There are many phenomena of interest in the atmosphere and ocean, only some of which are relevant for seasonal forecasting. One way of identifying the processes likely to be active is through scale analysis which identifies the important terms in the governing equations and highlights the importance of geostrophic balance. Simple arguments for Rossby waves are given. These waves are important in both atmosphere and ocean as a means of transferring energy over large distances. When the waves are embedded in a westerly flow it is possible for the waves to be stationary, giving rise to the possibility of a coherent remote response. A possible source of stationary atmospheric Rossby waves could be the deep convection over parts of the equatorial oceans where the sea surface temperatures are high. These stationary wave trains may interact with mid-latitude phenomena such as the storm tracks, so changing the occurrence and preferred locations of storms. This is an example of interaction between weather and lower frequency climate changes. Other teleconnections are introduced, such as the link between the Indian summer monsoon and Mediterranean climate. The area of the world where the interaction between the atmosphere and ocean is strongest is in the tropics. It is important to understand how the upper equatorial ocean works and how it is connected to the subtropical thermocline. The connection of the tropics to the subtropics gives a possible mechanism for low frequency variability of ENSO. Various theories of ENSO are introduced in which the importance of equatorial Kelvin and Rossby/planetary waves is highlighted. Simple models illustrate oscillatory behaviour in certain parameter regimes but damped oscillations in others. While these ideas are interesting in generating a framework within which to consider ENSO, the real test comes from the making of forecasts and determining by experience the limits of predictability.
4.1 The Role of the Atmosphere

4.1.1 Scales

The atmosphere and ocean are shallow layers of fluid around the Earth acted upon by gravitational attraction to the almost spherical solid Earth. Using the Earth parameters $a$, the radius, and $\Omega$, the rotation rate, $N$ the basic buoyancy frequency associated with the stable stratification, and typical scales for the phenomenon of interest for horizontal and vertical length, $L$ and $H$, respectively, and horizontal velocity, $V$, we have the following scaling relations:

$$H << a \quad \text{(shallow fluid)} \quad \text{and} \quad H << L$$

so that hydrostatic balance is a good approximation in the vertical. Also, important velocity scales and typical numbers for them are:

$$\frac{a\Omega}{465} \sim \frac{(gH)^{1/2}}{300} > \frac{\left(NH\right)^{1/2}}{100} > V \quad 20 \quad \text{m.s}^{-1}$$

The first number is the speed at which a parcel on the equator moves purely due to the rotation of the Earth. The second, in which the density height scale of the atmosphere (about 10 km) has been used, is the speed of external gravity waves. The next is the speed of internal gravity waves. The comparatively small value of the speed of motion relative to the Earth ($V$) emphasises the rapid rotation of the planet and the relatively small deviation of the atmosphere (and even more the ocean) in its motion from solid body rotation with the planet. The last inequality emphasises the strong stratification of the atmosphere (and ocean). Behaviour in the local vertical and horizontal directions is therefore very different. In the vertical there is generally stable stratification and a balance between the very large gravitational and pressure gradient forces. In the horizontal, the much smaller Coriolis force associated with the rotation of the Earth can be important. For synoptic scales, and indeed for larger scales away from the equator

$$\frac{V}{fL} < 1,$$

where $f = 2\Omega \sin \phi$, the Coriolis parameter, is twice the local vertical component of rotation of the Earth. This implies that the basic momentum balance in the horizontal is between the Coriolis and pressure gradient forces, and that $v$ is approximately geostrophic:

$$v \approx v_g = (\rho f)^{-1} k \times \text{grad } p$$
The potential temperature, \( \theta \), is the temperature air would have if it was taken adiabatically (no heat added) to a standard pressure (usually 1,000 hPa). In the absence of heat sources and sinks, it is conserved.

It is very useful to have another quantity that involves the dynamics and is conserved following the fluid in the absence of diabatic and frictional processes but, in these circumstances, the absolute circulation (\( C \)) around a closed material line on a constant \( \theta \)-surface is also conserved. However this is difficult to use directly. If the closed material line shape is used to make a material cylinder between this \( \theta \)-surface and its neighbour at \( \theta + \delta \theta \), then both the mass, \( m \), of the cylinder and \( \delta \theta \) are also conserved. Therefore the quantity

\[
C \times \delta \theta/m
\]

is conserved. Writing the circulation in terms of the absolute vorticity \( \zeta = f \mathbf{k} + \text{curl} \mathbf{v} \), a measure of the local rotation in the fluid, and using derivatives, this conserved quantity may be written:

\[
\rho^{-1} \zeta \cdot \nabla \theta.
\]

This is called the potential vorticity (PV) and is conserved moving with the fluid in the absence of heat sources and sinks. The PV involves the dynamics as well as the thermodynamics: from the derivation given here it is basically a measure of circulation on a \( \theta \)-surface divided by mass between isentropic surfaces.

In most large-scale motions of interest there is balanced dynamics involving the Coriolis force. The simplest example is geostrophic motion. In such cases the 3-D distribution of PV, along with suitable boundary conditions, can be inverted to give all the details of the balanced flow. The large stratification is associated with a large vertical component of \( \nabla \theta \), and so, on synoptic and larger scales, it is the vertical component of absolute vorticity, \( \zeta = f + \mathbf{k} \cdot \text{curl} \mathbf{v} \), that is most important for PV and so in the analysis of atmospheric motion.

4.1.2 Atmospheric Phenomena

A vast range of phenomena occur in the atmosphere and it is essential when modelling the system to consider which of these are to be simulated, and what are their characteristics that have to be represented in the model and diagnosed in atmospheric or model data. Probably the most fundamental of these on the larger scale is the Rossby wave. Its nature can be understood by considering the situation shown in Fig. 4.1. The equatorward initial perturbation in contours of the vertical component of absolute vorticity implies a positive vorticity anomaly. Associated with this will be cyclonic motion, as shown. The equatorward wind to the west implies a
Fig. 4.1 A simple description of Rossby wave dynamics. The basic situation is one with high absolute vorticity (or more generally PV as marked here) poleward and low absolute vorticity equatorward. The starting point is a local equatorward displacement of the absolute vorticity contours, leading to a cyclonic anomaly (represented by a +). This induces north-south flows as shown, and these in turn lead to vorticity anomalies as indicated in the panel below. The result is a westward movement of the cyclonic anomaly (+) and the development of a new anticyclone to the east. These features correspond, respectively, to westward “phase speed” and “eastward group velocity”. A basic westerly flow (u) will add on to them, giving a reduction of the former and an increase of the latter.

tendency to extend the initial positive vorticity anomaly in this direction: a westward “phase speed”. The poleward wind to the east of the original positive vorticity anomaly implies a tendency to create a negative vorticity anomaly there. This means that the region of wave activity extends in this direction: an eastwards “group velocity”. The propagation of the wave activity is measured by the group velocity which is therefore eastward. If a basic westerly flow is added, then the eastward group velocity becomes larger and the phase speed can become zero, depending on the wavelength. The discussion of Rossby waves given here can be extended to apply to PV on \( \theta \)-surfaces, to waves with their crests and troughs tilted from the north-south direction, and also to realistic flows on the spherical Earth, in which case propagation tends to be along great circle paths rather than east-west lines.

The existence of such stationary Rossby waves is very important because it means that there can be coherent remote responses to stationary wave sources such as mountains and regions of persistent deep convection such as the western tropical Pacific with its high SSTs. This response can occur on planetary scales in a wide arc on the eastern side of the wave source. Such responses lead to the climatological average waves and also to monthly or seasonal anomalies, which are normally associated with a sequence of height field anomalies of alternating sign.

An example for October 2000, which was one with record-breaking rainfall in England and Wales, is given in Fig. 4.2. The low height-field, cyclonic anomaly
over and to the east of the UK is seen as part of a wave pattern. The group velocity arguments suggest that the origin of the anomalous pattern should be sought to the west, in the Caribbean/Americas region.

The picture for October 2000 is probably not as simple as this might suggest. The North Atlantic storm-track extends from the coast of N America towards NW Europe. The weather in Europe is strongly dependent on the position and intensity of the storm-track. The anomalous large-scale flow in October 2000 will have influenced the storm-track. However the storms themselves will have fed back on the larger-scale flow through their vorticity and heat transports, thereby changing it. Fluctuations in the North Atlantic near surface westerly flow and in the storm-track are frequently characterised in terms of the North Atlantic Oscillation (NAO). There is much current interest in possibly predictable monthly to seasonal timescale behaviour of the NAO that may be related to the strength of the lower stratospheric vortex or to sea surface temperature (SST) patterns.

The absence of storms affecting a region can be associated with a phenomenon referred to as blocking, often characterised by a persistent deep positive height field anomaly. It is thought that blocking can occur as an interaction between weather systems and an anticyclonic anomaly, which may itself form part of an anomalous stationary Rossby wave train. Blocking is particularly important for Europe, being associated with anomalously dry or wet weather, depending on location, and warmer or colder weather, depending on the season.
In the tropical region a common occurrence is for frequent deep convection to occur in a large region for many days. This implies large latent heat release in this region. The response of the atmosphere to this heating usually has the general characteristic flow pattern which is shown schematically in Fig. 4.3. The middle tropospheric latent heating is balanced by adiabatic cooling associated with ascent. Off the equator this implies vortex stretching and the generation of cyclonic vorticity below and vortex shrinking and the generation of anticyclonic vorticity above. As in the Rossby waves argument, the lower and upper tropospheric circulations extend and move to the west, the latter process continuing until the parts of the circulations with, respectively, poleward and equatorward moving air are in the heating region. Through considerations of balance, associated with the change in the sense of the circulation with height, the mid-troposphere to the west of the heating must be warm. Such a pattern of circulations can be associated with anomalous heating in any month, perhaps associated with higher or lower SST than usual. It can also act as source for a Rossby wave train propagating into higher latitudes, perhaps like that seen in Fig. 4.2 for October 2000.

A particular example of such heating and associated global anomalies is the tropical Intra-Seasonal (Madden-Julian) Oscillation. Large regions of much intensified or weakened convection move slowly from the western Indian Ocean to the west Pacific and perhaps continue to the dateline on a monthly timescale. Again this offers the possibility of predictive power both in the tropics and in higher latitudes.

When, such as in the Asian Monsoon, the summer tropical heating region extends to high enough latitudes there is an interaction with the extra-tropical westerlies. These westerlies flow down the sloping $\theta$ surface on the western side of the circulations with their mid-tropospheric warmth, enhancing the descent there. This descent can be further enhanced and localised by topography. Radiative cooling, in the absence of convective heating, can then produce further enhancement, leading

---

Fig. 4.3 A schematic showing the response to large-scale tropical convective heating. The convection is shown by a cloud and is assumed to span the equator.
to very strong local descent. Rodwell and Hoskins (1996) proposed that this is the basic mechanism for the summer climate of the Mediterranean, which is therefore seen as part of the Asian Summer Monsoon. Such remote associations are very important in providing a context for considering seasonal anomalies and their forecasting. Indeed, significant weakening of the Asian summer monsoon may allow North Atlantic weather systems to enter the Mediterranean and then move into southern Europe as in the summer of 2002.

4.2 The Role of the Ocean in the Climate System

The fundamental role of the oceans in the climate system is to (1) act as a buffer for mitigating transients, (2) contribute to the required pole-to-equator heat transport, and (3) provide hidden “memory” in the coupled atmosphere-ocean-land system.

Over much of the planet, the ocean can be considered to be well-represented as a surface mixed layer whose temperature \( T \) obeys a simple heat conservation law:

\[
\rho \cdot c_p \cdot h \cdot \frac{\partial T}{\partial t} = Q_s(t)
\]

Here \( \rho \) is the density and \( c_p \) the heat capacity of seawater. The depth of the mixed layer \( h \) varies in space and time – it is typically thin during summer months, and thick during the winter. The wintertime depth can reach hundreds of meters or more at high latitudes, while during the summer a depth of 10–20 m might be found. Figure 4.4 shows the climatological mean profile of temperature at 40°N, 170°W in the ocean for February and August. During the winter, the cooling and wind cause the ocean to be well-mixed down to 200 m, while during the summer,

![Fig. 4.3](image1.png)

**Fig. 4.3** Climatological temperature values at 40°N, 170°W in the North Pacific for February (heavy) and August (light) (Data from NOAA NODC World Ocean Atlas 2005)
the very shallow surface layers warm up considerably. The surface fluxes ($Q_s$) may include a wide range of frequencies, including diurnal cycles, synoptic atmospheric weather, seasonal cycles and longer term climate changes. Due to thermal inertia, the ocean mixed layer will damp the high frequencies, providing a “reddening” of the spectrum. The deeper the mixed layer, the more pronounced the reddening. For the problem of seasonal climate prediction, modelling this mixed layer behaviour over the open ocean captures most of the essential physics over much of the ocean. Theories and models exist for simulating the 1-dimensional behaviour of turbulent mixed layers under the combined effects of heating and wind.

The above heat balance assumes that there is no heat flux out through the bottom of the mixed layer. For many problems, this treatment is adequate, but such an approximation will not permit any transport of heat from one latitude to another. The circulation in the ocean can carry heat into or out of the mixed layer. In regions of strong currents this extra heating can become important. The mixed layer budget becomes

$$\rho \cdot c_p \cdot h \cdot \frac{\partial T}{\partial t} = Q_s(t) - Q_c$$

In the time mean, a balance must exist between this circulation-induced heating and the surface heat flux, or $Q_s = Q_c$. It is this spatial variation of the circulation-induced heating that enables a net ocean heat transport. Circulation-induced heating can be caused by vertical motion at the bottom of the mixed layer, horizontal currents or turbulent mixing. There are a few distinct regions where the ocean currents strongly affect the surface heat balance: the western boundary currents, such as the Gulf Stream and Kuroshio, the Antarctic circumpolar current, and the tropics, particularly within the equatorial wave guide. The tropics get special attention in the seasonal to interannual climate problem not only because the ocean dynamics plays this strong role, but because the atmosphere responds strongly to the ocean-induced changes. In the mid-latitudes, the oceans carry heat, but the atmospheric response to this heating is not as strong, and does not cause secondary effects that influence the circulation on these timescales.

4.2.1 The Thermocline – Setting the Stage for El Niño

While the ocean is very deep, most of the important dynamics for seasonal to interannual timescales happen within the relatively thin warm region at the top of the ocean known as the thermocline.\(^1\) Oceanographers now understand that the

\(^1\) The term ‘thermocline’ originally referred to the region of strong thermal gradient, but recently is sometimes associated with the entire upper ocean through the development of the ventilated thermocline theory (see, for instance Pedlosky 1996).
character and shape of the thermocline is described by a dynamical construct known as the ventilated thermocline (Luyten et al. 1983). In a static view, one would expect the thermocline to be deepest at the equator, where the warmest surface waters are found. But instead, the thermocline almost vanishes along the eastern end of the equator in both the Pacific and Atlantic. Cold water from below the main thermocline is exposed to the surface in a feature commonly referred to as the “cold tongue”. Figure 4.5 and Figure 4.6 show climatological temperature sections of the top 500 m of the ocean along the dateline and equator, respectively, for February and August. Note how in Fig. 4.5 the 20°C isotherm is deepest at about 20° latitude along the dateline, but that in Fig. 4.6, it comes very close to surface at the eastern end of the equator. We can see that close to the surface, seasonal effects matter, but once deeper than a hundred meters or so, the seasonal effects are smaller.

As we will see later in this section, El Niño models function by predicting perturbations that happen to the thermocline. Its structure affects the sensitivity of the surface temperature to the subsurface ocean variability, which in turn affects the coupling between the ocean and atmosphere. Models for El Niño have shown sensitivity to the sharpness and tilt of the thermocline. A new body of research has emerged on how long-term variations in the thermocline occur and how they might influence the evolution of El Niño.

The equatorial thermocline connects to the subtropical thermocline through a circulation system known as the subtropical cells (STCs) or shallow overturning circulation. Work by McCreary and Lu (1994) has shown that the equatorial “cold tongue” is not simply an accident of having a thermocline and easterlies along the equator, but is an essential property of the STC circulation.

These cells have a three-dimensional circulation structure, with largely poleward surface branches and equatorward sub-surface flow. When the water flows along the surface, it is in constant contact with the atmosphere, and its properties are altered by the surface fluxes of heat and freshwater. Once removed from the surface, turbulent mixing and heating is much smaller, and the water is found to conserve its property over long distances and long times. A common approximation is that the flow is adiabatic, and flows along surfaces of constant density or “isopycnals”.

McCreary and Lu showed that at about 15° north and south latitudes in the Pacific, the lower branch of the cell has a net flow toward the equator, when integrated completely across the basin. Where does this water go? Except for a small leakage through the Indonesian throughflow, there is no horizontal outlet. The water coming in at this depth must rise to the surface (upwelling) within this tropical band. With the mean easterly Trade winds, the necessary upwelling conditions apply at the eastern end of the equator, along the American coasts, and in a few isolated regions such as the Peru upwelling and the Costa Rican dome.
Fig. 4.4 Climatological temperature sections along the dateline for February and August (Data from NOAA NODC World Ocean Atlas 2005)

Fig. 4.5 Climatological temperature sections along the equator for February and August (Data from NOAA NODC World Ocean Atlas 2005)
Further studies with more complex ocean circulation models have investigated the source waters of the equatorial under current or equatorial cold tongue, and are largely in agreement that the source waters of the equatorial under current and equatorial cold tongue lie well within the subtropical gyres. Observations support the canonical view of the STCs (Johnson and McPhaden 1999).

In summary, the large-scale circulation of the top several hundred meters of the oceans creates a thermocline with warm water overlying cold, and this thermocline is constrained to be tilted along the equator, with the cold water showing up at the surface at the eastern end. This sets the stage upon which fluctuations act to produce El Niño and La Niña.

### 4.2.2 Variations on the Thermocline

Once we understand that the thermocline should exist, and that it should surface at the equator, it seems natural to ask whether this outcropping is stable, or whether the system can experience an oscillation or be disturbed by local wind and weather effects. The theory of McCreary and Lu basically states that “since on the average, \( x \text{ kg/sec} \) of cool water converges toward the equator from both hemispheres, on the average \( x \text{ kg/sec} \) of cool water must surface”. One might think that this sets the temperature of the cold tongue, and that El Niño arises from changes in this process. But this applies only over a suitable averaging interval. We know that this flow can take decades to close the loop. For changes that take only a few months or even a few years, another theory is needed that permits significant variations, and it is these variations that are now known to be the root cause of El Niño and most of tropical climate variability.

For this theory, we turn to the dynamics of internal gravity waves as modified by the special features of planetary rotation – the equatorial Kelvin and Rossby waves that propagate signals east and west in the equatorial waveguide (see Moore and Philander 1977 or Gill 1982 for a synopsis of equatorial wave dynamics).

Because of the special nature of the Coriolis effect near the equator, low frequency planetary waves take on distinct properties, with one wave type propagating eastward (the Kelvin wave) and a set of others with westward propagation (the equatorial Rossby waves). Each wave describes the evolution of the thermocline perturbation \( h \), the zonal current \( u \), and the meridional current \( v \). The Kelvin and Rossby waves are different “modes” of the system. Each mode has a different pattern in the north-south direction, but they tend to keep all the action

---

2 Rothstein et al. (1998); Harper (2000); Huang and Liu (1999); Malanotte-Rizzoli et al. (2000); Rodgers et al. (2003); Fukumori et al. (2004).
near the equator. Mode 0 is the Kelvin wave, with a phase speed and group velocity to the east with the same speed as an internal gravity wave. Modes 1, 3, 5, ... are antisymmetric, with $h = u = 0$ at the equator, and a maximum in meridional velocity at the equator. Modes 2, 4, 6, ... are symmetric with perturbations in height and zonal current that have local maxima on the equator, and no meridional flow across the equator. Note that the gravest symmetric Rossby mode ($n = 2$) propagates to the west with $1/3$ the speed of the Kelvin wave. The higher mode Rossby waves propagate slower still.

Figure 4.7 shows the meridional structure of the Kelvin wave. The Kelvin wave is somewhat special because it has no meridional current ($v = 0$), and the meridional structure for $h$ and $u$ are the same. Note that the successively higher Rossby modes have amplitude extending further from the equator, and have a more oscillatory behaviour.

![Fig. 4.6 Meridional structure functions for the equatorial waves. These functions depict the relative amplitude of pressure perturbations from the Kelvin (K) wave and the first three symmetric Rossby waves (2, 4 and 6) from linear solutions with an internal gravity wave speed of 3 m s$^{-1}$. Anti-symmetric Rossby modes also exist, but are not shown.](image)

In the ocean, the Kelvin and Rossby waves are largely forced by the wind. Weakening the trade winds in the centre of the Pacific will generate Kelvin waves that cause the thermocline to deepen while at the same time driving Rossby waves that cause the thermocline to shallow. Looking along the equator, one would see the deepening signal head off to the east and a shallowing signal heading west.

Other than the slightly strange meridional structure and modified phase speeds, the Kelvin-Rossby wave set behave like internal gravity waves in a channel. Only one wave can send signals to the east, while all the rest send signals westward. Unlike a bounded channel, however, the reflection properties of the Kelvin and Rossby waves are different. When Kelvin waves reach the eastern end of the equator, they propagate poleward along the coast. These coastal Kelvin waves shed
some Rossby wave energy as they go, but to a large extent, much of the energy in the equatorial Kelvin wave is propagated out of the equatorial zone.

At the western boundary, Rossby waves have a somewhat complicated reflection. Cane and Sarachick (1977) demonstrated that Rossby waves reflect into equatorial Kelvin waves. When the zonal mass flux in the Rossby wave is integrated in the meridional direction, there can be no net accumulation of mass. The Kelvin wave that is reflected has sufficient amplitude to balance this mass convergence.

What we see from the wave reflections is therefore a “leaky” system: Rossby waves propagating westward return their energy in a Kelvin wave travelling eastward, but these Kelvin waves leak their energy to higher latitudes when they reach the eastern boundary. This leak means that the system can not simply resonate like a closed channel, with waves forever bouncing back and forth. Instead, any sustained oscillations of this system must involve forcing.

These wave dynamics describe the forced and freely propagating linear response of the upper ocean to imposed wind stresses. They describe the motion of a simplified representation of the thermocline. These motions might be interesting in and of themselves, but until and unless they change the SST, they will not have any influence over how the atmosphere evolves in time. But if the motions do alter the SST, and this change in SST causes changes in the atmosphere that further change the winds, then there can be a feedback loop between anomalous winds, anomalous currents in the ocean, SST perturbations and finally back to the anomalous winds. This loop is known as a closed feedback loop, in which perturbations in the system propagate from one variable to another. Closed feedback loops can lead to instabilities in the system.

The most important modification needed to our theory is to improve the representation of the ocean temperature. Although the upper ocean may be treated as two distinct layers to explain the essential dynamics, the actual ocean thermocline is a region of continuous gradients of temperature and salinity. The forced wave motions that alter the thermocline therefore introduce a continuous change in the surface temperature.

We concentrate on the eastern equatorial Pacific, because this is where the thermocline outcrops and causes the “cold tongue”. Bjerknes (1966) noted that westerly wind anomalies in the central part of the Pacific would drive Kelvin waves to the east that would deepen the thermocline and carry warmer water to the east. Both effects will cause warming of the SST in the east. Theories, models and observations of the atmospheric response to the warmer surface temperature agree that such warmer SSTs will lead to further strengthening of the westerly winds. Thus a positive feedback loop exists that can extract energy from the system. This source of energy is key to overcoming the “leaky” nature of the wave reflection arguments made earlier.

A second important point should be noted about the interaction of the waves with the SST: in the western Pacific, and away from the zone where the thermocline outcrops, changes in the depth of the thermocline do not have a perceptible
influence on the surface temperature. Although changes have been introduced into the ocean, and waves are sending signals around, the ocean has “sequestered” the information from the atmosphere. This is a key ingredient of the delayed oscillator mechanism, which we discuss next.

### 4.2.3 The Delayed Oscillator Theory of Enso

We now have a view of the system where the thermocline tilts up to the east and exposes cold water to the surface. Changes in the position of this thermocline are reflected in changes to SST which perturb the atmosphere. These changes give rise to a positive feedback through the winds to drive an unstable growth. If $h_e$ is the thickness of the thermocline in the eastern Pacific, and $\hat{\tau}$ is the zonal wind stress averaged across the basin, we have

$$
\hat{\tau}(t) = A h_e(t)
$$

$$
\frac{\partial h_e}{\partial t} = B \hat{\tau}
$$

where $A$ and $B$ are proportionality constants. Then,

$$
\frac{\partial h_e}{\partial t} = AB \hat{\tau}
$$

This is a simple view of the Bjerknes instability. In his original paper, Bjerknes noted how this feedback can explain the emergence of El Niño events, but he then remarked on the difficulty in finding a reason for the system to turn around and go from warm to cold. (Or, for that matter from cold to warm, as for example at the end of La Niña.) Since we have a rationale for the thermocline to be exposed to the surface in the eastern Pacific over the long term, such perturbations as described by Bjerknes can not take over and control the result forever. This, plus the observed preference for El Niño to occur every 3–7 years led to a search for a mechanism that could explain an oscillation in the equatorial system. One solution can be found in the delayed action oscillator.

As explained above, the winds that perturb the ocean by driving Kelvin waves to the east also drive Rossby waves to the west. They propagate on the deeper thermocline, hidden from the atmosphere. When they reach the western boundary, they reflect into equatorial Kelvin waves, and propagate back to the east. In this case, the equation for $h_e(t)$ must be modified to include the effects of the Rossby waves. The eastward propagation of Kelvin waves along the equator and the poleward propagation along an eastern boundary can be clearly seen in Fig. 3.3. This figure also shows the westward propagation of Rossby waves from the eastern boundary and their generation in mid-ocean and westward propagation to the western boundary.
The important feature of the Rossby waves is that a wind that drives a shallow- ing Kelvin wave will drive a deepening Rossby wave, and that a deepening Rossby wave reflects into a deepening Kelvin wave. Thus, the evolution of the height field in the east is a combination of the Bjerknes instability and information from some “old” Rossby wave:

$$\frac{\partial h_e(t)}{\partial t} = ABh_e(t) - C\hat{\tau}(t - \Delta t)$$

The factor $C$ includes the effects of how the wind drives the Rossby wave, how efficiently the western boundary reflection works and how the Kelvin wave alters the thermocline thickness. The time $t-\Delta t$ reflects the fact that the height at present time is influenced by the wind that existed in the past – at the time that the Rossby wave was first generated.

But once again, since $\hat{\tau}$ at any moment is presumed to be proportional to the SST anomaly, which is presumed to be proportional to the thermocline displacement, we can combine all these proportionality factors and arrive at:

$$\frac{\partial h_e(t)}{\partial t} = ABh_e(t) - Dh_e(t - \Delta t)$$

This equation is a differential-difference equation that describes the basic delayed action oscillator. Under certain conditions, this equation can lead to growing oscillations.

In their original proposal Schopf and Suarez (1988) include a cubic damping term which is intended to reflect the fact that SST can not grow without bounds: In our advective model, if the thermocline floods in completely from the west, the surface temperature can not get much above 30°C, because that is the warmest water available. Similarly, because the process works as an uncovering of the thermocline, if too much water is brought up, the surface will see the relatively uniform intermediate water that lies just below the thermocline. Their proposal is therefore

$$\frac{\partial h_e(t)}{\partial t} = ABh_e(t) - Dh_e(t - \Delta t) - \gamma h^3_e(t)$$

This cubic term means that the system will not grow without bound, but will undergo regular oscillations of a fixed amplitude. The modification to a non-linear system is not fundamental to our understanding of the mechanics of the delayed action oscillator.

By rescaling time with the growth rate $AB$, and the dimensional $h_e$ with $(AB/\gamma)^{1/2}$, we see that the oscillator depends on two parameters:

$$\frac{\partial h_e(t)}{\partial t} = h_e(t) - \alpha h_e(t - \delta) - h^3_e(t)$$
where \( \alpha = D / AB \) and \( \delta = AB \Delta t \). The system described may undergo self-sustained or damped oscillations, depending on the location of the base system in the parameter space described by \( \alpha \) and \( \delta \). When oscillations are present, they have a period in excess of twice the delay. They are typically far greater than this. See McCreary and Anderson (1991) for a full coverage of the various types of response as a function of \( \alpha \) and \( \delta \).

The delayed action oscillator succeeds in describing a mechanism whereby a preferred periodicity for El Niño may exist. As one can see, there are several parameters which are not easy to quantify, and attempts to diagnose whether the system should exhibit self-sustained oscillations or not have been made, but they are inconclusive. Instead, experiments with numerical models have been designed to examine the point, but in the end there is less to be learned from examining the stability question than there is in understanding the elements of the system.

The key elements of the delayed oscillator are:

1. Coupled instability in the east via Bjerknes mechanism
2. Low frequency Rossby wave generation that perturbs the thermocline
3. Reflection of the thermocline displacements into an equatorial Kelvin wave
4. “Coupled reflection” at the east

This last point is an interesting twist on what one would expect if gravity waves bounced back and forth across a closed basin. Instead of a period that is set by the time it takes a wave to go back and forth across the basin, the delayed oscillator operates with a period which is at least twice that time. Recall that the reflection of a deepening Kelvin wave at the eastern boundary causes a weak set of deepening Rossby waves. In the delayed oscillator, the Rossby waves are not generated at the coast, but through the coupling process whereby deepening Kelvin waves give rise to shallowing Rossby waves. This phase reversal is key to the period-doubling inherent in the system.

The delayed action oscillator theory demonstrates that El Niño arises from a coupled instability in the ocean-atmosphere system. Neither an ocean-only nor atmosphere-only model can explain El Niño and its dominant frequency. It implies that the memory in the system lies in the thermocline, off the equator. In Chapters 5 and 6, we discuss the nature of the prediction system, and how model initialization, and particularly ocean data assimilation is essential to successful forecasts of El Niño. One of the main reasons for this lies with the information contained in the ocean thermocline and the dynamics of how that information propagates through the system, only later to show up as changes in the surface temperature.

### 4.2.4 The Recharge Paradigm

An alternative to the delayed oscillator theory is the recharge paradigm for El Niño. In this view, it is recognized that the Rossby and Kelvin waves cross the basin far
more quickly than El Niño changes to La Niña. When fast-moving waves are forced slowly, it is hard to recognize them as waves at all. Instead of describing the changes in the state as due to wave propagation, perhaps we can describe the ocean as in quasi-steady state.

Jin (1997) was able to use this property to develop a simpler system of equations that describe an oscillator. When the waves are fast, the equatorial Kelvin wave can be written in the very simple form:

\[ h_e(t) = h_w(t) + a_1 \hat{\tau} \]

where \( h_e \) is the thermocline height at the eastern end of the equator, and \( h_w \) is the height at the west. \( \hat{\tau} \) is the average zonal wind stress across the basin, and \( a_1 \) is a proportionality factor.

Anderson and Gill (1975) demonstrated how the steady circulation of the ocean (the so-called Sverdrup flow) is established by the net effect of Rossby waves. In the recharge view, the explicit treatment of the Rossby waves of the delayed oscillator are replaced with Sverdrup flow, which causes mass to converge toward the equator, thereby setting up changes in the thermocline in the west. An equation for the thermocline thickness in the west is then

\[ \frac{dh_w}{dt} = -rh_w - \beta \hat{\tau} \]

where \( r \) is a damping factor, and \( \beta \) is a proportionality factor that builds in the different projection of the winds onto the modes as well as a number of other effects.

The coupling in the recharge oscillator occurs through the SST, as in the delayed action oscillator: the stress is proportional to the temperature in the east. In the delayed oscillator, the relationship between the thermocline depth and the SST is treated as due to several factors:

\[ \frac{dT_e}{dt} = -a_2 T_e + a_3 h_e + a_4 \tau_e \]

where \( T_e \) is the SST, which is damped by surface fluxes to some equilibrium, \( a_3 h_e \) reflects the contribution of upwelling, and \( a_4 \tau_e \) represents an advective feedback due to wind stress local to the east.

These terms are then related to two variables as primary \( h_w \) and \( T_e \): the winds (both \( \hat{\tau} \) and \( \tau_e \)) are made proportional to \( T_e \). If time is scaled with the Bjerknes instability growth rate and \( h_w \) scaled appropriately, the coupled set of ordinary differential equations can be written

\[ \frac{dT_e}{dt} = T_e + h_w \]
\[ \frac{dh_w}{dt} = -\hat{r}h_w - bT_e \]
The recharge oscillator and delayed action oscillator share the same ocean dynamics (low frequency forced modes), and depend on two parameters. In both views, one parameter sets how fast the western basin fills in relation to the time-scale of the Bjerknes instability. In the delayed oscillator, it is the Rossby wave propagation time, while in the recharge oscillator, it is the damping parameter $\hat{r}$. Both theories also have a free parameter describing how strongly the conditions in the west influence the SST in the east.

Given their shared view of ocean dynamics and their ultimate dependence on two similar parameters, it is not possible to differentiate the two theories based on observations or model experiments. For most questions, they share similar challenges. For instance, it has been noted that the recharge paradigm depends on the latitude at which one wishes to compute the Sverdrup flow. But the delayed oscillator can consider more than one meridional Rossby mode, with higher modes extending further poleward and travelling at slower speeds. The delayed oscillator is criticized because it may be possible for Rossby waves to propagate through the Indonesian archipelago.\(^3\) But in the quasi-steady Sverdrup flow of the recharge paradigm, the buildup of mass in the west may be returned poleward in western boundary flows or may pass through Indonesia just as the low frequency Rossby waves. In short, there is little to be gained from differentiating these two views.

### 4.2.5 Conclusion

The ocean is but one part of the climate system. We have discussed how the ocean takes up heat to buffer the high frequency changes induced by the atmosphere. Next we noted that the large scale circulation created by the combined effects of winds and surface heating does not drive the entire ocean uniformly, but leads to a rather shallow circulation that is described by the ventilated thermocline. This thermocline connects to the equator via the shallow tropical cells, and an inevitable consequence of the atmospheric forcing is that this thermocline will emerge at the eastern end of the equator (at least in the Pacific).

Coupling to the atmosphere and a simple deterministic view of the atmosphere led us to discover that this tilted thermocline is perhaps not a stable stationary state, but can possibly have unstable, self-sustained oscillations. These oscillations lie at the heart of El Niño and La Niña. Whether or not dynamics such as the delayed oscillator are strong enough to cause spontaneous changes to the system,\(^3\) Schopf and Suarez (1990) show that a reflection efficiency as low as 15% is sufficient to permit oscillations.
it is clear that other perturbations to this tilted thermocline are capable of causing significant changes in the equatorial surface temperatures. Storms, sub-seasonal variations, and other unpredictable features of the tropical atmosphere will all leave their imprint on the thermocline. Some may lead to expressions in the SST that will give rise to coupled instability, some will pass through the system as sub-surface Kelvin waves with little hope of making a sustained change in the climate system.

The challenge to the problem of seasonal climate prediction via dynamical models is to build models that must capture all of the essential physics – the mixed layers, ventilated thermocline, shallow tropical cells, wave dynamics, and thermodynamics of how the thermocline emerges, and they must be able to be initialized with the important information that contains the dynamics of the evolution.

### 4.3 The Nature of the Prediction Problem

The problem of seasonal climate prediction is one of attempting to simulate the seasonal average of the weather, not the individual fronts, cold snaps, or storms. These “weather” events have been shown to have no predictability beyond a week or two (see also Chapter 3). If the climate is the sum of weather, but the weather is unpredictable, does this not imply that the climate is unpredictable? In fact, the answer is no, the climate can be predictable considerably longer than the weather. If a forecast system fails to predict a storm 10 days from now, but predicts one 12 days from now, the forecast is wrong, but the average number of storms in the next month will be correct. The prediction problem relies upon the fact that some parts of the system evolve slowly, while others are of short duration. If the short events are unpredictable, but an equation can be written to describe the slow evolution, then the high frequency component can be considered unpredictable “noise”, and the challenge is to describe the effect of noise on the solution of the slow equations.

In Chapter 3 and in the previous sections of this chapter we discussed some of the current theories for El Niño/La Niña, which involve the propagation of signals on the ocean thermocline, transformation of these signals to SST anomalies, then coupling to the atmosphere, modification of the winds and driving of the ocean. We derived equations for this slowly evolving part of the system. The oscillator theory is very simplified, however, and much can disturb the process. Each El Niño develops differently, and the magnitude can vary greatly from one event to the next. The examination of this irregularity is fundamental to understanding the prediction problem, because it lies at the core of understanding the “predictability limit”.
4.3.1 Predictability Limits

The predictability limit is a concept that describes our recognition that we do not do as well with models as we can, but that even a perfect model and perfect initialization will be unable to forecast the climate forever. If the models are inherently flawed, then the predictability limit may be a gross overestimate of how long we can make a successful forecast, but it is useful to try to approximate this limit. If, for instance, it can be demonstrated that no model/initialization system can forecast for more than 2 months, why bother to build better and better models and more and more expensive observing systems? If, on the other hand, it can be shown that forecasts of up to 3 years can be made, then we had better put a lot more effort into our models, observing systems and initialization methodology.

Unfortunately, there is no absolute way to define a predictability limit. We can study how models behave, using the “perfect model” technique. To study predictability, we want to know how fast a perfect model diverges from nature. Unfortunately, although we can know what nature did over the past, we can not construct a perfect model. Instead, we can examine predictability by replacing nature with a model simulation. For this model simulation, there does exist a perfect model – the model itself. There also exists perfect initial conditions and perfect initialization.

If we run the same computer code on the same computer many times with widely different initial conditions, the solutions will enclose a wide region of phase space that describes the climate and its variability. One should see the seasonal march of temperatures, for instance, but the model simulation for a specific day will vary considerably from one run to the other. This spread in the results of a random collection of model runs is known as the “saturation”.

If we run the same computer code on the same computer many times with exactly identical initial conditions, the model will produce identical results forever, unless a coding error exists. There will be no spread between results. But if we introduce a very tiny error in the initial conditions, the model runs will diverge. At first, if we repeat this experiment over and over with many initial conditions that differ by small amounts, we find that the spread in these results is far smaller than the saturation. But eventually, solutions with even the tiniest of initial errors will reach a spread indistinguishable from saturation. When this occurs, we have reached the predictability limit.

Thus, we can define the predictability limit in the context of a model. But does this model represent nature? Is it close enough? If the experiments are repeated with another model, will the results be the same? If they are, is it because the two models reflect nature, or because they share a systematic bias or systematic error that leads to this behaviour? These are questions which confront the theoretician trying to deduce a predictability limit.
4.3.2 Enso Irregularity and Predictability

Is there a relationship between the fact that El Niño is irregular and its predictability? This irregularity reflects the complexity of the coupled ocean-atmosphere system and hints at the difficulties in predicting ENSO. Is it due to noise in the system, the inability to adequately specify the initial conditions, inherent deficiencies in the models, or to not-yet-understood fundamentals of the physical system.

Theories on the cause of ENSO irregularity can be broadly grouped into three categories that are related to their assumption about the strength and validity of the underlying oscillator and the importance of noise. We have presented the delayed action and recharge oscillators as theories for the dominant periodicity of El Niño, but there are debates as to whether they actually operate. We know that in certain parameter ranges, the equations for these simple systems will describe robust oscillations, while in others, the only solution will be a decaying, damped oscillation. The first is self-sustained, the latter requires some external forcing to keep the system going. The three categories of theory on El Niño irregularity split into a view that the oscillators are self-sustained, that they are damped, or that they are essentially neutral. The role of non-linearity and noise is markedly different in each case, and our view of predictability is different in each.

The first view argues for the importance of non-linearity within the tropical coupled system. The non-linearity arises from strong air–sea feedback that puts the coupled mode in an unstable dynamic region. In this regime, El Niño can not only be described as due to a self-sustained oscillator, but it can interact non-linearly with either the annual cycle or other coupled modes. A common model that is cited in this regime is the Zebiak-Cane coupled model, which can be configured to exhibit strong non-linearities and chaotic behaviour. In this view, the loss of predictability is primarily due to the uncertainty in the initial conditions or in non-linearities in the atmospheric response to the ocean. It relies upon fairly robust ocean wave dynamics that provide the underlying timescales for the problem.

The opposing view to this is the stochastic ENSO theory in which “weather” noise generated by the internal dynamics of the atmosphere plays a fundamental role in not only giving rise to ENSO irregularity, but also in maintaining ENSO variance. In this view, the coupled mode is in a damped regime, and thus the ENSO cycle cannot be self-sustained without external noise forcing. The oscillator describes a tendency for the system to have a preferred period, but does not explain much about the appearance of any single event. It is the exact pattern of the noise and how it forces the weakly coupled modes that determine whether a large or small El Niño or La Niña will next appear. The cyclic nature of the underlying oscillator merely alters the odds a little in favour of one side or the other. In this view, the role of the equatorial Kelvin wave and the equatorial air-sea coupling is important, and the off-equatorial ocean dynamics seems less vital.
In between these two viewpoints is the view that ENSO is very close to the dividing line between self-sustained and damped behaviour. Its behaviour is governed by the temporal characteristics of the single, most dominant coupled mode plus the influence of weather noise. In this scenario of ENSO, predictability comes from the oscillatory nature of the dominant mode, while the loss of predictability is primarily due to noise influence. Different from the stochastic ENSO theory where the noise influences the non-modal growth of the coupled system, the role of the noise in this case is to disrupt the regular oscillation of the dominant mode. In this regime, the ocean wave dynamics and reflection properties must also be sufficient to sustain the oscillation.

An extension of this view is the notion that over decades, the system can wander across the dividing line between self-sustained oscillations and a damped regime – so that predictability may vary from 1 decade to the next (Kirtman and Schopf 1998). This concept of time-varying predictability is an important one to bear in mind when considering the skill of previous forecasts and whether this means that our current forecasts are “better” or “worse” than before.

Pinpointing exactly where in the parameter regime ENSO resides in reality is difficult, if not impossible, given the available observations. Many of the recent studies on this issue are based on relatively simple coupled model simulations and prediction experiments. Some of the evidence supporting stochastic ENSO theory is based on the finding that in the damped regime the coupled model forced by stochastic processes produces the best fit to observed ENSO statistics. But in a non-linear system such as the delayed oscillator with cubic damping, the system will appear as damped, while in fact it will spontaneously generate oscillations. Other evidence comes from the finding that there is a lack of support for a continuous ENSO cycle in the observations. In particular, there is little observational evidence that the initiation of an ENSO event relies on the memory of a previous event, though the termination of an event is generally consistent with the delayed oscillator mechanism. The break in the cycle suggests that the system is in a damped regime and the onset of ENSO relies on external influences. Other studies dispute the stochastic hypothesis by providing evidence that seems to be more consistent with the self-sustained ENSO theory. As demonstrated in Schopf and Suarez (1988) and discussed in Jin (1997), a system with a stable, periodic oscillation in the absence of noise can become irregular with the addition of stochastic forcing, and will present statistics that appear to be more stable. Chen et al. (2004) provide retrospective forecasts of ENSO over a 148-year period and show that all prominent ENSO events can be re-forecasted at lead-times up to 2 years. Such a long predictability is in better agreement with the self-sustained ENSO theory than the stochastic theory. However it remains to be tested in the crucible of an actual forecast.

What has emerged in the consideration of the theory is the conclusion that noise has a profound influence on the system and that ocean wave dynamics are essential to obtaining predictive skill as is the proper description of the air-sea
coupling. The non-linear, strongly oscillating view gives the most optimistic view of predictability, the stochastic version gives the most pessimistic.

Finally, the debate over where the system lies may have less importance for the practical forecaster than for the theoretician. If weather noise has an influence on the system, there are two parts to consider: what is the role of noise that occurred in the past, and what can we do about the future weather? The past weather noise has become stamped on the ocean and is propagating in the system. If we had a good observing system and initialization method, one could hope to capture all the influences of the past noise, and march forward to a good simulation. This means that we need more than a single simple metric for the ocean initial state. It is insufficient to look at the depth of the thermocline in the west and make a prediction. It will not work to describe the average amplitude of the gravest westward propagating Rossby wave in the ocean. The past noise is inherent in the very complex and complete ocean state, and extracting as much of this as possible is the key job of the data assimilation systems.

If one might hope to capture the effects of past “noise” or weather with a good observing system, what can we do about the weather events that are going to occur over the upcoming seasons that we are attempting to predict? There is evidence that some features, such as the Madden-Julian oscillation, may be able to be predicted for more than a week, but beyond that time, one has to consider these effects as unknowable. It is ultimately these disturbances that will limit the predictability of seasonal means. Perfect models and perfect initialization will never be able to overcome their effect. Experience with idealized model predictability studies seems to show that the limit of predictability is significantly longer than we currently realize with today’s prediction systems. Much work remains to be done, advancing the models and refining the initialization systems.