute only a very shallow layer floating on the cold water below. The winds, by causing variations in the depth of the thermocline, can expose cold water to the surface. For example, intense westward trade winds along the equator during La Niña drive the warm waters westward while bringing cold water to the surface in the east, thus contributing to the temperature pattern shown in Fig. 1C. The pattern in Fig. 1B appears when a relaxation of those winds during El Niño allows the warm water to flow back eastward. The winds—intense during La Niña, relaxed during El Niño—cause a continual horizontal redistribution of the warm surface waters, which results in the temperature fluctuations seen in Fig. 1. This redistribution is effected by changes in the complex system of tropical currents and undercurrents—some are westward and some are eastward—and involves waves of various types that propagate back and forth across the ocean basin. The responses of the Indian Ocean to the abrupt changes in monsoonal winds, the Atlantic to the gradual seasonal fluctuations of the trades, and the Pacific to the wind variations associated with El Niño and La Niña provide a wealth of information about the oceanic response to different wind-stress patterns. The instruments developed for measurements of the surface and subsurface temperature and current fluctuations associated with the various phenomena made it possible by the early 1990s to continually monitor the tropical Pacific Ocean with an array of instruments (shown in Fig. 2) (*3*–*5*). The measurements are relayed by satellite to certain locations in the United States. The information was of crucial importance for the prediction of the El Niño of 1997.

The availability of detailed measurements of the different oceanic phenomena in the tropics led to the development of a hierarchy of models (*6*). Some are highly idealized and serve to study isolated aspects of complex reality, for example, the spectrum of waves that travel along the equator. The most sophisticated models—the oceanic counterparts of the atmospheric models used for weather forecasts—are capable of realistic simulations of the tropical Pacific over many decades, provided that the surface winds are specified. These models are now used to interpolate the measurements from the array of instruments shown in Fig. 2, thus providing very detailed descriptions of conditions in the Pacific (*7*). That information is needed to initialize coupled ocean-atmosphere models for the prediction of climate variability on seasonal and interannual time scales, including the prediction of El Niño.

Interactions Between the Ocean and Atmosphere

The explanation for El Niño and La Niña involves a circular argument: Changes in sea surface temperature are both the cause and consequence of wind fluctuations. It follows that interactions between the ocean and atmosphere can amount to positive feedbacks. Consider, for example, conditions during La Niña, when intense trade winds keep the warm surface waters along the equator in the far west, thus maintaining a zonal temperature gradient that contributes to the intense winds. A modest disturbance in the form of a brief burst of westerly winds near the date line—winds that are common in the western equatorial Pacific—will generate currents that transport some of the warm water eastward, thus decreasing the zonal temperature gradient. The resultant weakening of the trade winds will cause more warm water to flow eastward, causing even weaker winds. That is how El Niño can develop from a modest initial disturbance. Such positive feedbacks between the ocean and atmosphere occurred in the equatorial Pacific in early 1997 when the eastward progression of warm waters was accompanied by the simultaneous weakening of the easterly trades (*8*). (The oceanic aspects of these developments depend on the wind forcing and differ from Kelvin waves that are unforced modes of oscillation of the ocean.)

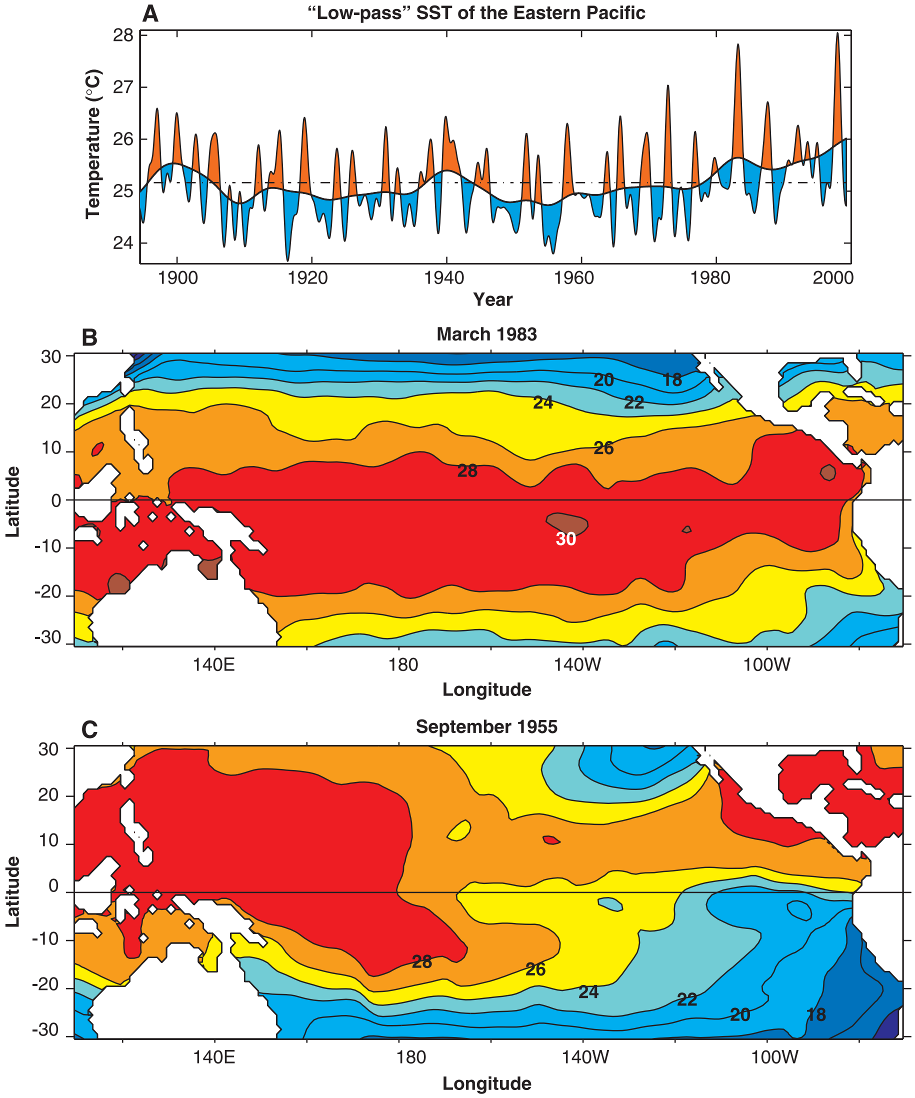

Fig. 1. (**A**) The interannual oscillations in sea surface temperature (SST) at the equator in the eastern Pacific (averaged over the area 5°S to 5°N, 80° to 120°W) shown on the background of the decadal fluctuation (obtained by means of a low-pass filter) after removal of the annual cycle and higher frequency variability. The horizontal dot-dashed line is the time average for the record. Sea surface temperatures (°C) at the peaks of (**B**) El Niño in March 1983 and (**C**) La Niña in September 1955. Some of the differences between (B) and (C) (high temperatures that extend far north in September and far south in March) are attributable to the seasonal cycle.

Westerly wind bursts may have been a factor in the development of El Niño in 1997, but there have been occasions when the bursts failed to produce El Niño or when the bursts were absent and El Niño nonetheless appeared. Furthermore, these irregular, sporadic wind fluctuations can account for neither the dominant period nor the spatial structure of the continual oscillation shown in Fig. 1. That oscillation is a natural mode of the coupled ocean-atmosphere system, and determination of its properties requires stability analyses of the interactions between the two media. A powerful tool for such studies is a model (*9*) that deals only with departures (in the winds, temperatures, and so on) from a specified background state described by parameters such as H and τ (which were defined earlier). A convenient classification of the various modes of oscillation is in terms of the processes that determine sea surface temperatures, the only oceanic parameter to affect the atmosphere on the time scales of interest here [see (*10*) and (*11*) for an alternative classification in terms of nondimensional parameters and for a comprehensive summary of research on this topic]. The following is a brief discussion of two important idealized modes with features evident in the observations.

A thermocline at such a great depth that winds are unable to bring cold, deep water to the surface precludes interactions between the ocean and atmosphere unless the thermocline shoals, or the time-averaged westward winds intensify, thus elevating the thermocline in the east while deepening it in the west by driving surface waters westward. The second of these two possibilities favors a type of mode shown schematically in Fig. 3A. Sea surface temperature variations occur mainly in the east, in response to vertical movements of the thermocline induced by wind fluctuations in the west. (The thermocline is so deep in the west that its vertical movements leave local sea surface temperatures unaffected.) At the long periods of these modes—3 years and more—the winds excite a continuum of waves (rather than a few isolated ones) that brings a narrow equatorial zone into equilibrium with the winds while, off the equator, the ocean lags behind the winds. This delayed response of the ocean ensures a continual oscillation. The off-equatorial elevations of the thermocline in the west (Fig. 3A) (their presence is attributable to the winds that transfer warm water from those regions to the east during El Niño) disperse into waves that ultimately reach the east, where they elevate the thermocline and initiate a change from El Niño to La Niña. The time dependence of this mode, which can be associated with eastward phase propagation, is captured by the equation

$$T_t = aT + bT(t - d)$$

The first two terms represent the positive feedbacks between the wind and the ocean that amplify temperature disturbances, T, with time, t. The third term represents the delayed response of the ocean. Its presence gives rise to oscillations for a certain range of values for the constants a, b, and d (*12, 13*). The crucial role of the delayed response of the ocean (future developments can be anticipated by observing the ocean, especially movements of the thermocline) is the reason for the importance of the array in Fig. 2. The time scale of the Southern Oscillation is determined primarily by oceanic processes.

The mode in Fig. 3B, unlike that of Fig. 3A, involves no vertical movements of the thermocline but depends on the entrainment of cold water across a shallow thermocline. The winds that converge onto a warm disturbance can cause cooling on its eastern side (by means of westward advection and upwelling of cold water) and warming on the western side. The resultant mode, which drifts westward, is associated with relatively short periods of a year or two and, along with a mode that is antisymmetrical about the equator (*14*), influences the response of the equatorial ocean to seasonal forcing (*5, 15*).

The coupled ocean-atmosphere modes that are possible in reality are hybrids of those in Fig. 3 and involve sea surface temperature variations induced by vertical movements of the thermocline and by the entrainment of water across the thermocline. Horizontal advection can also substantially influence sea surface temperatures (*16*), broadening the spectrum of modes even further. Figure 4 shows how the properties of the most unstable mode depend on the background state,

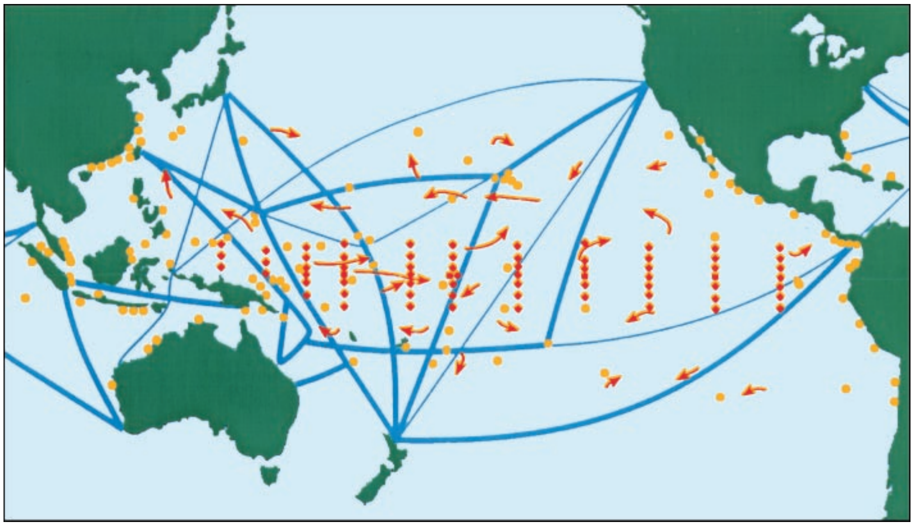

Fig. 2. The array of instruments that monitor oceanic conditions. The blue lines indicate the tracks of commercial ships that deploy instruments that measure temperature to a depth of a few hundred meters. The arrows show drifting buoys that measure the temperature and the wind, whose movements, tracked by satellites, yield information about surface currents. The yellow dots represent tide gages that measure sea level depending on the temperature of the water column. The red diamonds and squares (buoys moored to the ocean floor) show locations where temperature and currents, respectively, are measured over the upper few hundred meters of the ocean. (The data are available at www.pmel.noaa.gov/toga-tao/realtime.html.)

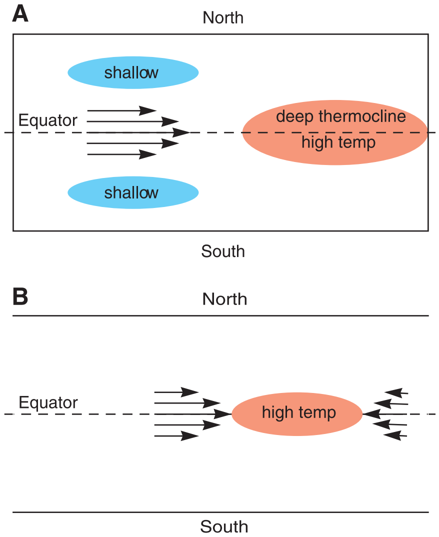

Fig. 3. A schematic diagram of the spatial structure of two idealized modes. Arrows indicate winds, shaded areas show changes in thermocline depth, and "temp" refers to surface temperature. These conditions, during El Niño, correspond to departures from a background state. Mode (**A**), known as the delayed oscillator, is prominent when the background state has a deep thermocline and intense westward winds. Mode (**B**), which is prominent when the background state has a shallow thermocline, has a relatively short period of a year or two and has westward phase propagation. The structure of the Southern Oscillation in reality is a hybrid of these two.

described by the values of the parameters H and τ (*17*). (The equatorial temperature gradient of the background state is shown in Fig. 4C as the value of the surface temperatures in the east; the value in the west is assumed to remain constant.) These results confirm the previous statements concerning the stabilizing effects of increasing H or weakening τ. Changes in the values of those parameters alter the structure of the modes. Each point on Fig. 4 is associated with a distinct spatial structure that is a hybrid of the idealized types in Fig. 3. Near point E, where periods are short and where entrainment across the thermocline has a strong influence on surface temperatures, the structure tends toward that in Fig. 3B. Near point D, where the period is long, vertical movements of the thermocline are mainly responsible for surface temperature changes, and the structure resembles that in Fig. 3A.

The best check for results such as those shown in Fig. 4 are measurements that describe the properties of the Southern Oscillation under a variety of background climatic conditions. Paleorecords from sources such as lake deposits and corals are beginning to provide partial information of the required type, but accurate data that describe El Niño and the background climate during a certain period in the past are lacking. However, the results in Fig. 4 can assist with the interpretation of paleorecords. For example, deposits in a lake in southwestern Ecuador indicate that, whereas El Niño today occurs every 3 to 5 years, the interval between successive events was far longer, on the order of a decade, some 7000 years ago (*18*). In Fig. 4, such a change requires a move in the direction of point D—a background state with intense winds, a deep thermocline, and low sea surface temperatures in the east. This result could help resolve the debate about the climate of the eastern equatorial Pacific during the early Holocene.

Present-day estimates of the values of H, τ, and other parameters indicate that current conditions in the tropical Pacific Ocean correspond to the general area of points A and B in Fig. 4, where interactions between the ocean and atmosphere are close to neutral stability. Random disturbances are therefore responsible for sustaining the Southern Oscillation and for contributing to its irregularity. Exactly how close the ocean-atmosphere interactions are to neutral stability is a matter of debate. Some investigators argue that the interactions are sufficiently unstable for the irregularities to be the result mainly of nonlinearities (*19*). Other investigators (*20, 21*) assume that the interactions are strongly damped and explore the nonnormal modes of the coupled ocean-atmosphere system; given an initial random disturbance with the appropriate spatial structure, those modes amplify more rapidly than the normal modes. Those who claim that westerly wind bursts are essential "triggers" of El Niño implicitly assume that the system is damped and that random triggers at different times result in the impression of a continual oscillation. In reality, the background state is probably changing gradually and continually, causing the interactions to be strongly damped on some occasions, unstable on other occasions, and neutral for much of the time. This possibility of a changing background state (*22*) has recently generated much debate in connection with the possible effect, on El Niño, of future global warming.

The Modulation of the Southern Oscillation

In Fig. 1A, the properties of the Southern Oscillation appear to have changed during the 1980s and 1990s: La Niña episodes were very weak or practically absent during those decades, whereas El Niño attained unprecedented amplitudes in 1982 and 1997 and was unusually prolonged in 1992 (if the horizontal dot-dashed line is used as the reference). To some investigators (*23*), this change is attributable to global warming. Others (*24–26*) interpret the record without invoking any changes in the properties of El Niño. They interpret the fluctuations in Fig. 1A as the random fluctuations of a stationary time series, so that the Southern Oscillation is an unchanging, weakly damped ocean-atmosphere mode made irregular by random atmospheric disturbances, which also contribute to the decadal fluctuation (*27*) (the smooth bold line in Fig. 1A). This argument, which is strictly about statistical matters, does not preclude the possibility that the properties of the mode are changing gradually, but finds that the available time series are too short to confirm such a conclusion. An alternative approach, based on the results in Fig. 4, explores how the decadal fluctuation, which is assumed to be distinct from the interannual variability, can modulate that variability by gradually changing the val-

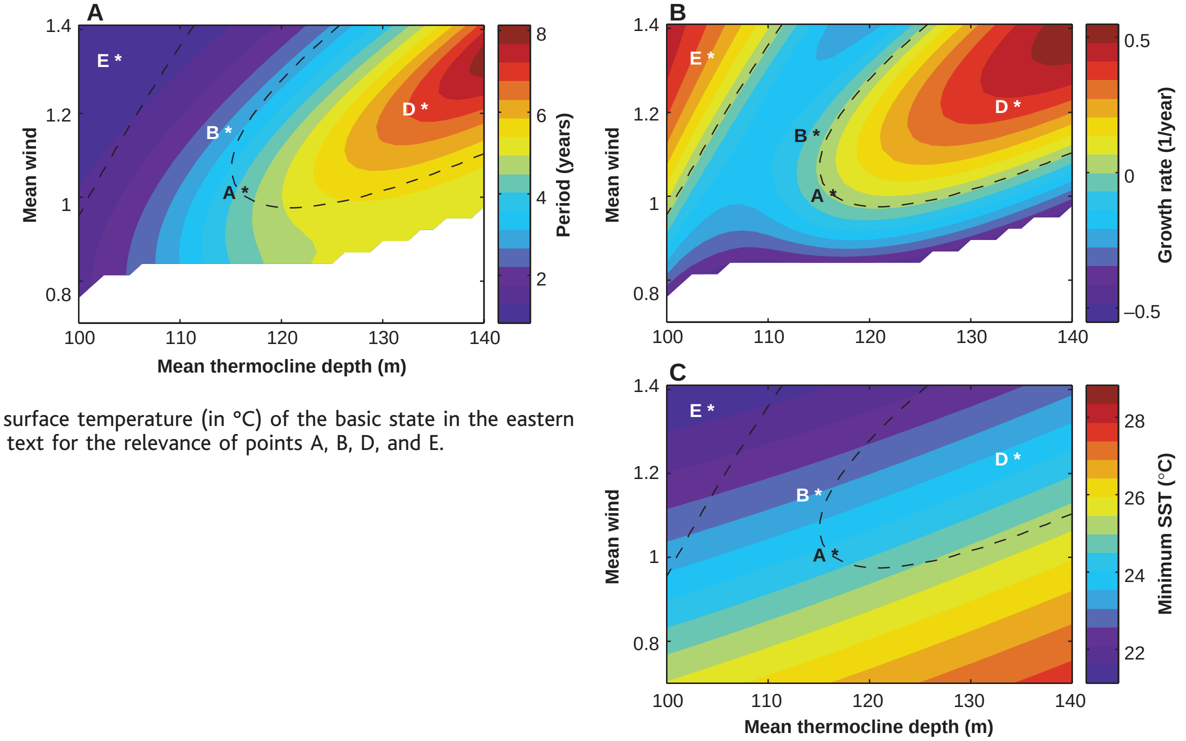

Fig. 4. (**A**) The period (in years) and (**B**) growth rates (in 1/years) of the most unstable oscillations as a function of thermocline depth (in meters) along the horizontal axis and the intensity of easterly equatorial winds (in units of 0.5 cm^2/s^2). Dashed lines indicate zero growth rate or neutral stability; modes with a coherent structure are absent from the stable white area. (**C**) The minimum sea surface temperature (in °C) of the basic state in the eastern Pacific Ocean. See the text for the relevance of points A, B, D, and E.

ues of the parameters H and τ. In this approach, the reference line for interpreting the interannual oscillation is no longer the time-averaged temperature of the past century (the horizontal dot-dashed line in Fig. 1A) but is the slowly undulating decadal fluctuation. With this choice, La Niña is present throughout the record, and the warm conditions that start in 1992, rather than an exceptionally prolonged El Niño, amount to the persistence of background conditions. Those conditions are in part responsible for the exceptionally high surface temperatures of El Niño in 1982 and 1997. The debate about changes in El Niño therefore becomes a debate about the reality of the decadal fluctuation. Its cause, a matter currently under investigation, probably involves the processes that maintain the thermocline, for example, exchanges between the tropical and extratropical oceans (*28, 29*). The effect of the fluctuation on El Niño can be inferred from Fig. 4.

A weakening of the easterly winds and a deepening of the thermocline accompanied the relatively warm conditions in the eastern tropical Pacific during the 1980s and 1990s (*30*), causing the values of H and τ to change from the area of point B in Fig. 4 during the 1960s and 1970s to the vicinity of point A in the 1980s and 1990s. Such a move should have increased the period of the oscillation (from ~3 years to 5 years) and should have altered the structure of the mode, moving it closer toward that in Fig. 3A. Spectral analysis suggests that such an increase in period occurred, although the statistical significance is low because of the brevity of the time series. Furthermore, descriptions of El Niño based on data from the 1960s and 1970s (*31*) emphasize westward phase propagation, a property of the mode in Fig. 3B; subsequently, eastward phase propagation, a feature of the mode in Fig. 3A, has been more common, starting with El Niño of 1982, which was surprising because its development started in the west rather than in the east.

A regular Southern Oscillation whose properties are modulated by a gradually changing background state is an idealization that filters out several important observed features. In reality, the presence of random disturbances that can accelerate (or retard) developments makes the Southern Oscillation irregular and alters its structure temporarily by promoting processes not favored by the background state. (For example, wind bursts can cause advection, say, to have a strong influence on sea surface temperature.) Another complication is the nonsinusoidal nature of the Southern Oscillation. In reality, La Niña is not simply the negative of El Niño. The intense El Niño events of 1982 and 1997 were brief and lasted on the order of a year, after which cold La Niña conditions persisted for several years. In both cases, El Niño started as an eastward surge of warm waters, associated with a deepening of the thermocline driven by westerly winds that progressed eastward. The return of cold surface waters to the eastern Pacific started with an eastward traveling elevation of the thermocline along the equator, but then involved additional processes different from those that brought warm water to the east. These asymmetrical aspects of the Southern Oscillation are as yet neither explained nor simulated by any coupled ocean-atmosphere model.

The Background State

Global warming is bound to affect El Niño by altering the background climate. An investigation of this matter requires consideration of the climate of today, especially the intriguing asymmetries of the tropical Pacific evident in Fig. 1C: Water is colder in the east than in the west along the equator, and in the east, water is warmest to the north of the equator even though the distribution of sunlight is symmetrical about the equator and is independent of longitude. Time-averaged cloudiness, rainfall, and other variables all have these asymmetries, which depend on continental geometries and on a variety of feedbacks, some similar to those involved in El Niño. The east-west asymmetry is attributable to westward winds that depend on global factors and, in addition, on zonal temperature gradients that in turn depend on the winds. The north-south asymmetry is confined to the east because the shallow thermocline there permits northward cross-equatorial winds to create a north-south temperature gradient that in turn maintains the winds. An additional feedback involves low-level stratus clouds that form over the cold waters south of the equator, thus reinforcing the low surface temperatures. The various processes that maintain the background have been studied by using highly idealized models that focus on a few feedbacks in isolation (*32*) and coupled general circulation models of the ocean and atmosphere (*33*). The development of the latter models, which should be capable of realistically reproducing the background state plus superimposed fluctuations, is a major challenge because the various feedbacks amplify not only random disturbances (that may in fact occur) but also errors in the models. In a strictly atmospheric model, specified sea surface temperatures provide a strong constraint. However, if that model is only a component of a coupled ocean-atmosphere model, then that constraint is absent and, because of the feedbacks, the results are very sensitive to the treatment of complex phenomena and processes such as convection, clouds, radiation, and turbulent mixing. Inadequate treatment of those processes causes some models to have weak winds and an absence of cold equatorial waters, whereas in other models the winds are too intense and the waters are too cold. Some models have cold water along the equator but have unrealistic bands of warm water to the north and south. If a model has an unrealistic background state, then its natural modes of oscillation are also unrealistic. The properties of the simulated oscillations (and the results in Fig. 4) can help identify the reasons for the behavior of a model. For example, in some models, the parameterization of convection causes sea surface temperature changes to induce changes in the winds very close to those temperature changes. Such models are biased toward modes of the type in Fig. 3B with relatively short periods. Models in which wind changes are far to the west of the temperature changes are biased toward modes with a long period, similar to those in Fig. 3A.

The development of coupled ocean-atmosphere models is progressing rapidly, and several of those models reproduce interannual fluctuations whose gross features resemble the observed Southern Oscillation (*34, 35*). The simulations nonetheless fail to be in the realistic neighborhood of points A and B in Fig. 4. For example, in a recent model with an oceanic component that has high spatial resolution, the Southern Oscillation has a period of 2 years (*36*). The simulated mode is in the vicinity of point E in Fig. 4, probably because of a thermocline that is too shallow. If simulations of the current background state are poor, then projections of how global warming will alter that background state, and hence influence El Niño, have enormous uncertainties. Several such projections have been made and yield results that differ from one another (*36–38*). At this time, it is impossible to decide which, if any, are correct.

References and Notes

1. J. M. Wallace *et al.*, *J. Geophys. Res.* **103**, 14241 (1998).
2. K. E. Trenberth *et al.*, *J. Geophys. Res.* **103**, 14291 (1998).
3. D. Halpern, *J. Geophys. Res.* **92**, 8197 (1987).
4. S. P. Hayes *et al.*, *Bull. Am. Meteorol. Soc.* **72**, 339 (1991)
5. M. J. McPhaden *et al.*, *J. Geophys. Res.* **103**, 14169 (1998).
6. T. N. Stockdale, A. J. Busalacchi, D. E. Harrison, R. Seager, *J. Geophys. Res.* **103**, 14325 (1998).
7. M. K. A. Ji and A. Leetmaa, *Bull. Am. Meteorol. Soc.* **75**, 569 (1995).
8. M. J. McPhaden, *Science* **283**, 950 (1999).
9. S. E. Zebiak and M. A. Cane, *Mon. Weather Rev.* **115**, 2262 (1987).
10. J. D. Neelin *et al.*, *J. Geophys. Res.* **103**, 14261 (1998).
11. F.-F. Jin and J. D. Neelin, *J. Atmos. Sci.* **50**, 3477 (1993).
12. P. S. Schopf and M. J. Suarez, *J. Atmos. Sci.* **45**, 549 (1988).
13. D. S. Battisti and A. C. Hirst, *J. Atmos. Sci.* **46**, 1687 (1989).
14. P. Chang and S. G. Philander, *J. Atmos. Sci.* **51**, 3627 (1994).
15. T. Li and S. G. Philander, *J. Clim.* **9**, 2986 (1996).
16. J. M. Picaut and T. Delcroix, *J. Geophys. Res.* **100**, 18393 (1995).
17. A. Fedorov and S. G. Philander, *Geophys. Res. Lett.*, in press.
18. D. T. Rodbell *et al.*, *Science* **283**, 516 (1999).
19. M. A. Cane, S. E. Zebiak, S. C. Dolan, *Nature* **321**, 827 (1986).
20. A. M. Moore and R. Kleeman, *J. Clim.* **12**, 1199 (1999).

21. C. Penland and P. D. Sardeshmukh, *J. Clim.* **8**, 1999 (1995).
22. B. P. Kirtman and P. S. Schopf, *J. Clim.* **11**, 28043 (1998).
23. K. E. Trenberth and T. J. Hoar, *Geophys. Res. Lett.* **24**, 3057 (1997).
24. D. E. Harrison and N. K. Larkin, *Geophys. Res. Lett.* **24**, 1775 (1997).
25. B. Rajagopalan, U. Lall, M. A. Cane, *J. Clim.* **10**, 2351 (1997).
26. C. Wunsch, *Bull. Am. Meteorol. Soc.* **80**, 245 (1999).
27. Y. Zhang, J. M. Wallace, D. S. Battisti, *J. Clim.* **10**, 1003 (1997).
28. Z. Liu, *J. Phys. Oceanogr.* **23**, 1153 (1993).
29. J. P. McCreary and P. Lu, *J. Phys. Oceanogr.* **24**, 466 (1994).
30. T. P. Guilderson and D. P. Schrag, *Science* **281**, 240 (1998).
31. E. M. Rasmusson and T. H. Carpenter, *Mon. Weather Rev.* **110**, 354 (1982).
32. H. A. Dijkstra and J. D. Neelin, *J. Clim.* **8**, 1343 (1995).
33. S. G. Philander *et al.*, *J. Clim.* **9**, 2958 (1996).
34. C. R. Mechoso *et al.*, *Mon. Weather Rev.* **123**, 2825 (1995).
35. P. Delecluse *et al.*, *J. Geophys. Res.* **103**, 14357 (1998).
36. A. Timmermann *et al.*, *Nature* **398**, 694 (1999).
37. S. Tett, *J. Clim.* **8**, 1473 (1995).
38. T. R. Knutson, S. Manabe, D. Gu, *J. Clim.* **10**, 138 (1997).
39. This work was supported by the National Oceanic and Atmospheric Administration (grant NA56GP0226).

REVIEW

Mantle Convection and Plate Tectonics: Toward an Integrated Physical and Chemical Theory

Paul J. Tackley

Plate tectonics and convection of the solid, rocky mantle are responsible for transporting heat out of Earth. However, the physics of plate tectonics is poorly understood; other planets do not exhibit it. Recent seismic evidence for convection and mixing throughout the mantle seems at odds with the chemical composition of erupted magmas requiring the presence of several chemically distinct reservoirs within the mantle. There has been rapid progress on these two problems, with the emergence of the first self-consistent models of plate tectonics and mantle convection, along with new geochemical models that may be consistent with seismic and dynamical constraints on mantle structure.

Subsolidus (*1*) convection of the rocky, 2900-km-thick mantle of Earth (*2*) is the driving mechanism for plate tectonics and associated geological activity on the surface of our planet, including continental drift, earthquakes, volcanoes, and mountain building (*3*). Mantle convection and plate tectonics are one system, because oceanic plates are the cold upper thermal boundary layer of the convection. The slow motion of plates and the mantle is powered by radiogenic heating and by the slow cooling of our planet over its 4.5-billion-year history (*4*). Mantle convection and plate tectonics provide the central framework linking the subdisciplines of solid Earth science, including geochemistry, seismology, mineral physics, geodesy, tectonics, and geology. A successful model must thus satisfy constraints from all of these fields. For example, seismic waves provide a direct probe of structures inside Earth, chemical analyses of erupted lavas and other volcanic products give information about different compositions that exist in the mantle, and laboratory experiments determine the properties and deformation mechanisms of rocks at the high pressures (≤136 GPa) and temperatures (≤4000 K) of Earth's mantle.

Despite three decades of research, some first-order questions regarding the plate-mantle system remain largely unresolved. The first of these is why Earth developed plate tectonics at all, given that other terrestrial planets such as Mars and Venus do not currently exhibit this behavior (*3*). The problem is complex because rocks exhibit many different deformation mechanisms, ranging from brittle failure to viscous creep, depending on the pressure, temperature, differential stress, and past history of deformation (*5, 6*), and many of these mechanisms are not well quantified. The second major question is how the differing chemical compositions of erupted magmas, which require several chemically distinct reservoirs within the mantle (*7*), can be reconciled with other observational and dynamical constraints favoring whole-mantle convection (*8*), which should mix chemical heterogeneities.

There has been substantial recent progress toward resolving these two issues, such that a self-consistent physical and chemical model capable of integrating plate tectonics, geochemical observations, and other constraints may be possible within the next decade. This review focuses on these two issues. However, there has also been important progress in other areas, notably (i) understanding how the present state of the mantle, as imaged by seismic tomography, is related to the recent (~140 million year) history of plate motions and recent geological events (*9–11*); (ii) basic mantle dynamics, particularly the influence of non-Newtonian rheology (*12–14*); and (iii) plume dynamics, including formation from a hot thermal boundary layer (*15*) and interaction with plates (*16, 17*).

The Plate Problem

Plate tectonics does not arise from convection by viscous creep with temperature-dependent viscosity. Instead, a rigid lid is formed with convection taking place beneath it. The other terrestrial planets display a rigid lid, although plate tectonics may have existed in the past on Mars (*18*) or Venus (*19*), with Venus perhaps fluctuating between plate tectonics and a rigid lid (*20*). A rigid lid forms because the temperature drops from 1600 K (mantle) to 300 K (surface) over the upper thermal boundary layer, and the thermally activated processes responsible for mantle deformation (*2*) become many orders of magnitude slower as temperature decreases (*14*). Our understanding of rigid lid convection has been greatly refined by several recent studies (*13, 21, 22*).

Earth does not have a rigid lid. Instead, convection reaches the surface, and oceanic plates are the upper thermal boundary layer participating in the convective motion (*23*). Additional deformation mechanism(s) are responsible for breaking the stiff outer lid into strong plates separated by weak boundaries (*24, 25*). Field observations and laboratory experiments give guidance as to these mechanisms (*5, 6*). Faults form by brittle failure near the surface, but increasing pressure (hence friction) with depth limits faults to the upper ~15 km of the plates, which are some tens of kilometers thick. Below ~15 km, deformation is probably accommodated by distributed microcracks (*5*). Laboratory experiments indicate that the yield strength of rocks is hundreds of megapascals and is weakly dependent on pressure (*26*). Plastic flow, by lattice processes such as dislocation glide, occurs at higher pressures and stresses (*6*). Thermally activated viscous creep (*2, 14*) takes over at higher mantle temperatures and greater depths.

In addition to the above mechanisms,

Department of Earth and Space Sciences, University of California, Los Angeles, CA 90095, USA.