

CHAPTER 10***The general circulation as turbulence***

Copyright 2003 David A. Randall

10.1 Energy and enstrophy cascades

Consider small-scale three-dimensional motion. The relevant momentum equation is

$$\frac{\partial \mathbf{V}}{\partial t} = -(\mathbf{V} \cdot \nabla) \mathbf{V} - \nabla \left(\frac{\delta p}{\rho_0} \right) + g \frac{\delta \theta}{\theta_0} \mathbf{k} + \nu \nabla^2 \mathbf{V}. \quad (10.1)$$

Here ν is the (molecular) kinematic viscosity, which we assume to be a constant; we have used the Bousinesq approximation, for simplicity; and we have neglected the effects of rotation. We also adopt the Bousinesq form of the continuity equation:

$$\nabla \cdot \mathbf{V} = 0. \quad (10.2)$$

Taking $\nabla \cdot$ (10.1), and using (10.2), we obtain a diagnostic equation for δp :

$$\nabla^2 \left(\frac{\delta p}{\rho_0} \right) = \nabla \cdot \left[-(\mathbf{V} \cdot \nabla) \mathbf{V} + g \frac{\delta \theta}{\theta_0} \mathbf{k} + \nu \nabla^2 \mathbf{V} \right]. \quad (10.3)$$

This shows that δp is entirely determined by the motion. It plays only a *passive* role. We can eliminate it by taking $\nabla \times$ (10.1):

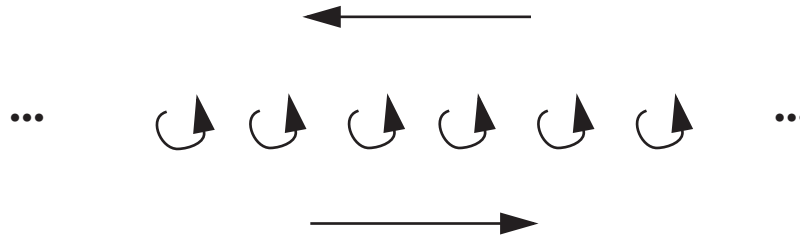
$$\frac{\partial \boldsymbol{\omega}}{\partial t} = -(\mathbf{V} \cdot \nabla) \boldsymbol{\omega} + (\boldsymbol{\omega} \cdot \nabla) \mathbf{V} + \nabla \times \left(g \frac{\delta \theta}{\theta_0} \mathbf{k} \right) + \nu \nabla^2 \boldsymbol{\omega}. \quad (10.4)$$

Here $\boldsymbol{\omega}$ is the three-dimensional vorticity vector, and we have used (10.2).

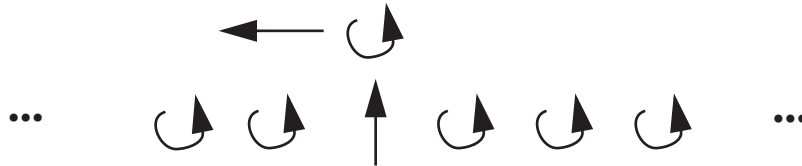
When ν is sufficiently small (i.e. when the Reynolds number, $\frac{VL}{\nu}$, is sufficiently large), the flow described by (10.4) becomes turbulent, due to shearing instability. The mechanism of shearing instability is illustrated in Fig. 10.1. The upper panel of the figure shows a balanced vortex sheet which is considered to extend to infinity in both directions. The sheet is in balance because each vortex is advected up by its neighbor to the left, and down by its neighbor to the right, so that no net vertical motion occurs. If one vortex is perturbed

upward, however, as in the middle panel, it is carried to the left by the combined effects of the vortices left behind. After being displaced to the left, it experiences a net upward advection, away from the sheet. This means that the initial upward perturbation is amplified, and so the balanced sheet is unstable. The tilted lines in the bottom panel show a new shear zone, which is also unstable. After some time, the flow becomes highly disordered. This is the mechanism that leads to turbulence. It can work in either two or three dimensions.

Infinite vortex sheet (balanced)



Disturb one vortex...



This leads to:

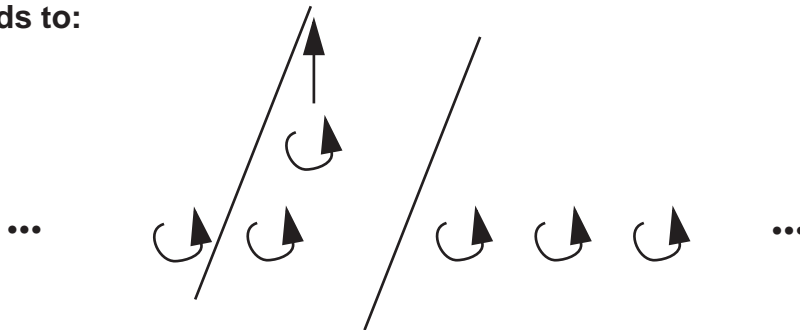


Figure 10.1: The basic mechanism of shearing instability.

The simplified discussion of shearing instability given above refers only to vorticity advection, i.e. to the $-(\mathbf{V} \cdot \nabla)\omega$ term of (10.4). The buoyancy term of (10.4) can make a strong contribution to vorticity production, and so you may wonder whether it represents a second mechanism for the production of turbulence. The answer is “not really;” buoyancy acts as an indirect source of turbulence, but not as a direct source. The buoyancy term of (10.4) is linear; as discussed later, turbulence is intrinsically nonlinear. Buoyancy generates highly organized coherent structures that contain regions of strong shear, e.g., the vortex rings associated with thermals. The shear in these buoyancy-driven structures sets the stage for shearing instability. In short, buoyancy creates conditions in which turbulence can

develop through shearing instability, but buoyancy does not actually produce the turbulence itself. Further discussion of the buoyancy term of (10.4) is outside the scope of this course.

The $(\boldsymbol{\omega} \cdot \nabla) \mathbf{V}$ term of (10.4) represents both stretching and twisting; it vanishes in two-dimensional flows, and is not essential for the generation of turbulence. It can be written as $|\boldsymbol{\omega}| \partial \mathbf{V} / \partial s$, where s is a curvilinear coordinate in the direction of the vorticity vector. Note that the vector velocity appears. We can subdivide $|\boldsymbol{\omega}| \partial \mathbf{V} / \partial s$ into two effects:

- i) Stretching, $\mathbf{e}_s |\boldsymbol{\omega}| \partial v_s / \partial s$, where v_s is the component of \mathbf{V} in the direction of $\boldsymbol{\omega}$ and \mathbf{e}_s is a unit vector parallel to $\boldsymbol{\omega}$; see Fig. 10.2.

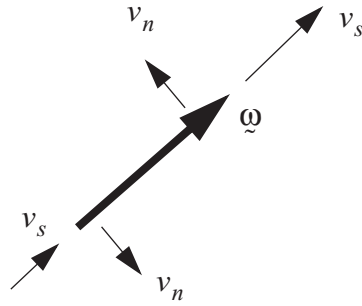


Figure 10.2: The components of the velocity normal and tangent to the vorticity vector, and the stretching and twisting processes associated with these velocity components.

Positive stretching ($\partial v_s / \partial s > 0$) causes $\partial \omega / \partial t > 0$. There is a tendency for the vorticity field and $\partial v_s / \partial s$ to be positively correlated, because viscosity causes convergence into regions of positive vorticity, and divergence from regions of negative vorticity. This will be proven below. A consequence is that stretching causes $|\boldsymbol{\omega}|$ to increase in an average sense. As demonstrated below, we can also say that the stretching term causes the mean of the squared vorticity to increase; the squared vorticity is called the “*enstrophy*.”

- ii) Twisting, $\mathbf{e}_n |\boldsymbol{\omega}| \partial v_n / \partial s$, where v_n is the component of \mathbf{V} normal to the vector $\boldsymbol{\omega}$ and \mathbf{e}_n is a unit vector perpendicular to $\boldsymbol{\omega}$. Twisting changes the direction of $\boldsymbol{\omega}$, but not its magnitude. You should prove this as an exercise.

To derive an equation for the time change of the enstrophy, we dot (10.4) with the vorticity vector, $\boldsymbol{\omega}$:

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{1}{2} |\boldsymbol{\omega}|^2 \right) = & -(\mathbf{V} \cdot \nabla) \left(\frac{1}{2} |\boldsymbol{\omega}|^2 \right) + |\boldsymbol{\omega}|^2 \frac{\partial v_s}{\partial s} + \boldsymbol{\omega} \cdot \left[\nabla \times \left(g \frac{\delta \theta}{\theta_0} \mathbf{k} \right) \right] \\ & + \nu \nabla \cdot [\boldsymbol{\omega} \cdot (\nabla \boldsymbol{\omega})] - \nu [(\nabla \boldsymbol{\omega}) \cdot \nabla] \cdot \boldsymbol{\omega} . \end{aligned} \quad (10.5)$$

Here we have exposed the effects of vortex stretching on the enstrophy, using

$$\boldsymbol{\omega} \cdot [(\boldsymbol{\omega} \cdot \nabla) \mathbf{V}] = |\boldsymbol{\omega}|^2 \frac{\partial v_s}{\partial s} , \quad (10.6)$$

which you should prove as an exercise, and also we have used

$$\begin{aligned} \boldsymbol{\omega} \cdot [\nu \nabla^2 \boldsymbol{\omega}] &= \nu \boldsymbol{\omega} \cdot [\nabla \cdot (\nabla \boldsymbol{\omega})] \\ &= \nu \{ \nabla \cdot [\boldsymbol{\omega} \cdot (\nabla \boldsymbol{\omega})] - [(\nabla \boldsymbol{\omega}) \cdot \nabla] \cdot \boldsymbol{\omega} \} . \end{aligned} \quad (10.7)$$

The term $-\nu [(\nabla \boldsymbol{\omega}) \cdot \nabla] \cdot \boldsymbol{\omega}$ represents enstrophy dissipation, and is always an enstrophy sink.

Now use the continuity equation to rewrite (10.5) in flux form, drop the buoyancy term, and combine the two flux-divergence terms:

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{1}{2} |\boldsymbol{\omega}|^2 \right) = & \nabla \cdot \left\{ -\mathbf{V} \frac{1}{2} |\boldsymbol{\omega}|^2 + \nu [\boldsymbol{\omega} \cdot (\nabla \boldsymbol{\omega})] \right\} + |\boldsymbol{\omega}|^2 \frac{\partial v_s}{\partial s} \\ & - \nu [(\nabla \boldsymbol{\omega}) \cdot \nabla] \cdot \boldsymbol{\omega} + \boldsymbol{\omega} \cdot \left[\nabla \times \left(g \frac{\delta \theta}{\theta_0} \mathbf{k} \right) \right] . \end{aligned} \quad (10.8)$$

In a steady state, the left-hand side of (10.8) vanishes, and when we average over the whole domain the flux divergence terms on the right-hand side vanish as well, so that we are left with

$$\int_M \left[|\boldsymbol{\omega}|^2 \frac{\partial v_s}{\partial s} - \nu [(\nabla \boldsymbol{\omega}) \cdot \nabla] \cdot \boldsymbol{\omega} \right] dM = 0 . \quad (10.9)$$

Because the viscous term of (10.9) is a sink of enstrophy, we can conclude that, on the average, the stretching term has to be a source of enstrophy, i.e.

$$\int_M |\boldsymbol{\omega}|^2 \frac{\partial v_s}{\partial s} dM > 0 . \quad (10.10)$$

Note that this follows from our *assumption* of a steady state. Recall that the rate of production of vorticity by vortex stretching is $|\boldsymbol{\omega}| \partial v_s / \partial s$. Eq. (10.10) implies that in a

statistically steady turbulence there is a tendency for $\partial v_s / \partial s$ to be positive where $|\omega|$ is larger than average, and for $\partial v_s / \partial s$ to be negative where $|\omega|$ is smaller than average. This means that the rich get richer -- strong vortices are stretched and made even stronger.

We conclude that the net effect of $(\omega \bullet \nabla) \mathbf{V}$ is to increase $|\omega|$ in an average sense. The advective terms of the kinetic energy equation do not change the domain-averaged kinetic energy, however. *A process that increases the enstrophy without affecting the kinetic energy tends to shift the spectrum of kinetic energy towards shorter scales.* To see this, note that the ratio of the kinetic energy to the enstrophy has the units of a length squared:

$$\frac{\frac{1}{2}|\mathbf{V}|^2}{|\omega|^2} \sim L^2. \quad (10.11)$$

We can interpret L as the width of the most energetic vortices. Circulations with lots of vorticity per unit kinetic energy have their energy concentrated in relatively small vortices.

When $|\omega|^2$ increases while $\frac{1}{2}|\mathbf{V}|^2$ remains constant, L must decrease. This systematic migration of the kinetic energy from larger to smaller scales, due to the action of the stretching term, is called a kinetic energy “*cascade*”. This term evokes a waterfall in which a single stream falls off a cliff and repeatedly splits on the way down, as the water strikes rocks or other obstructions. The kinetic energy cascade is like a “flow” of kinetic energy from larger scales to smaller scales. Because viscosity acts most effectively on small scales, vortex stretching leads to kinetic energy dissipation.

10.2 Nonlinearity and scale interactions

Scale interactions are intrinsically nonlinear, i.e. they can only arise from the nonlinear terms of the equations, such as the stretching term. To see how this works from a mathematical perspective, consider the following simple example. Suppose that we have two modes given by

$$A(x) = \hat{A}e^{ikx} \text{ and } B(x) = \hat{B}e^{ilx}, \quad (10.12)$$

respectively. Here the wave numbers of modes A and B are denoted by k and l , respectively. If we combine A and B linearly, e.g., form

$$\alpha A + \beta B, \quad (10.13)$$

where α and β are spatially constant coefficients, then no “new” waves are generated; k and l continue to be the only wave numbers present. In contrast, if we multiply A and B together, which is a nonlinear operation, then we generate the new wave number, $k + l$:

$$AB = \hat{A}\hat{B}e^{i(k+l)x}. \quad (10.14)$$

Other nonlinear operations such as division, exponentiation, etc., will also generate new wave numbers.

10.3 Two-dimensional turbulence

In two dimensions, the stretching/twisting term of (10.4) is zero, because $\frac{\partial}{\partial s}$ is zero.

It follows that *vorticity and enstrophy are both conserved under inertial processes in two-dimensional turbulence*. Here the term “inertial processes” refers to advection and rotation, only. Of course, kinetic energy is also conserved under inertial processes. Since both kinetic energy and enstrophy are conserved under inertial processes in two-dimensional flows, L is also conserved. The implication is that *kinetic energy does not cascade in frictionless two-dimensional flows*; it “stays where it is” in wave number space.

When the effects of viscosity are included in two-dimensional turbulence, they act most effectively on the smallest scales, which are the scales on which the enstrophy is concentrated. Friction therefore removes or “dissipates” enstrophy quite effectively at the small-scale end of the spectrum. In order for enstrophy dissipation to continue, additional small-scale enstrophy must be supplied by nonlinear transfer of enstrophy from larger scales to the smaller scales. It is as if the enstrophy dissipation “pulls” enstrophy from larger to smaller scales, through the nonlinear terms. This is an *enstrophy cascade*. Note that L will tend to increase as a result of enstrophy dissipation. This means that the scale of the most energetic eddies will increase with time. This is a *kinetic energy “anti-cascade.”* Further discussion is given below.

We conclude that in three-dimensional flow both kinetic energy and enstrophy cascade and are dissipated, while in a two-dimensional flow enstrophy cascades and is dissipated, but kinetic energy is nearly conserved.

The exchanges of energy and enstrophy among scales in two-dimensional turbulence were studied by Fjortoft (1953), who obtained some very fundamental and famous results, which can be summarized in a simplified way as follows. Consider three equally spaced wave

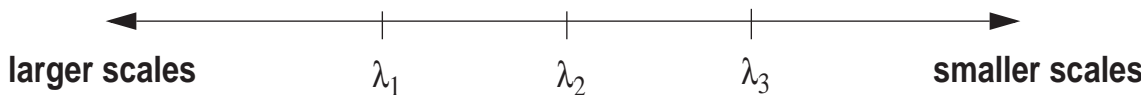


Figure 10.3: Diagram used in the explanation of Fjortoft’s (1953) analysis of the exchanges of energy and enstrophy among differing scales in two-dimensional motion.

numbers, as shown in Fig. 10.3. By “equally spaced” we mean that

$$\lambda_2 - \lambda_1 = \lambda_3 - \lambda_2 \equiv \Delta\lambda. \quad (10.15)$$

The enstrophy, E , is

$$E = E_1 + E_2 + E_3, \quad (10.16)$$

and the kinetic energy is

$$K = K_1 + K_2 + K_3 . \quad (10.17)$$

It can be shown that

$$E_n = \lambda_n^2 K_n , \quad (10.18)$$

where λ_n is a wave number, and the subscript n denotes a particular Fourier component.

Consider an inertial process such that kinetic energy and enstrophy are redistributed in a two-dimensional frictionless flow, i.e.

$$K_n \rightarrow K_n + \delta K_n , \quad (10.19)$$

$$E_n \rightarrow E_n + \delta E_n . \quad (10.20)$$

Because kinetic energy and enstrophy are both conserved under two-dimensional inertial processes, we have

$$\sum \delta K_n = 0 , \quad (10.21)$$

$$\sum \delta E_n = 0 . \quad (10.22)$$

From (10.21) we see

$$\delta K_1 + \delta K_3 = -\delta K_2 , \quad (10.23)$$

Note from (10.18) that

$$\delta E_n = \lambda_n^2 \delta K_n . \quad (10.24)$$

From (10.22) and (10.24), we obtain

$$\begin{aligned} \lambda_1^2 \delta K_1 + \lambda_3^2 \delta K_3 &= -\lambda_2^2 \delta K_2 \\ &= \lambda_2^2 (\delta K_1 + \delta K_3) . \end{aligned} \quad (10.25)$$

Collecting terms in (10.25), we find that

$$\frac{\delta K_3}{\delta K_1} = \frac{\lambda_2^2 - \lambda_1^2}{\lambda_3^2 - \lambda_2^2} . \quad (10.26)$$

Using (10.15), we can simplify (10.26) to

$$\frac{\delta K_3}{\delta K_1} = \frac{\lambda_2 + \lambda_1}{\lambda_3 + \lambda_2} < 1. \quad (10.27)$$

This shows that the energy transferred to higher wave numbers (δK_3) is less than the energy transferred to lower wave numbers (δK_1). This conclusion rests on both (10.21) and (10.22), i.e. on both kinetic energy conservation and enstrophy conservation. The implication is that kinetic energy actually “migrates” from higher wave numbers to lower wave numbers, i.e. from smaller scales to larger scales. This is sometimes called an “anti-cascade.”

We now perform a similar analysis for the enstrophy. As a first step, we note from (10.24) and (10.27) that

$$\begin{aligned} \frac{\delta E_3}{\delta E_1} &= \frac{\lambda_3^2}{\lambda_1^2} \left(\frac{\lambda_2 + \lambda_1}{\lambda_3 + \lambda_2} \right) \\ &= \frac{(\lambda_2 + \Delta\lambda)^2 \left(\lambda_2 - \frac{1}{2}\Delta\lambda \right)}{(\lambda_2 - \Delta\lambda)^2 \left(\lambda_2 + \frac{1}{2}\Delta\lambda \right)}. \end{aligned} \quad (10.28)$$

To show that this ratio is greater than one, we demonstrate that $\frac{\delta E_3}{\delta E_1}$ can be written as $a \cdot b \cdot c$, where a , b , and c are each greater than one. We can choose:

$$a = \frac{\lambda_2 + \Delta\lambda}{\lambda_2 + \frac{1}{2}\Delta\lambda} > 1, \quad b = \frac{\lambda_2 - \frac{1}{2}\Delta\lambda}{\lambda_2 - \Delta\lambda} > 1, \quad c = \frac{\lambda_2 + \Delta\lambda}{\lambda_2 - \Delta\lambda} > 1. \quad (10.29)$$

The conclusion is that enstrophy does cascade to higher wave numbers in two-dimensional turbulence. In the presence of viscosity, such a cascade ultimately leads to enstrophy dissipation.

10.4 *Quasi-two-dimensional turbulence*

Large-scale motions are quasi-two-dimensional, so it is reasonable to suppose that an enstrophy cascade occurs for them just as it does for purely two-dimensional motion. To investigate this, we use the momentum equation in isentropic coordinates:

$$\frac{\partial \mathbf{V}}{\partial t} + \eta \mathbf{k} \times \mathbf{V} + \nabla_\theta (K + M) = 0. \quad (10.30)$$

Here M is the Montgomery Stream Function,

$$\eta \equiv \zeta + f \quad (10.31)$$

is the absolute vorticity, where

$$\zeta \equiv (\nabla_{\theta} \times \mathbf{V}) \cdot \mathbf{k} \quad (10.32)$$

and we assume no heating ($\dot{\theta} = 0$) and no friction. Note that ζ is the vorticity along an isentropic surface. Taking the curl of (10.30) gives

$$\frac{\partial \zeta}{\partial t} + (\mathbf{V} \cdot \nabla_{\theta})\eta = -\eta(\nabla_{\theta} \cdot \mathbf{V}). \quad (10.33)$$

Since f is independent of time, we can write this as

$$\frac{D\eta}{Dt} = -\eta(\nabla_{\theta} \cdot \mathbf{V}). \quad (10.34)$$

Recall that the continuity equation in isentropic coordinates is

$$\frac{\partial m}{\partial t} + \nabla_{\theta} \cdot (m\mathbf{V}) = 0, \quad (10.35)$$

which is equivalent to

$$\frac{Dm}{Dt} = -m(\nabla_{\theta} \cdot \mathbf{V}), \quad (10.36)$$

where $m \equiv -\partial p / \partial \theta$. It is remarkable that (10.34) and (10.36) have exactly the same forms. Mass and absolute vorticity have to abide by the same rules! By combining (10.34) and (10.36), we find that

$$\frac{Dq}{Dt} = 0. \quad (10.37)$$

Here $q \equiv (\zeta + f)/m$ is the Ertel potential vorticity.

Since $m = -\partial p / \partial \theta$, we can think of m as the pressure-thickness or mass per unit area between two isentropic surfaces. In this analogy, we have

$$\frac{D}{Dt} \left(\frac{\eta}{\delta p} \right) = 0. \quad (10.38)$$

This looks very much like

$$\frac{D}{Dt}\left(\frac{\eta}{h}\right) = 0, \quad (10.39)$$

which is the conservation law for potential vorticity in “shallow water” of depth h (see the Appendix on the shallow water equations).

From (10.37) we immediately find that

$$\frac{D}{Dt}(q^2) = 0. \quad (10.40)$$

The quantity q^2 is called the “potential enstrophy.” According to (10.40), the potential enstrophy is conserved following a particle. From (10.40), we can easily show that

$$\overline{mq^2}^\theta = \text{constant}. \quad (10.41)$$

Here $\overline{(\quad)}^\theta$ denotes an average over an isentropic surface.

For a *nondivergent* flow (i.e., nondivergent on an isentropic surface) with no heating, conservation of q implies conservation of η , because (10.36) implies that in such a case m is constant following the motion. It follows that

$$L^2 \equiv \frac{\overline{K}^\theta}{\overline{\eta^2}^\theta} = \text{constant}. \quad (10.42)$$

This is the same conclusion that we reached earlier, for the case of two-dimensional turbulence.

For a more general divergent flow, we can write

$$\overline{\eta^2}^\theta < \overline{\left(\frac{m_{\max}}{m}\right)\eta^2}^\theta = m_{\max} \overline{\left(\frac{\eta^2}{m}\right)}^\theta = m_{\max} \overline{(mq^2)}^\theta = \text{constant}. \quad (10.43)$$

Here m_{\max} is an appropriately chosen constant upper bound on m . According to (10.43)

$\overline{\eta^2}^\theta$ has an upper bound. Then from (10.42) we see that L^2 has a lower bound. This shows that even for divergent motion the kinetic energy cannot cascade into arbitrarily small scales; it tends to stay in the large scales. This is a partial explanation for the “smoothness” of the general circulation.

Charney (1971) showed that even though real atmospheric motions are not two-dimensional, the constraint of quasi-geostrophy causes large-scale geostrophic turbulence to behave much like idealized two-dimensional turbulence, so that potential enstrophy cascades to smaller scales and is dissipated while kinetic energy is conserved (is not dissipated) and

anti-cascades to larger scales.

When kinetic energy is conserved while enstrophy decreases due to dissipation, the effect is that L must increase. (Our conclusion above that L remains constant was based on the assumption that both kinetic energy and enstrophy are invariant.) Because the total amount of “room” available to the eddies is fixed (the planet is not getting any bigger), the only way that L can increase is if the “number of eddies” decreases; for this reason, “eddy mergers” tend to occur in two-dimensional or geostrophic turbulence.

In an idealized numerical simulation, McWilliams (1984) showed that through vortex mergers pure two-dimensional turbulence gradually organizes itself into a single large cyclone-anticyclone pair. His results are shown in Fig. 10.4.

As mentioned earlier, it has been hypothesized that in a real atmosphere, the cyclonic member of the pair would be killed off by convective mixing, so that only the anticyclonic member would survive. The anticyclonic Great Red Spot on Jupiter, and similar phenomena observed elsewhere on Jupiter and the other giant gaseous planets of the outer solar system, may be the end products of a kinetic energy anti-cascade in quasi-two-dimensional turbulence.

10.5 Dimensional analysis of the kinetic energy spectrum

We now turn to an analysis of the distribution of kinetic energy with scale, i.e. the kinetic energy “spectrum”. Consider a three-dimensional flow. Let ε be the rate of kinetic energy dissipation per unit mass, and ν the kinematic viscosity:

$$\begin{aligned}\varepsilon &\sim L^2 T^{-3}, \\ \nu &\sim L^2 T^{-1}.\end{aligned}\tag{10.44}$$

Kolmogorov (1941) hypothesized that:

- 1) For locally homogeneous and isotropic three-dimensional turbulence, the turbulence statistics are determined by only two dimensional parameters: ε and ν .
- 2) There exists an inertial subrange within which energy is neither generated nor dissipated, but just “passes through,” like a little town on a big highway, where nobody stops except perhaps to buy gas. The inertial subrange consists of a range of scales — a portion of wave-number space. Within the inertial subrange, the turbulence statistics are determined, as functions of wave number, by only one dimensional parameter that characterizes the particular flow in question: the dissipation rate, ε . Note that ε represents dissipation that occurs on scales outside (shorter than those of) the inertial range.

The *viscous subrange* consists of scales for which both ε and ν are important. Together, the viscous subrange and the inertial subrange make up the homogeneous isotropic

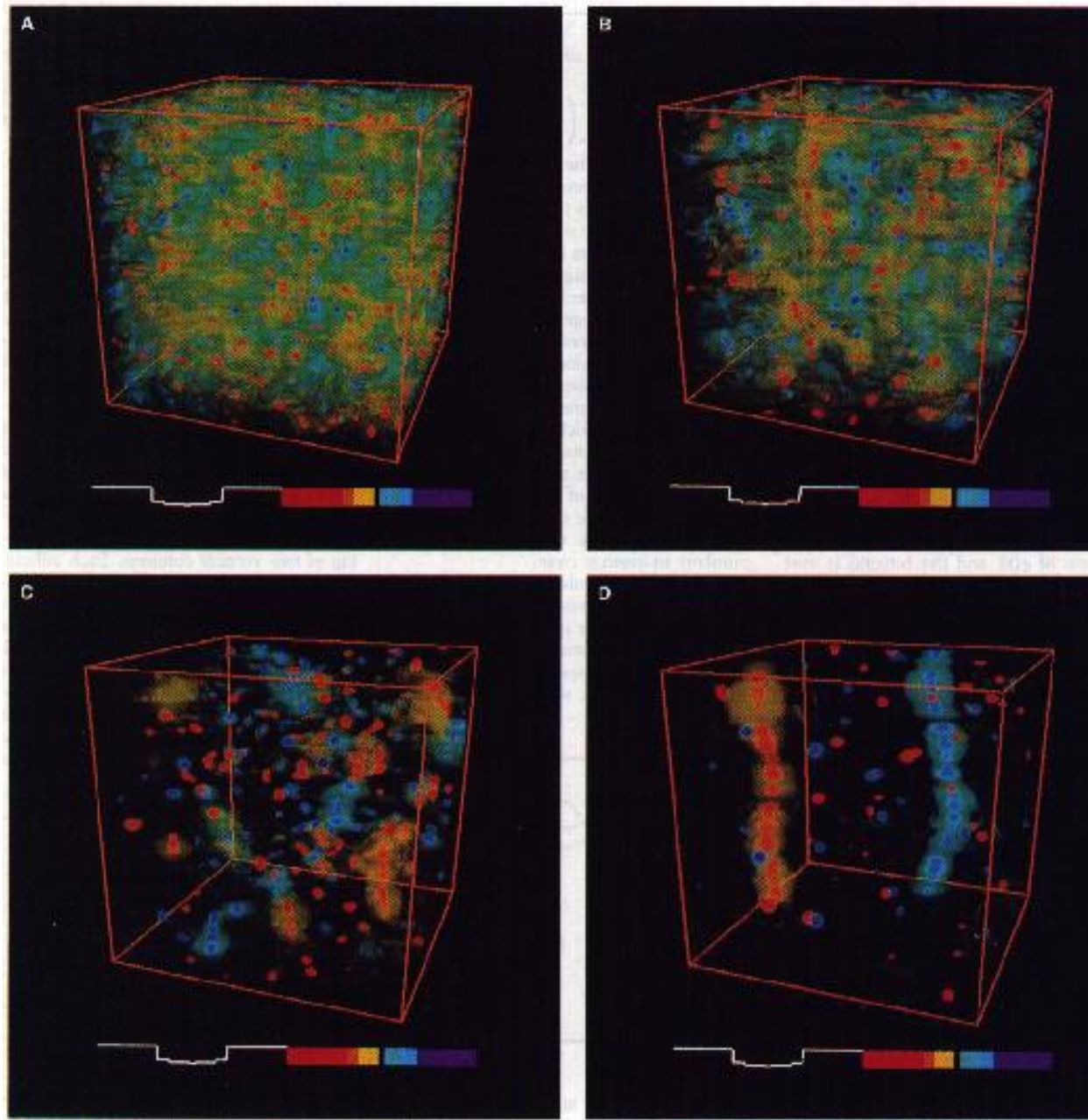


Figure 10.4: A time sequence of the vorticity distribution in an idealized numerical simulation of quasi-two-dimensional turbulence, by McWilliams (1984). In panel A the vorticity is highly disorganized. Through a process of enstrophy dissipation and vortex mergers, the vorticity gradually organizes itself into a single pair of vortices -- one cyclonic and one anti-cyclonic.

component of the turbulence. The smallest scale in the inertial subrange, which is the same as the largest scale in the viscous subrange, can be estimated by forming a length from ν and ϵ :

$$\lambda_k = \left(\frac{v^3}{\varepsilon} \right)^{\frac{1}{4}}. \quad (10.45)$$

This is called the Kolmogorov microscale. The inertial subrange consists of homogeneous isotropic eddies that are larger than λ_k . If ε increases for a given v , then λ_k has to become smaller. For the Earth's atmosphere, a typical value of λ_k is in the range 10^{-2} m to 10^{-3} m.

In the Earth's atmosphere, kinetic energy is typically generated at lower wave numbers, and migrates to higher wave numbers by way of the inertial subrange, and is finally dissipated at the smallest scales. This is illustrated in Fig. 10.5. We want to find $K(k)$, the

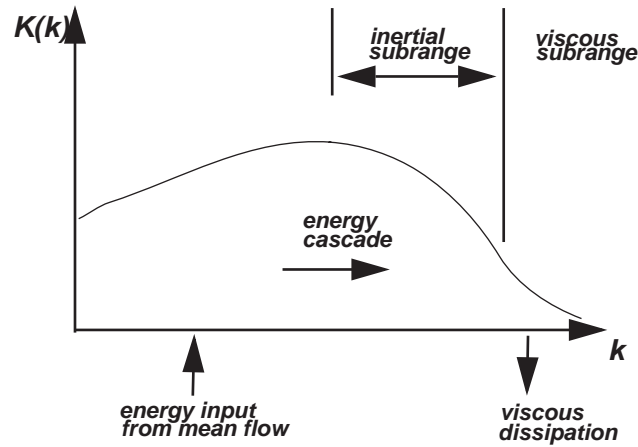


Figure 10.5: Sketch illustrating the flow of kinetic energy through wave-number space, from a low-wave number region with an energy source, to a high-wave number region in which dissipation removes kinetic energy.

spectrum of energy for an inertial subrange. This is essentially the Fourier transform of the spatial distribution of energy. It has units of energy per unit mass per unit wave number. Dimensionally,

$$K(k) \sim L^3 T^{-2}. \quad (10.46)$$

Since the turbulence statistics in the inertial subrange depend only on ε , $K(k)$ can depend only on ε and k . Assuming that

$$K(k) \sim \alpha \varepsilon^m k^n, \text{ where } \alpha = \text{nondimensional constant}, \quad (10.47)$$

we find $m = 2/3$, $n = -5/3$. In other words,

$$K(k) \sim \alpha \varepsilon^{\frac{2}{3}} k^{-\frac{5}{3}}. \quad (10.48)$$

This conclusion, due to Kolmogorov, is well supported by measurements.

We know that kinetic energy does not cascade in two-dimensional turbulence, but we have seen that enstrophy does cascade. Suppose that in two-dimensional turbulence there exists an inertial subrange, presumably on rather large scales, in which *enstrophy* is neither generated nor dissipated, and that in this inertial subrange the turbulence statistics are determined by the rate of *enstrophy* dissipation. We can show by methods similar to those used to derive (10.48) that the kinetic *energy* spectrum follows k^{-3} . This conclusion is supported by numerical simulations of two-dimensional turbulence. Note that k^{-3} falls off more steeply with increasing k than does $k^{-5/3}$. This means that there is less kinetic energy at high wave numbers in two-dimensional turbulence than in three-dimensional turbulence. This is consistent with our earlier conclusion that kinetic energy dissipation is weak in two-dimensional turbulence and strong in three-dimensional turbulence.

An analysis very similar to that used to derive (10.48) can be used to determine the kinetic energy spectrum for scales longer than those at which kinetic energy is generated. We assume that in an inertial subrange upscale from the kinetic energy source, the kinetic energy spectrum depends only on the strength of the source and the wave number. The conclusion is that upscale from the energy source, the kinetic energy spectrum follows $k^{-5/3}$. This is true for either two-dimensional or three-dimensional turbulence.

We can now make a more refined version of Fig. 10.5, in which we indicate both k_B and k_C . This is Fig. 10.6, in which four distinct inertial subranges are indicated. The “low”

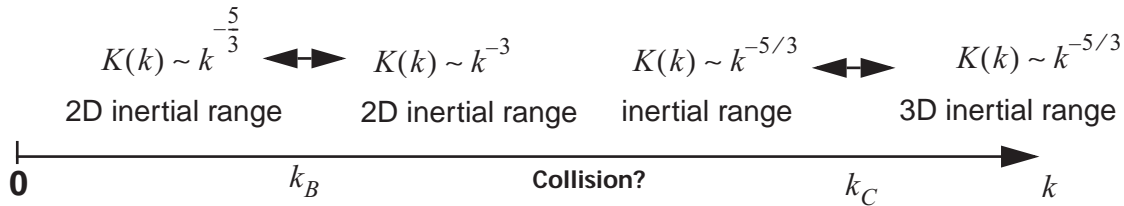


Figure 10.6: Schematic indicating the kinetic energy spectrum in the Earth's atmosphere, as implied by dimensional analysis. Baroclinic instability adds energy at wave number k_B , and convection adds energy at wave number k_C . See text for details.

wave number k_B denotes the scale at which baroclinic instability acts as a kinetic energy source. The “high” wave number k_C denotes the scale at which convection acts as a kinetic energy source. For $k > k_C$, we have a three-dimensional inertial range with $K(k) \sim k^{-5/3}$, and an energy cascade. For k less than k_B , energy flows upscale, away from the baroclinic energy source. For k slightly greater than k_B , we have a two-dimensional inertial subrange with $K(k) \sim k^{-3}$ and energy flowing down-scale. For k slightly less than k_C , we have a

three-dimensional inertial range with $K(k) \sim k^{-5/3}$ and energy flowing upscale. Clearly, the shape of the spectrum has to change somewhere between k_B and k_C .

Lilly (1998) discussed observations that seem to depict such a change in the spectral slope, as shown in Fig. 10.7. The spectrum follows $k^{-5/3}$ for scales less than about 100 km,

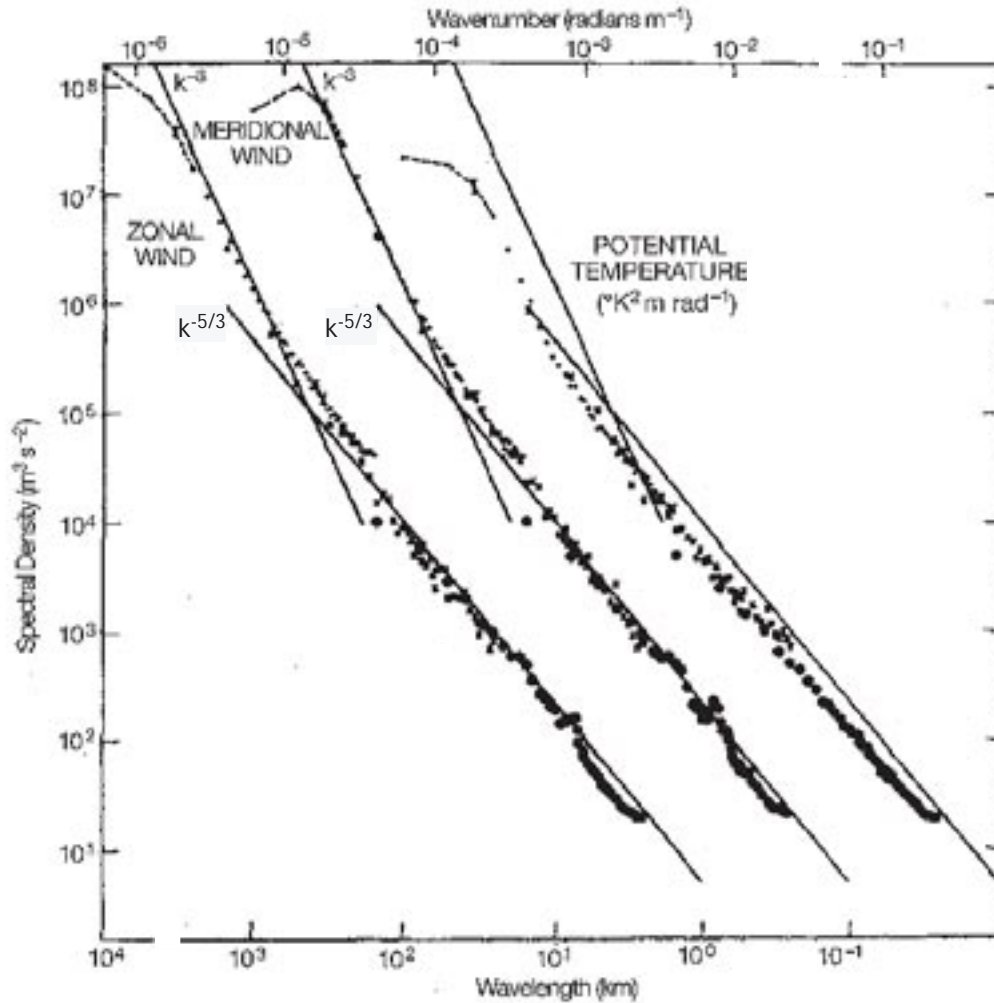


Figure 10.7: Spectra of zonal and meridional winds and potential temperature near the tropopause, obtained from aircraft data over many parts of the world. The meridional wind spectrum is displaced one decade to the right, and the potential temperature spectrum two decades. From Lilly

and follows k^{-3} for scales greater than about 100 km. There is a “kink” in the spectrum close to 100 km. A second kink is visible at a scale of several thousand kilometers. The correspondence between Fig. 10.6 and Fig. 10.7 should be clear.

Consider a planet on which baroclinic instability is inactive, perhaps because solar heating is weak and internal heat sources are dominant, primarily driving small-scale convection. In such a case, k_B drops out of the problem, and we might expect to see $K(k) \sim k^{-5/3}$ all the way from k_C up to near planetary scales. Jupiter's atmosphere may work something like this.

There is one more complication. Rhines (1975) pointed out that sufficiently large vortices will feel the β effect, which exerts a "restoring force" opposing large meridional excursions by the particles making up the vortex. This means that the β effect will tend to resist meridional widening of vortices beyond some limit. Rhines suggested that the eddies will in fact begin to behave as Rossby-wave packets for scales large enough so that the characteristic eddy velocity, U , is comparable to the phase speed of a Rossby wave. Dimensional analysis suggests that this occurs for

$$k \sim k_\beta \equiv \sqrt{\frac{\beta}{U}}. \quad (10.49)$$

This scale is sometimes called the "Rhines barrier." For $k < k_\beta$, further meridional broadening is resisted by β , while longitudinal broadening can continue. The eddies therefore become elongated in the zonal direction and ultimately give rise to alternating zonal jets of width k_β^{-1} , like those seen on Jupiter. In recent years there have been a number of numerical modeling studies which tend to support this idea (e.g. Huang and Robinson, 1998).

In summary, vorticity and enstrophy are conserved in two-dimensional flow but not in three-dimensional flow. Kinetic energy is conserved under inertial processes in both two-dimensional and three-dimensional flows. Because both energy and enstrophy are conserved in two-dimensional flows, a two-dimensional motion field "has fewer options" than do three-dimensional flows. Because kinetic energy does not cascade in two-dimensional flow, the motion remains smooth and is dominated by "large" eddies.

10.6 Observations of the kinetic energy spectrum

Boer and Shepherd (1983) analyzed observations to examine the spectra of kinetic energy, enstrophy, and available potential energy, and also the exchanges of kinetic energy among various scales. Following a suggestion of Baer (1972), they used the two-dimensional index associated with the spherical harmonics as a measure of scale, much as Blackmon did in the work described in Chapter 7. The vertically integrated spectra obtained by Boer and Shepherd are shown in Fig. 10.8, for kinetic energy and enstrophy only. The slope of the kinetic energy spectrum is plotted as a function of height in Fig. 10.9. A k^{-3} behavior is evident, particularly at the upper levels. Boer and Shepherd evaluated the exchanges of kinetic energy among the various scales, as shown in Fig. 10.10; similar computations were reported by Chen and Wiin-Nielsen (1978). The smaller scales generally experience a kinetic energy cascade towards even smaller scales, as would be expected for three-dimensional turbulence, but the larger scales experience an inverse cascade, as would be expected for two-dimensional turbulence.

It has also been suggested that the formation and maintenance of a blocking high

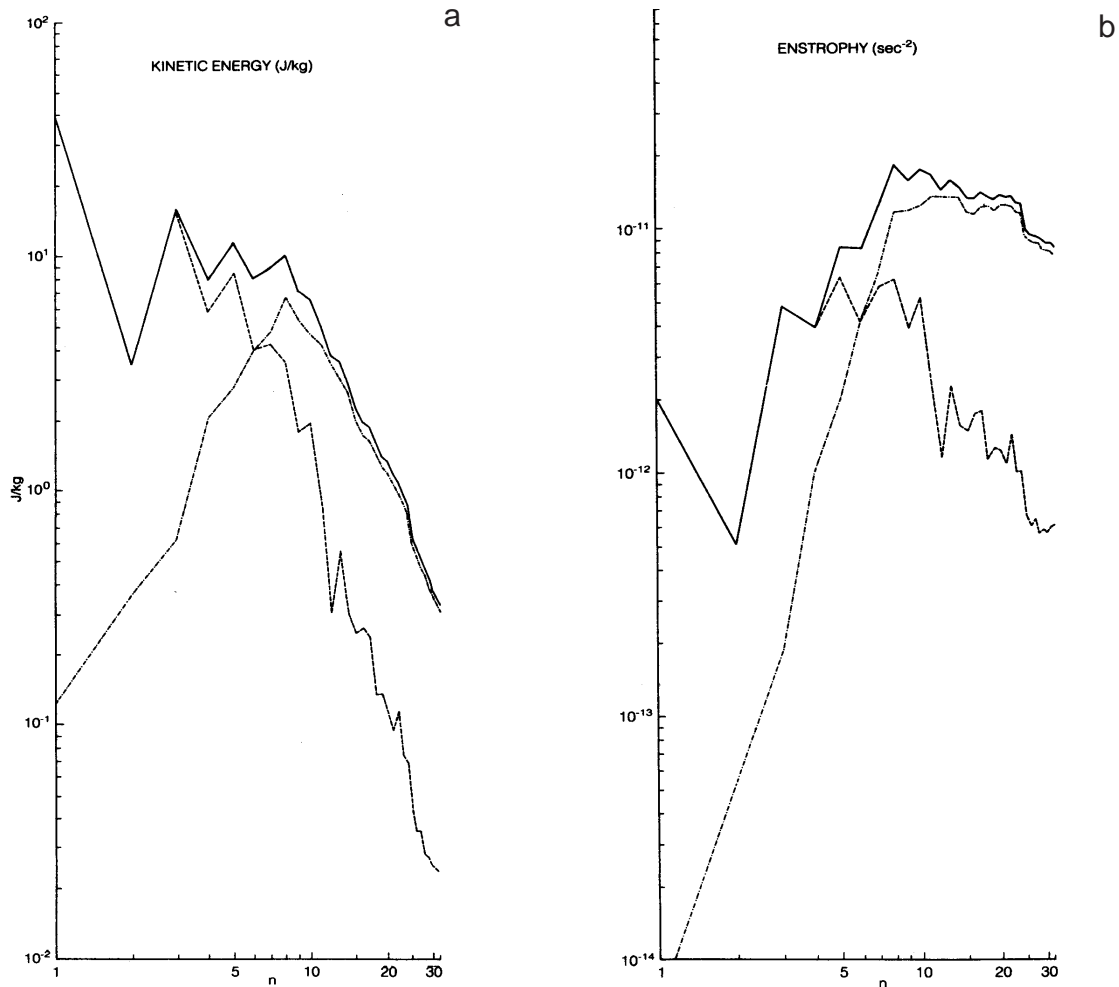


Figure 10.8: The observed vertically integrated spectra of kinetic energy (panel a) and enstrophy (panel b). The solid lines denote total, the dashed stationary, and the dashed-dot transient. From Boer and Shepherd (1983).

represents an example of up-scale energy transfer. As mentioned in Chapter 8, the formation of a block is associated with the advection of subtropical potential vorticity into middle latitudes. The “agent” that carries out this advection is a rapidly developing cyclone, e.g. over the Gulf Stream (Hoskins et al. 1985). Moreover, it is the interaction of the block with small-scale eddies that allows the block to maintain itself for an extended period (Shutts, 1986).

10.7 The general circulation as a blender

10.7.1 What does the blender blend?

To the extent that the eddies of the general circulation act like turbulence, it is natural to ask whether they “mix” things, and if so, which things. A variable that is mixed by turbulence is a conservative variable. Linear momentum is not very well conserved; it is subject to a variety of non-conservative effects, including the coriolis acceleration and pressure gradients. The angular momentum $M \equiv a \cos \phi (u + \Omega a \cos \phi)$, first discussed in

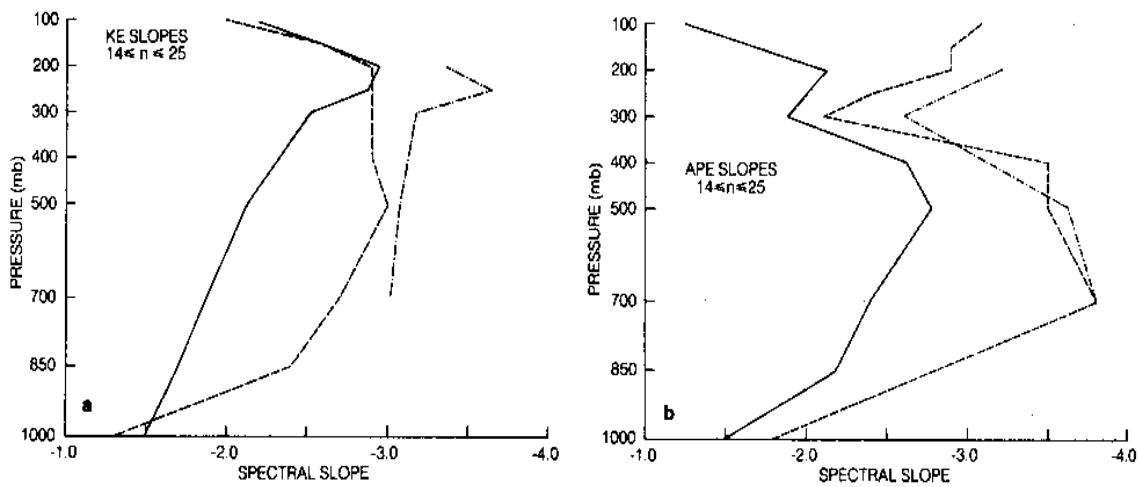


Figure 10.9: Slope of straight line fitted to the kinetic energy spectrum, for two-dimensional indices in the range 14 to 25. The dash-dotted line represents the results of Baer (1972), and the dashed line shows the results of Chen and Wiin-Nielsen (1978). From Boer and Shepherd (1983).

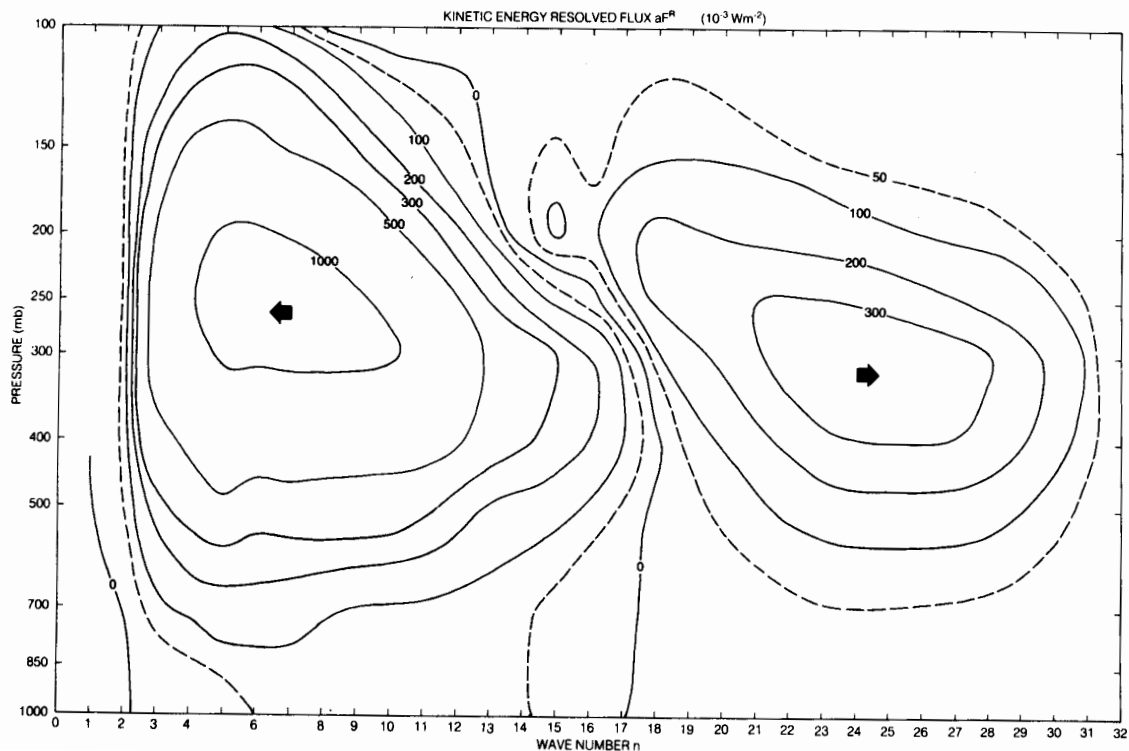


Figure 10.10: Observed nonlinear kinetic energy exchanges as a function of height. From Boer and Shepherd (1983).

Chapter 2, is somewhat more conservative, because it is not affected by the coriolis acceleration; a little thought shows that mixing of angular momentum is impossible, however, because a finite angular momentum at the pole would imply infinite zonal wind and vorticity there.

We are thus led to look for evidence that the *eddies mix the Ertel potential vorticity*, which is a very conservative dynamical variable. In the absence of heating, particles stay on their isentropic surfaces, so we would expect the large-scale turbulence to homogenize potential vorticity (and other conservative variables) along isentropic surfaces. Fig. 10.11, from Sun and Lindzen (1994), provides evidence that this does happen.

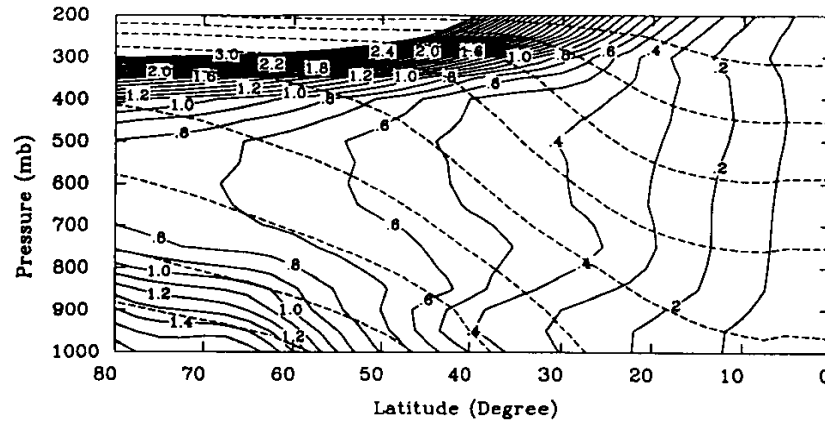


Figure 10.11: Potential vorticity (solid lines) and potential temperature (dashed lines) plotted as functions of pressure and latitude. From Sun and Lindzen (1994).

10.7.2 Dissipating enstrophy but not kinetic energy

Sadourny and Basdevant (1985) suggested an interesting approach to representing the effects of quasi-two-dimensional, geostrophic turbulence in the momentum equation. The issue is that in such a flow enstrophy is dissipated but kinetic energy is not. How can we formulate a “frictional” term in the momentum equation that has this property? To examine the idea of Sadourny and Basdevant, we start from the equation of motion in “invariant” form, using isentropic coordinates:

$$\frac{\partial \mathbf{V}}{\partial t} + q \mathbf{k} \times (m \mathbf{V}) + \nabla \left(\frac{1}{2} \mathbf{V} \cdot \mathbf{V} + M \right) = 0. \quad (10.50)$$

Here q is the potential vorticity, as before. We want to include friction in such a way that it does not affect the kinetic energy. This can be done by introducing a parameter, D , as follows:

$$\frac{\partial \mathbf{V}}{\partial t} + (q - D) \mathbf{k} \times (m \mathbf{V}) + \nabla \left(\frac{1}{2} \mathbf{V} \cdot \mathbf{V} + M \right) = 0 \quad (10.51)$$

We can interpret $(q - D)$ as a modified potential vorticity. When we dot (10.51) with \mathbf{V} to form the kinetic energy equation, the term involving D drops out, *regardless of the form of*

D . This means that D does not contribute to the tendency of the kinetic energy.

We want to choose the form of D in such a way that potential enstrophy is dissipated. The first step is to construct the potential vorticity equation, by taking the curl of (10.51) and using continuity:

$$\frac{Dq}{Dt} = \frac{1}{m} \nabla \bullet (Dm \mathbf{V}) . \quad (10.52)$$

The right-hand side of (10.52) vanishes, as it should, for $D \equiv 0$. Let the potential enstrophy averaged over an entire isentropic surface (globally) be given by

$$Z(\theta) = \frac{1}{S} \iint_S \frac{q^2}{2} m \, dS . \quad (10.53)$$

Then (10.52) implies that

$$\frac{dZ(\theta)}{dt} = - \iint_S (D_m \mathbf{V} \bullet \nabla q) dS . \quad (10.54)$$

To guarantee dissipation of $Z(\theta)$, we choose

$$D = \tau \mathbf{V} \bullet \nabla q , \quad (10.55)$$

with $\tau \geq 0$. By substituting (10.55) into (10.54), we find that

$$\frac{dZ(\theta)}{dt} = -\tau \iint_S (m \mathbf{V} \bullet \nabla q)^2 dS . \quad (10.56)$$

This guarantees that $Z(\theta)$ decreases with time for $\tau > 0$. We can recover potential vorticity conservation and potential enstrophy conservation by setting $\tau = 0$. From (10.55), we see that

$$q - D = q - \tau (\mathbf{V} \bullet \nabla q) \equiv q_{\text{anticipated}} . \quad (10.57)$$

Here $q_{\text{anticipated}}$ can be interpreted as the value of q that we expect or *anticipate* by looking upstream to see what value of q is being advected towards us. Eq. (10.57) is equivalent to

$$\frac{q_{\text{anticipated}} - q}{\tau} = -\mathbf{V} \bullet \nabla q . \quad (10.58)$$

For this reason, the idea is referred to as the “anticipated potential vorticity method” to parameterize the effects of geostrophic turbulence in the momentum equation. Sadourny and

Basdevant (1985) showed that this approach gives realistic kinetic energy and enstrophy spectra in a numerical model.

10.7.3 The Gent-McWilliams theory of tracer transports along isentropic surfaces

In isentropic coordinates, the conservation of an arbitrary tracer of mixing ratio τ is expressed by

$$\left(\frac{\partial}{\partial t}\right)_{\theta} (m\tau) + \nabla_{\theta} \cdot (m\mathbf{V}\tau) + \frac{\partial}{\partial \theta} (m\dot{\theta}\tau) = mS(\tau), \quad (10.59)$$

$S(\tau)$ is the source or sink of tracer per unit mass and per unit time. Putting $\tau = 1$ and $S(1) = 0$, we recover the continuity equation:

$$\left(\frac{\partial}{\partial t}\right)_{\theta} m + \nabla_{\theta} \cdot (m\mathbf{V}) + \frac{\partial}{\partial \theta} (m\dot{\theta}) = 0. \quad (10.60)$$

In the absence of heating, (10.60) reduces to

$$\left(\frac{\partial}{\partial t}\right)_{\theta} m + \nabla_{\theta} \cdot (m\mathbf{V}) = 0. \quad (10.61)$$

We can write the mass flux $m\mathbf{V}$ as the sum of its zonal mean and the departure from the zonal mean:

$$\begin{aligned} m\mathbf{V} &= [m\mathbf{V}] + (m\mathbf{V})^* \\ &= [m][\mathbf{V}] + [m^* \mathbf{V}^*] + (m\mathbf{V})^*. \end{aligned} \quad (10.62)$$

Here we have used

$$[m\mathbf{V}] = [m][\mathbf{V}] + [m^* \mathbf{V}^*], \quad (10.63)$$

which is similar to a relation discussed way back in Chapter 2. Recall that $[m^* \mathbf{V}^*]$ is called the “bolus mass flux,” and $\frac{[m^* \mathbf{V}^*]}{[m]}$ is called the “bolus velocity.” In a similar way, the flux of a tracer τ can be written as

$$\begin{aligned} m\mathbf{V}\tau &= ([m\mathbf{V}] + (m\mathbf{V})^*)([\tau] + \tau^*) \\ &= [m\mathbf{V}][\tau] + [m\mathbf{V}]\tau^* + (m\mathbf{V})^*[\tau] + (m\mathbf{V})^*\tau^*. \end{aligned} \quad (10.64)$$

Zonal averaging gives

$$[m\mathbf{V}\tau] = [m\mathbf{V}][\tau] + [(m\mathbf{V})^* \tau^*]. \quad (10.65)$$

Substituting from (10.63), we obtain

$$[m\mathbf{V}\tau] = [m][\mathbf{V}][\tau] + [m^* \mathbf{V}^*][\tau] + [(m\mathbf{V})^* \tau^*]. \quad (10.66)$$

This shows that *the mean value of the tracer, $[\tau]$, is advected by both the mean mass flux and the bolus mass flux*. The third term can be interpreted as “diffusion,” because the mass flux involved, i.e. $(m\mathbf{V})^*$, has zero mean. Substituting (10.66) into the zonal mean of (10.60), we obtain

$$\begin{aligned} \left(\frac{\partial}{\partial t}\right)_\theta [m\tau] + \frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} \{([m][v] + [m^* v^*])[\tau] \cos \varphi\} + \frac{\partial}{\partial \theta} [m\dot{\theta}\tau] \\ = [mS(\tau)] - \frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} ([m\mathbf{V})^* \tau^*] \cos \varphi \quad . \end{aligned} \quad (10.67)$$

Similarly, the zonally averaged continuity equation can be written as

$$\left(\frac{\partial}{\partial t}\right)_\theta [m] + \frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} \{([m][v] + [m^* v^*]) \cos \varphi\} + \frac{\partial}{\partial \theta} [m\dot{\theta}] = 0. \quad (10.68)$$

In order to make use of these ideas, we have to determine $[m^* v^*]$. Gent and McWilliams (1990) suggested

$$[m^* v^*] = -\frac{\partial}{\partial \theta} \left(\frac{\kappa}{a} \frac{\partial}{\partial \varphi} [p] \right), \quad (10.69)$$

where κ is a non-negative parameter. Substitution of (10.69) into (10.68) gives

$$\begin{aligned} \left(\frac{\partial}{\partial t}\right)_\theta [m] + \frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} ([m][v] \cos \varphi) + \frac{\partial}{\partial \theta} [m\dot{\theta}] \\ = \frac{1}{a^2 \cos \varphi} \frac{\partial}{\partial \varphi} \left\{ \frac{\partial}{\partial \theta} \left(\kappa \frac{\partial}{\partial \varphi} [p] \right) \cos \varphi \right\}. \end{aligned} \quad (10.70)$$

In case κ is spatially constant, this reduces to

$$\begin{aligned}
& \left(\frac{\partial}{\partial t} \right)_{\theta} [m] + \frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} ([m][v] \cos \varphi) + \frac{\partial}{\partial \theta} [m\dot{\theta}] \\
& = \frac{\kappa}{a^2 \cos \varphi} \frac{\partial}{\partial \varphi} \left\{ \left(\frac{\partial}{\partial \varphi} \frac{\partial}{\partial \theta} [p] \right) \cos \varphi \right\} .
\end{aligned} \tag{10.71}$$

Since

$$[m] = -\frac{1}{g} \frac{\partial}{\partial \theta} [p], \tag{10.72}$$

this is equivalent to

$$\begin{aligned}
& \left(\frac{\partial}{\partial t} \right)_{\theta} [m] + \frac{1}{a \cos \varphi} \frac{\partial}{\partial \varphi} ([m][v] \cos \varphi) + \frac{\partial}{\partial \theta} [m\dot{\theta}] \\
& = \frac{-g\kappa}{a^2 \cos \varphi} \frac{\partial}{\partial \varphi} \left\{ \left(\frac{\partial}{\partial \varphi} [m] \right) \cos \varphi \right\} .
\end{aligned} \tag{10.73}$$

This equation describes “diffusion of $[m]$,” it shows that the effect of (10.69) is to drive $[m]$ towards uniformity on θ -surfaces, which implies driving $[p]$ towards uniformity on θ -surfaces. Recall that a state in which $[p]$ is uniform along θ -surfaces is one in which the available potential energy is zero. This means that *the effect of (10.69) is to remove available potential energy from the system*. An interpretation is that the available potential energy has been converted into subgrid-scale kinetic energy, through subgrid-scale baroclinic instability.

The theory outlined above, due to Gent and McWilliams (1990), is being used to parameterize eddy-transport in ocean models, with good results.

10.8 The limits of deterministic weather prediction

pre•dict \pri-'dikt\ vb

[L praedictus, pp. of praedicere, fr. prae- pre- + dicere to say — more at DICTION]

vt

(1632)

:to declare in advance; esp: foretell on the basis of observation, experience, or scientific reason

~ vi

:to make a prediction

When we make a weather forecast, we are solving an “initial value problem.” The current state of the atmosphere is the “initial condition.” The governing equations are integrated forward in time to predict the future state of the atmosphere.

Is it possible in principle to make a perfect (or arbitrarily accurate) forecast? Broadly speaking, there are three sources of forecast error:

1) Properties of the atmosphere itself

- The Uncertainty Principle of quantum mechanics (not important for this problem)
- Nonlinearity and instability (*very* important). As discussed below, these lead to sensitive dependence on initial conditions.

2) The observing system

- Imperfect measurements of the initial conditions (e.g., imperfect thermometers)
- Imperfect spatial coverage of the initial conditions
- Mistakes

3) The model

- Wrong equations
- Imperfect resolution

We focus on the first and most fundamental source of error, i.e., properties of the atmosphere itself that limit predictability.

Lorenz (1963) discussed the predictability of a “deterministic nonperiodic flow.” A system is said to be *deterministic* if its future evolution is completely determined by a set of rules. Models of the atmosphere (e.g. the primitive equations) are examples of sets of rules. The atmosphere is, therefore, a deterministic system. The behavior of the atmosphere is obviously nonperiodic; its previous history is not repeated. The predictability of periodic flows is a rather boring subject. If the behavior of the atmosphere was periodic, the weather would certainly be predictable!

How does nonperiodic behavior arise? The forcing of the atmosphere by the seasonal and diurnal cycles is at least approximately periodic. For linear systems, periodic forcing always leads to a periodic response. For non-linear systems, however, periodic forcing can lead to a nonperiodic response. *Nonperiodic behavior arises from nonlinearity.*

Lorenz (1963) studied an idealized set of nonlinear convection equations and found that for some values of the parameters all of the steady and periodic solutions are unstable. The model exhibits nonperiodic solutions. The equations of the model are remarkably simple:

$$\begin{aligned}
 \dot{X} &= -\sigma X + \sigma Y, \\
 \dot{Y} &= -XZ + rX - Y, \\
 \dot{Z} &= XY - bZ.
 \end{aligned}
 \tag{10.74}$$

Here

$$\sigma = 10, b = \frac{8}{3}, \text{ and } r = 24.74 \tag{10.75}$$

are parameters. The numerical values given in (10.75) are particular choices (not unique ones) that lead to nonperiodic behavior. A solution of (10.74) is shown in Fig. 10.12. The state of the model is plotted in a phase space. Most of the time the solution is near one of two “attractors,” i.e. at a randomly chosen point in the integration the probability of finding the solution near one of the attractors is very high. Occasionally the solution wanders from one attractor to the other. Partly because of the appearance of this plot, the solution is sometimes called the “Butterfly Attractor.”

This example illustrates the important point that even a *simple* nonlinear system can be unpredictable. Lack of predictability and complex behavior are not necessarily due to complexity in the definition of the system itself.

Here is why there is a finite limit to deterministic predictability. Two slightly different

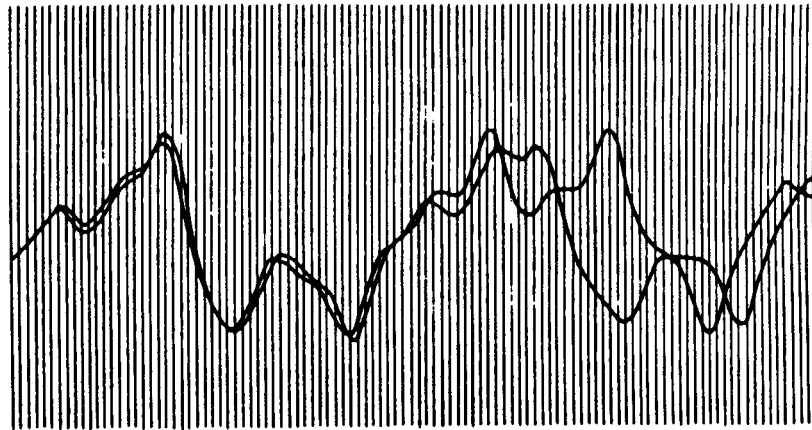


Figure 10.13:How two weather patterns diverge. From nearly the same starting point, Lorenz noticed that two numerical solutions grew farther and farther apart until all resemblance disappeared. From Gleick (1987).

states of the atmosphere diverge from each other with time because the atmosphere is *unstable*. See the sketch in Fig. 10.13. This leads to *sensitive dependence on the initial conditions, on a scale-by-scale basis*. Systems that exhibit sensitive dependence on initial conditions are called “chaotic.” The sensitive dependence of the state of the atmosphere on its past history suggests that the flap of a butterfly’s wings in China could noticeably change the weather in North America a few days later. This is a second reason for calling Fig. 10.13 the

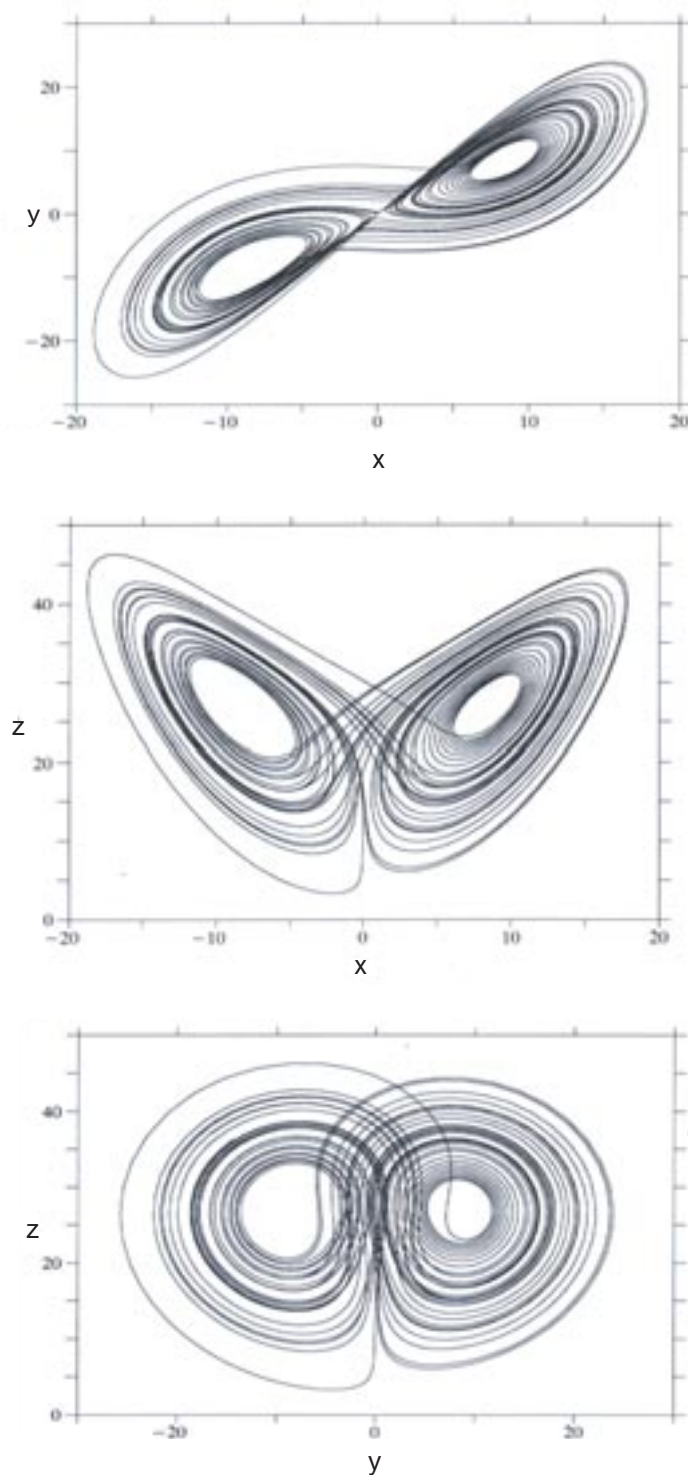


Figure 10.12: The Butterfly Attractor of Lorenz (1963), obtained as the solution of (10.74).
From Drazin (1992).

Butterfly Attractor.

There are many kinds of instability, acting on virtually all spatial scales. Small-scale shearing instabilities act on scales of meters or less. Buoyant instabilities, including cumulus instability, occur primarily on scales of a few hundred meters to a few kilometers. Baroclinic instability occurs on scales of thousands of kilometers.

Although Lorenz discovered the importance of sensitive dependence on initial conditions, he was not the first to discover it; Poincaré (1912) recognized the phenomenon, and even discussed the fact that it makes long-term weather forecasting impossible. James Clerk Maxwell was also aware, during the 19th century, that as a result of instabilities, deterministic physical laws do not necessarily permit deterministic predictions (Harman, 1998, pp. 206-208). Lorenz was the first to realize that the phenomenon of sensitive dependence on initial conditions permits complex unpredictable behavior to occur even in very simple systems. He emphasized the importance of nonlinearity, in addition to instability.

Recall that small-scale eddies produce fluxes that modify larger scales. In this way, errors on small scales can produce errors on larger scales through *nonlinear* processes, as discussed in Chapter 7. Recall that scales cannot interact in linear systems.

We conclude that it is the combination of instability and nonlinearity that limits our ability to make skillful forecasts of the largest scales of motion. This idea is illustrated in Fig. 10.14. Both instability and nonlinearity are properties of the atmosphere itself; we cannot

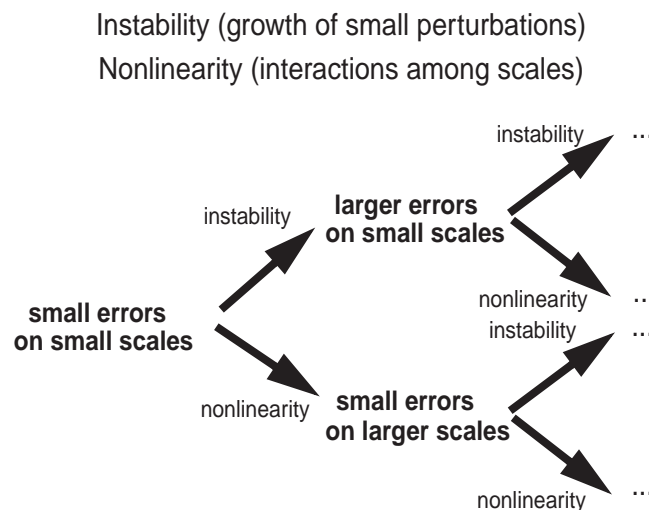


Figure 10.14: Sketch illustrating the role of instability in leading to error growth, and of nonlinearity in leading to the movement of error from small scales to larger scales.

make them “go away” by improving our models or our observing systems. It is the properties of the atmosphere itself that lead to an *intrinsic limit of deterministic predictability*. We can say that sensitive dependence on initial conditions imposes a “predictability time” or “predictability limit” which is a limit beyond which the weather is unpredictable in principle. *This limit is a property of the atmosphere itself, and so it cannot be circumvented no matter what method we use to make our forecast; it is not a limitation that is intrinsic to numerical weather prediction or any other specific forecast technique. It applies to all methods.*

Errors on smaller scales double (grow proportionately) faster than errors on larger scales, simply because the intrinsic time scales of smaller-scale circulations are shorter. For example, the intrinsic time scale of a buoyant thermal in the boundary layer might be on the order of 20 minutes, that of a thunderstorm circulation might be on the order of one hour, that of a baroclinic eddy might be two or three days, and that of planetary wave number one might be one to two weeks. Because the predictability time is related to the eddy turn-over time, the predictability limit is a function of scale; larger scales are generally more predictable than smaller scales.

When we eliminate errors on smaller spatial scales by adding more observations, the range of our skillful forecasts is increased by a time increment approximately equal to the predictability time of the newly resolved smaller scales. Pushing the initial error down to smaller and smaller spatial scales is, therefore, a strategy for forecast improvement that yields *diminishing returns*. See Fig. 10.15.

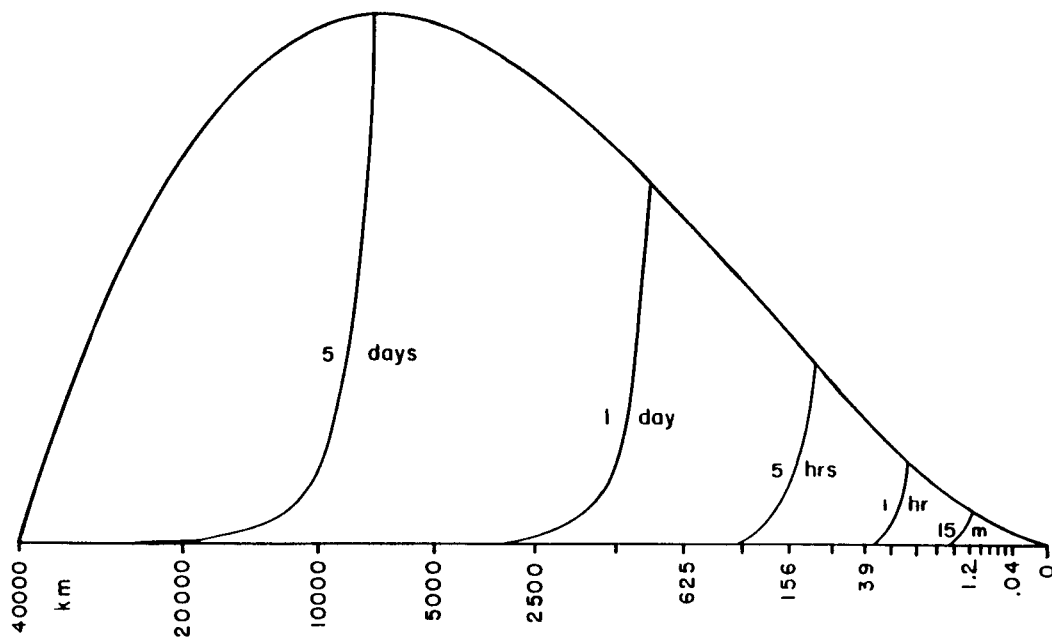


Figure 10.15: A simplified depiction of the energy spectrum $E(k)$ (upper curve), and the error-energy spectra (lower curves) at 15 minutes, 1 hour, 5 hours, 1 day, and 5 days, as interpolated from the result of a numerical study. The lower curves coincide with the upper curve, to the right of their intersections with the upper curve. Areas are proportional to energy. From Lorenz (1969).

As discussed below, estimates show that small errors on the smallest spatial scales can grow in both amplitude and scale to significantly contaminate the largest scales (comparable to the radius of the Earth) in about two to three weeks. Some aspects of atmospheric behavior may nevertheless be predictable on longer time scales. This is particularly true if they are forced by *slowly changing external influences*. An obvious example is the seasonal cycle. Another example is the statistical character of the weather anomalies associated with long-lasting sea surface temperature anomalies, such as those due to El Niño. This point will be discussed further, later in this chapter.

At this point, we can offer a definition of turbulence: *A circulation is said to be turbulent if its predictability time is shorter than the time scale of interest.* For example, baroclinic storms in midlatitude winter can be considered as turbulent eddies if we are interested in seasonal time scales, but they behave as highly predictable, orderly circulations if we are doing a one-day forecast. With this definition, turbulence is in the eye of the beholder.

10.9 Quantifying the limits of predictability

Three approaches to determine the limits of predictability were discussed by Lorenz (1969):

10.9.1 The dynamical approach

In this approach, two or more model solutions are produced, starting from similar but not quite identical initial conditions. This procedure is similar to what Lorenz did by accident, when he discovered sensitive dependence on initial conditions. The model is put in place of the atmosphere; no real data is used. Problems with this approach are: 1) truncation error; 2) imperfect equations; 3) lack of information about very small scales. Studies of this type suggest that the doubling time for small errors with spatial scales of a few hundred km is about 5 days. This implies that the limit of predictability is about two weeks.

One of the earliest examples of the dynamical approach is the study of Charney et al. (1965). They used several atmospheric general circulation models to study the growth of small

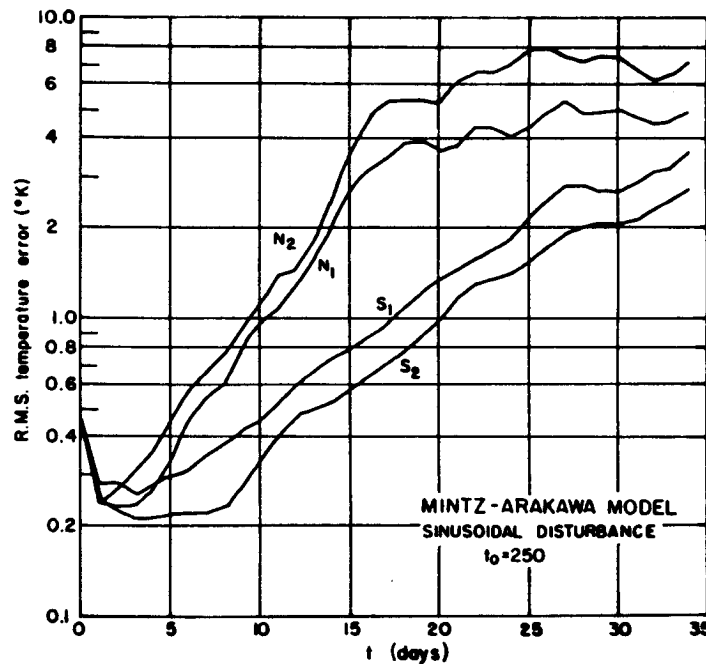


Figure 10.16: Root mean-square temperature error in January simulations performed with the two-level Mintz-Arakawa model. "N" and "S" denote the Northern and Southern Hemispheres, respectively. The subscripts "1" and "2" denote the two model levels. From Charney et al. (1966).

perturbations. Fig. 10.16 shows some of their results, obtained with an early version of the

UCLA general circulation model (GCM). The rms temperature error grows in both hemispheres, but more rapidly in the winter hemisphere, where the circulation is more unstable. The error does not continue to grow indefinitely; it stops growing at the point where the “forecast” is no better than a guess. We say that the error has “saturated.” A suitable guess might consist of a state of the system chosen at random from a very long record of such states; think of pulling a weather map at random out of a huge file full of many such maps.

A more modern example of the dynamical approach is was discussed by Shukla (1981, 1985). Fig. 10.17 is taken from the paper of Shukla (1985). The results shown in the figure are based on computations with a GCM. The model is used to perform multiple simulations, differing only through very small perturbations of the initial conditions. The results given in the two panels on the left side of the figure show the growth and saturation of errors in the sea level pressure for winter and summer. The winter errors grow more rapidly than the summer ones. The errors are large in middle latitudes, where baroclinic instability is active, especially in the winter. The errors are much smaller in the tropics.

Recall, however, that the tropical sea level pressure normally does not vary much. This means that a small error in the tropics can be important. To take this into account, the two panels on the right show the errors normalized by the temporal standard deviation of the sea level pressure. From this perspective, we see that the tropical errors actually grow more rapidly than those of middle latitudes, and saturate at about the same (normalized) values. Note that at all latitudes the saturation values are close to one. This means that the errors stop growing when they become as large as the standard deviation.

Fig. 10.18 is taken from Shukla (1981). For numerical experiments on the growth of errors, similar to those discussed above, the figure shows how error growth varies with zonal wave number. In each panel, the solid curve shows the growth of error in the numerical model, and the dashed curve shows the corresponding growth of error when the “forecast” used is simply persistence. We can say that the model forecast is no longer skillful when it is no better than a forecast based on persistence. After 30 days, the numerical model still has some skill at the lower wave numbers. For higher wave numbers, the model’s skill disappears more quickly.

10.9.2 The empirical approach

In the empirical approach, the atmosphere is put in place of a model; the atmosphere itself is used to predict the atmosphere. Lorenz (1969) examined the observational records of the 200, 500, and 850 mb heights, for the Northern Hemisphere only. He searched for pairs of similar states, or “analogs,” occurring within one month of the same day of the year. He chose 30 December 63 and 13 January 65 as the best available analogs within a five-year record.

The results of the empirical approach show that the doubling time for errors on the smallest resolved scales is less than eight days. They are thus reasonably consistent with the results of the dynamical approach.

Among the problems with the empirical approach are: 1) it is hard to find “good” analogs, in that the smallest error is about 1/2 of the average error; 2) we cannot experiment with the initial error, because we have to take what nature gives us; and 3) the data cannot be used to study the growth of errors on very small scales, simply because such scales are not adequately observed.

10.9.3 The dynamical-empirical approach

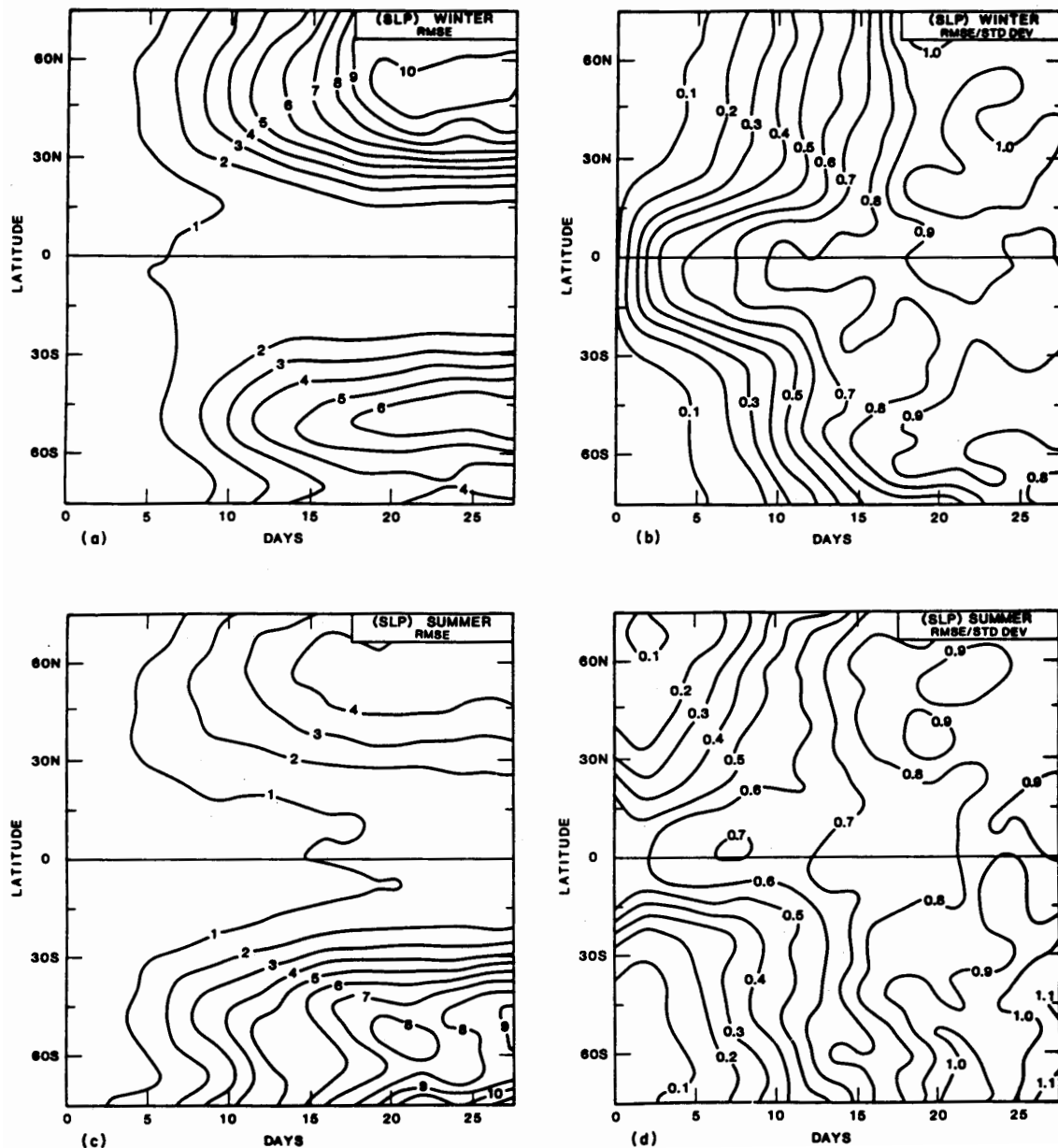


Figure 10.17: Zonal average of root-mean-square error (RMSE; left panels) and ratio (RMSE/STD; right panels) of root mean square and standard deviation of daily values for sea-level pressure. (a) RMSE and (b) RMSE/STD for six pairs of control and perturbation runs during winter; (c) RMSE and (d) RMSE/STD for three pairs of runs during summer. From Shukla (1985).

The “dynamical-empirical” approach is the most difficult of the three to understand. The basic idea, as first conceived by Lorenz (1969), is to derive an equation for the time change of “error kinetic energy” (using a model). We then Fourier transform the error energy equation, so that the spectrum of the kinetic energy appears. The key step is to prescribe this kinetic energy spectrum from observations, down to very small scales (~ 40 m). We then use

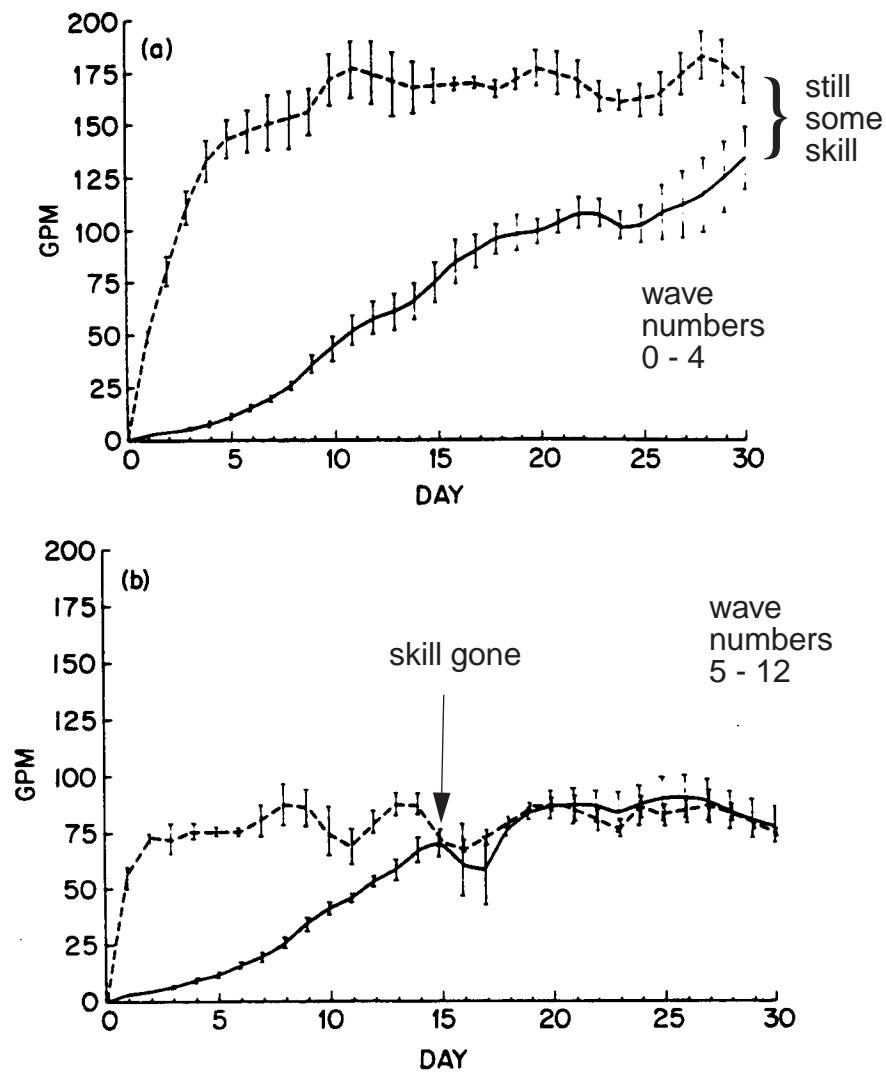


Figure 10.18: Root-mean-square error (solid line) averaged for six pairs of control and perturbation runs and averaged for latitude belt 40–60°N for 500-mb height for (a) wave numbers 0–4 and (b) wave numbers 5–12. Dashed line is the persistence error averaged for the three control run. Vertical bars denote the standard deviation of the error values. [From Shukla (1981).]

the resulting semi-empirical equation to draw conclusions about error growth. The dynamical-empirical approach shows that errors in the smallest scales amplify most quickly and soon dominate. Again, the dynamical-empirical approach suggests that the limit of deterministic predictability is about two to three weeks.

A modern example of the dynamical-empirical approach was described by Lorenz (1982). Suppose that we make a large number of one-day and two-day forecasts, as shown in Fig. 10.19. Let $z_{i,1}$ be the 1-day forecast for z on day i , and let $z_{i,2}$ be the 2-day forecast for z on day i (the same day). Let $E_{1,2}$ be the root-mean-square (rms) difference between the 1-

day and 2-day forecasts *for the same day*, averaged over the globe and over all verification days:

$$E_{1,2}^2 = N^{-1} S^{-1} \sum_{i=1}^N \int_S (z_{i,1} - z_{i,2})^2 dS. \quad (10.76)$$

Here the integral is over the area, S , and the sum is over the verification days, which are distinguished by subscript i . We can interpret $E_{1,2}$ as the average or typical *growth of the forecast error* between the first and second days of the forecasts. If all forecasts were perfect, $E_{1,2}$ would be zero. If the 1-day forecasts were perfect but the two-day forecasts were not, $E_{1,2}$ would be positive, etc.

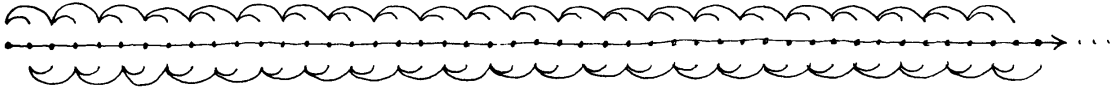


Figure 10.19: A long sequence of one-day and two-day forecasts.

More generally, we can compute the rms error growth between the j -day forecast and the k -day forecast, still verifying on the same day (i.e. day i), from

$$E_{j,k}^2 = N^{-1} S^{-1} \sum_{i=1}^N \int_S (z_{i,j} - z_{i,k})^2 dS. \quad (10.77)$$

We assume without loss of generality that $k \geq j$. It should be clear that $E_{j,k}$ compares j -day forecasts with k -day forecasts. If $E_{j,k} > 0$, the implication is that k -day forecasts are less skillful, on the average, than j -day forecasts. Note that $E_{0,k}$ compares “0-day forecasts,” which are actually analyses rather than forecasts, with k -day forecasts. If the k -day forecasts were perfect, $E_{0,k}$ would be zero. Because of the intrinsic limit of predictability and the inevitable small errors of the initial conditions, however, even a perfect model will give $E_{0,k} > 0$, for $k > 0$. Also, $E_{0,k}$ increases initially as k increases, but then it “saturates” when the k -day forecast becomes no better than a guess, as shown in Fig. 10.20. Even if the model's simulated climate is perfect, $E_{0,k}$ will be different from zero because of the limit of deterministic predictability.

As k becomes large, we expect $E_{j,k}$ to saturate, no matter what the value of j is. For $k \rightarrow \infty$, you might expect that $E_{j,k}$ “should be” independent of j . *It is not, at least not for “small” j (≤ 20).* The reason is that the model's climate is different from the real climate. Since the model is started from real data, a j -day forecast looks like the real world when j is

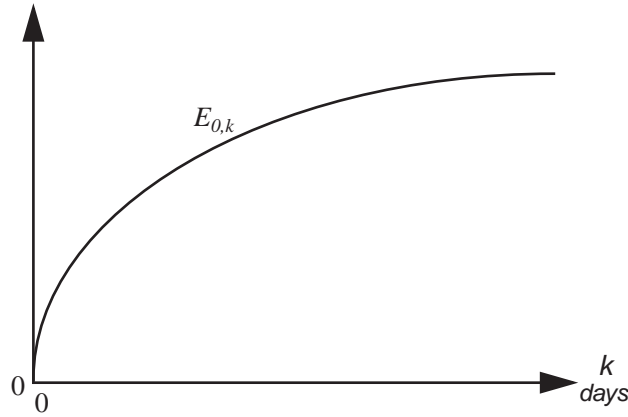


Figure 10.20: The growth of $E_{0,k}$ with time. Rapid initial growth is followed by saturation.

small. As j increases, the model goes to its own climate, and so the forecast increasingly departs from the ensemble of real-world states. Note, however, that $E_{j,k}$ becomes independent of j when both j and k are large, because then we are comparing the model's climate with itself.

As a special case, illustrated in Fig. 10.21, $E_{k-1,k}$ compares $(k-1)$ -day forecasts with k -day forecasts for the same day. For example, putting $k = 10$, $E_{9,10}$ compares nine-day forecasts with ten-day forecasts for the same day. Of course, nine-day forecasts and ten-day forecasts are both pretty bad, and so generally speaking they will be quite different from each other, even though they are supposed to represent the weather on the same day. For this reason, as $k \rightarrow \infty$, $E_{k-1,k}$ is not small. Instead, as $k \rightarrow \infty$, we expect $E_{j,k}$ to approach a constant, i.e., to become independent of the actual values of j and k , so long as $j \neq k$. We can say that $\lim_{k \rightarrow \infty} E_{j,k}$ (for $j \neq k$ and $j \neq 0$) measures the variability of z in the model's climate.

For a perfect model, we would have

$$\lim_{k \rightarrow \infty} E_{0,k} = \lim_{k \rightarrow \infty} E_{k-1,k}, \quad (10.78)$$

because the model's climate would be identical to the true climate. For an imperfect model, on the other hand, we expect

$$\lim_{k \rightarrow \infty} E_{0,k} > \lim_{k \rightarrow \infty} E_{k-1,k}, \quad (10.79)$$

because the model's climate differs from the true climate. This is a key point.

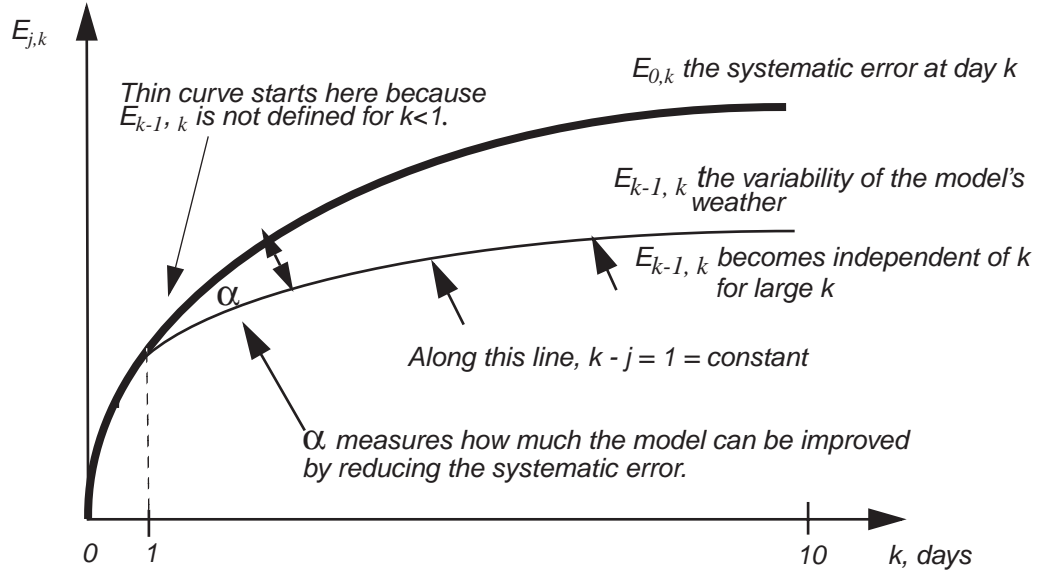


Figure 10.21: The growth of $E_{0,k}$ and $E_{k-1,k}$ with time.

To explore these ideas with real data, Lorenz chose 1 Dec 80 - 10 March 1981 (100 days) as the *verification* dates. Because he worked with 100 verification days, $N = 100$ in (10.77). He used an archive of forecasts and analyses performed with the ECMWF model. Since ECMWF routinely makes ten-day forecasts, he had 100 one-day forecasts, 100 two-day forecasts, etc., out to 100 ten-day forecasts. He plotted $E_{j,k}$ as shown in Fig. 10.22. Each point represents an average over 100 pairs of forecasts.

The thin curves in Fig. 10.22 appear to run nearly parallel to each other, in the sense that $\frac{dE}{dk} = f(E)$. In other words, for any given value of E the thin curves all have about the same slope. This means that the rate of error growth is strongly influenced by the size of the error; of course we know that this cannot be the only factor involved, but it can be the dominant one.

Lorenz wanted to estimate the growth rate of very small errors in forecasts performed with a perfect model. To do this, he essentially fit a curve to the error growth rates in the ECMWF forecasts. He hypothesized that, *for a perfect model*

$$\frac{dE}{dk} = aE - bE^2. \quad (10.80)$$

Here $E \equiv E_{j,k}$, where $k - j = \text{constant}$, and so $\frac{dE}{dk} \equiv (E_{j+1, k+1} - E_{j,k})$ per day. Notice that the second index minus the first index is the same for both E 's, so we are following the “thin curves.” Eq. (10.80) says that the rate of error growth is determined by the size of the error. For small E , the exponential growth rate is a . For larger E , saturation occurs so that

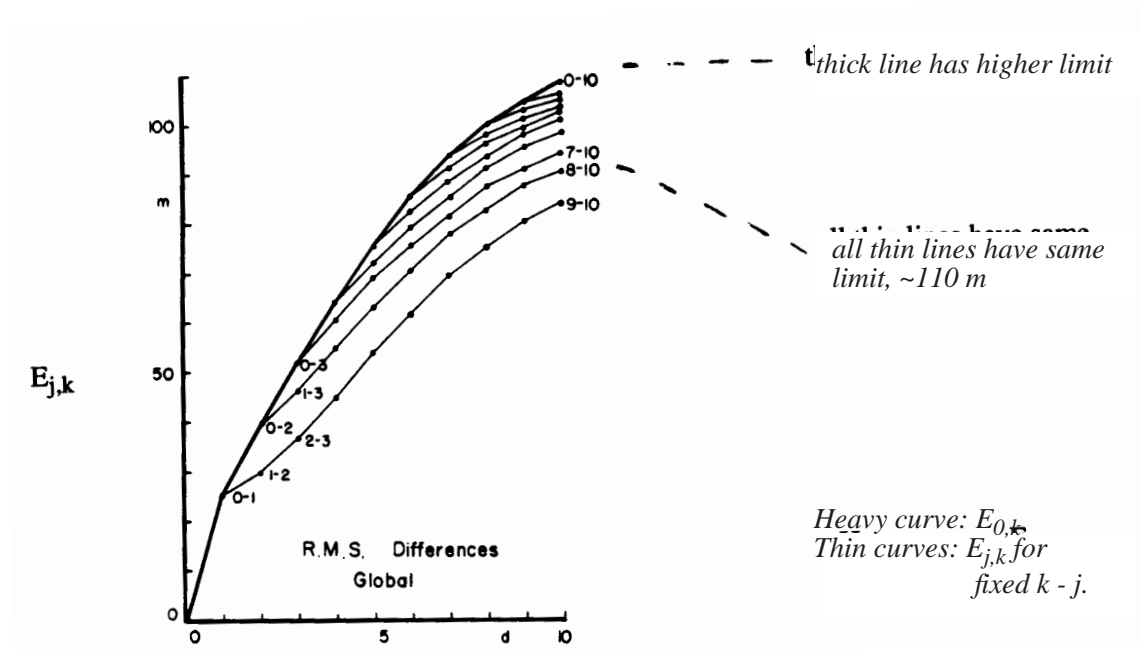


Figure 10.22: Global root-mean-square 500-mb height differences $E_{j,k}$, in meters, between j -day and k -day forecasts made by the ECMWF operational model for the same day, for $j < k$, plotted against k . Values of (j, k) are shown beside some of the points. Heavy curve connects values of $E_{0,k}$. The thin curves connect values of $E_{j,k}$ for constant $k-j$. From Lorenz (1982).

$$E_{sat} = \frac{a}{b}. \quad (10.81)$$

This means that a/b is the maximum error for the perfect model. Note from (10.80) that if the initial error is zero, there will be no error growth. That is what we expect from a “perfect model.”

We want to deduce a , the growth rate of *small errors*, from the data, even though the data do not contain truly small errors. We can do this by fitting the function $f(E) = aE - bE^2$ against $\frac{dE}{dk}$ as evaluated from the data. See Fig. 10.23. In performing this curve fit, we find the “best” values of a and b . This calculation can be interpreted as an example of the dynamical-empirical approach to determine the growth rate of small errors, because both a model and data have been used.

Fig. 10.24 shows the same data as Fig. 10.22, plotted in a different way. The dots correspond to the thin curves in Fig. 10.22, and so represent the “perfect model.” The crosses correspond to the thick curve in Fig. 10.22, and so represent the “imperfect model.” The

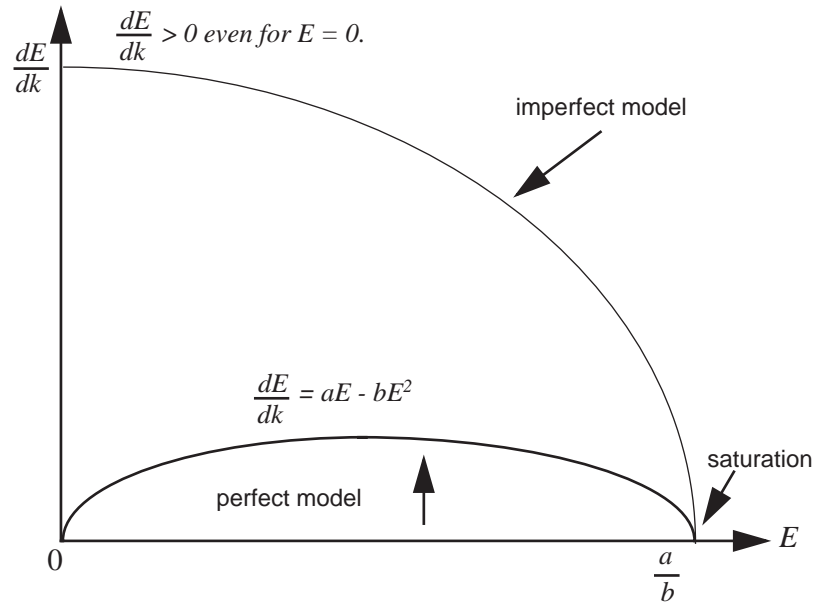


Figure 10.23:Diagram illustrating the expected variation of the error growth rate $\frac{dE}{dk}$ with the magnitude of the error, E .

conclusion of this exercise is that the doubling time for small errors on the smallest scales resolved by the model is 2.4 days.

In a later publication, Lorenz discussed a later set of more skillful forecasts made with an improved model and started from better observed initial conditions. He found that the heavy curve Fig. 10.22 came down, but the thin curves also came down. The gap between them was cut in half. The downward shift of the heavy curve, by itself, suggests that the newer model's simulated climate was more realistic than that of the older model. Alternatively, it could mean that the analyses are more realistic. The downward shift of the thin lines suggest that the newer model's weather is less active than that of the old model.

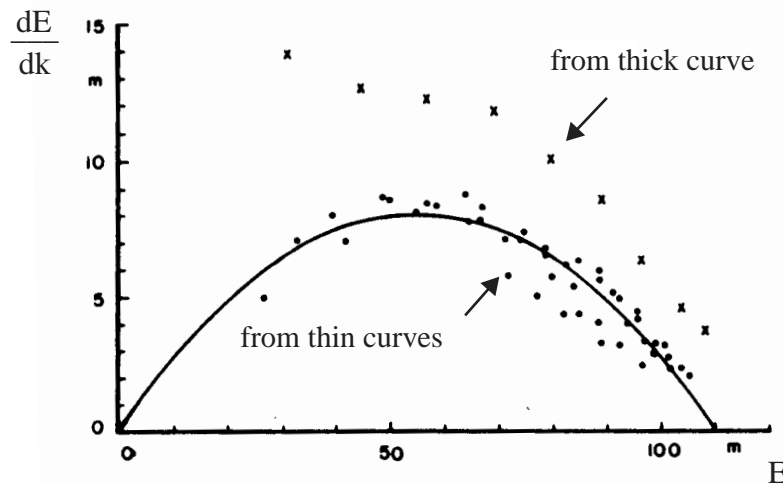


Figure 10.24: Increases in global root-mean-square 500-mb height differences, $E_{j+1,k+1} - E_{jk}$, plotted against average height differences $(E_{j+1,k+1} + E_{jk})/2$, in meters, for each one-day segment of each thin curve in Fig. 1 (large dots), and increases $E_{0,k+1} - E_{0k}$ plotted against average differences $(E_{0,k+1} + E_{0k})/2$, for each one-day segment of heavy curve in Fig. 1 (crosses). Parabola of “best fit” to large dots is shown.

10.10 Climate prediction

cli•mate \ˈkli-met-n

ME climat, fr. MF, fr. LL climat-, clima, fr. Gk klimat-, klima inclination, latitude, climate, fr. klinein to lean — more at LEAN]

(14c)

1: a region of the earth having specified climatic conditions

2a: the average course or condition of the weather at a place over a period of years as exhibited by temperature, wind velocity, and precipitation

b: the prevailing set of conditions (as of temperature and humidity) indoors (a climate-controlled office)

3: the prevailing influence or environmental conditions characterizing a group or period:

We can define “weather” as the instantaneous distribution of the atmospheric state variables, and “climate” as the long-term statistical properties of the same variables. The ocean, land-surface, cryosphere, and biosphere influence the climate. Together with the atmosphere, they form the climate system. All parts of the climate system interact, on sufficiently long time scales.

Is the climate state unique and deterministic, for a given set of “external parameters” such as the solar constant and the Earth’s present orbital configuration? A system possessing a single stable climate for a given set of external parameters is called “*transitive*.” Here the “external parameters” are any factors that influence the climate but are not normally considered part of the climate system. Examples are the rate of energy output of the sun, and the angular velocity of the Earth’s rotation. Some parameters, like the composition of “dry air,” may or may not be considered external.

A system possessing more than one stable climate for a given set of external parameters is called “*intransitive*.” For a given set of external parameters, an intransitive system can assume any of N possible states, where $N > 1$, and will remain in the same state indefinitely unless forced out by some *external* cause.

A system that exhibits occasional *unforced* transitions between two or more distinct climates is called “*almost intransitive*.” It may exhibit multiple climates if finite time-averaging intervals are used, but, an almost-intransitive system has only one climate if an indefinitely long averaging interval is used.

If long-range weather prediction is impossible, how can climate prediction be contemplated at all? If all memory of the initial conditions is “forgotten,” what is the point of solving an initial value problem? Two factors have the potential to make seasonal forecasting and/or climate change prediction possible. First, the system has components with very long memories, including especially the ocean. At present, however, we lack the observations needed to fully initialize these components of the system.

Second, the system responds in systematic and statistically predictable ways to changes in the external forcing. *This means that it is possible to make a prediction without solving an initial-value problem!* A good example is the seasonal cycle. If we *predict* systematic differences in weather between summer and winter in Fort Collins, is that prediction based on solution of an initial value problem? Suppose that the Sun’s energy output decreased significantly. If we *predict* that a general cooling of the Earth’s climate would ensue, is that prediction based on solution of an initial value problem?

These examples illustrate that predictions can be based on knowledge of an external forcing (e.g., the seasonal cycle, or a change in solar output) that changes “slowly” and in a predictable way. Here we have in mind that the forcing changes slowly enough so that the climate system has time to adjust to the variations.

We can thus distinguish two types of climate prediction.

- *Prediction of the equilibrated state (after transient adjustment period) that results from a steady forcing anomaly.* This is not an initial value problem.
- *Prediction of the transient adjustment period itself, as a response to steady or time-varying forcing.* This may or may not be an initial value problem, depending on how rapidly the forcing varies, compared with the internal adjustment times of the system.

In “Dynamical Extended-Range Forecasting” (DERF), long-range forecasts are made for the statistics of the weather, by long-range forecasts are not made for individual synoptic events. The period of the forecast ranges upwards from 10 days to a year or so. The physical basis is the “long memory” of the external boundary conditions that provide persistently

anomalous forcing to the atmosphere, such as sea surface temperature, ground wetness, sea ice, etc. (e.g., Palmer, 1993; Palmer and Anderson, 1995; Barnett, et al. 1996; Stockdale, et al., 1998).

An example of the ideas underpinning DERF is shown in Fig. 10.25, which is taken from the work of Shukla (1985). The upper panel shows the observed sea-surface temperature anomaly for January 1983, relative to long-term climatology. The warm temperatures in the tropical eastern Pacific are characteristic of El Niño, as will be discussed in detail later. The middle panel shows the associated rainfall anomaly as simulated by a general circulation model, and the lower panel shows the observed rainfall anomaly as inferred from outgoing long-wave radiation during winter 1983. This study illustrates that anomalous precipitation in the tropics can be successfully predicted on the basis of anomalous sea surface temperature patterns of the type associated with El Niño. It does not necessarily follow that we can also predict the effects of El Niño on precipitation elsewhere in the world; nor does it follow that all types of sea surface temperature anomalies produce predictable effects on the climate; nor does it follow that all climate anomalies are forced by or even associated with sea surface temperature anomalies.

Fig. 10.26, which is taken from Charney and Shukla (1981), illustrates that DERF may be more feasible for the tropics than for middle latitudes. The two panels of the figure show the simulated and observed zonally-averaged standard deviations, σ_m and σ_o , as functions of latitude, and their ratio, for the mean July sea-level pressure; and the rainfall. The observed values are for land stations. The simulation results are based on runs of an atmospheric general circulation model, and are for grid-points over land. The model runs used the same sea surface temperatures in all cases, but the initial conditions varied. For middle latitudes, the simulated and observed standard deviations are quite comparable, while in the tropics the simulated values are much smaller than observed. The implication is that something is causing the sea level pressure and rainfall to vary more in the real tropics than they do in the simulations. Charney and Shukla (1981) hypothesized that the explanation is that in the real world the sea surface temperatures vary noticeably from year to year, e.g. due to El Niño, whereas in the simulations no such variations occurred.

One important implication is that the observed variations of tropical sea level pressure and rainfall are largely due to changes in sea surface temperature; this suggests that DERF can work in the tropics. In the middle latitudes, however, it appears that most of the variability of sea level pressure and rainfall can be explained without resorting to sea surface temperature variations, since in fact the model was able to reproduce the observed variability of sea level pressure and rainfall even though no sea surface temperature perturbations were prescribed as forcing in the simulations. The implication is that in midlatitudes the “signal” of sea surface temperature forcing is swamped by other variability, which we might call “noise,” so that midlatitude DERF may be difficult.

Fig. 10.27, taken from Lau (1985), shows temporal variations of monthly indices of the observed sea surface temperature (top) and the simulated zonal wind at 200 and 950 mb, precipitation, east-west sea level pressure gradient across the South Pacific, 200 mb height and an index of what is called the “Pacific-North American pattern,” as obtained in two 15-year model simulations performed with an atmospheric general circulation model. The two runs are shown in the left- and right-hand panels, respectively. The smooth curves superimposed on the various time series were obtained using a kind of running mean.

The imposed sea surface temperature changes were exactly the same in the two runs, which differed only through perturbations of their initial conditions. The main point here is that the two runs give very similar results, showing that the statistics indicated were

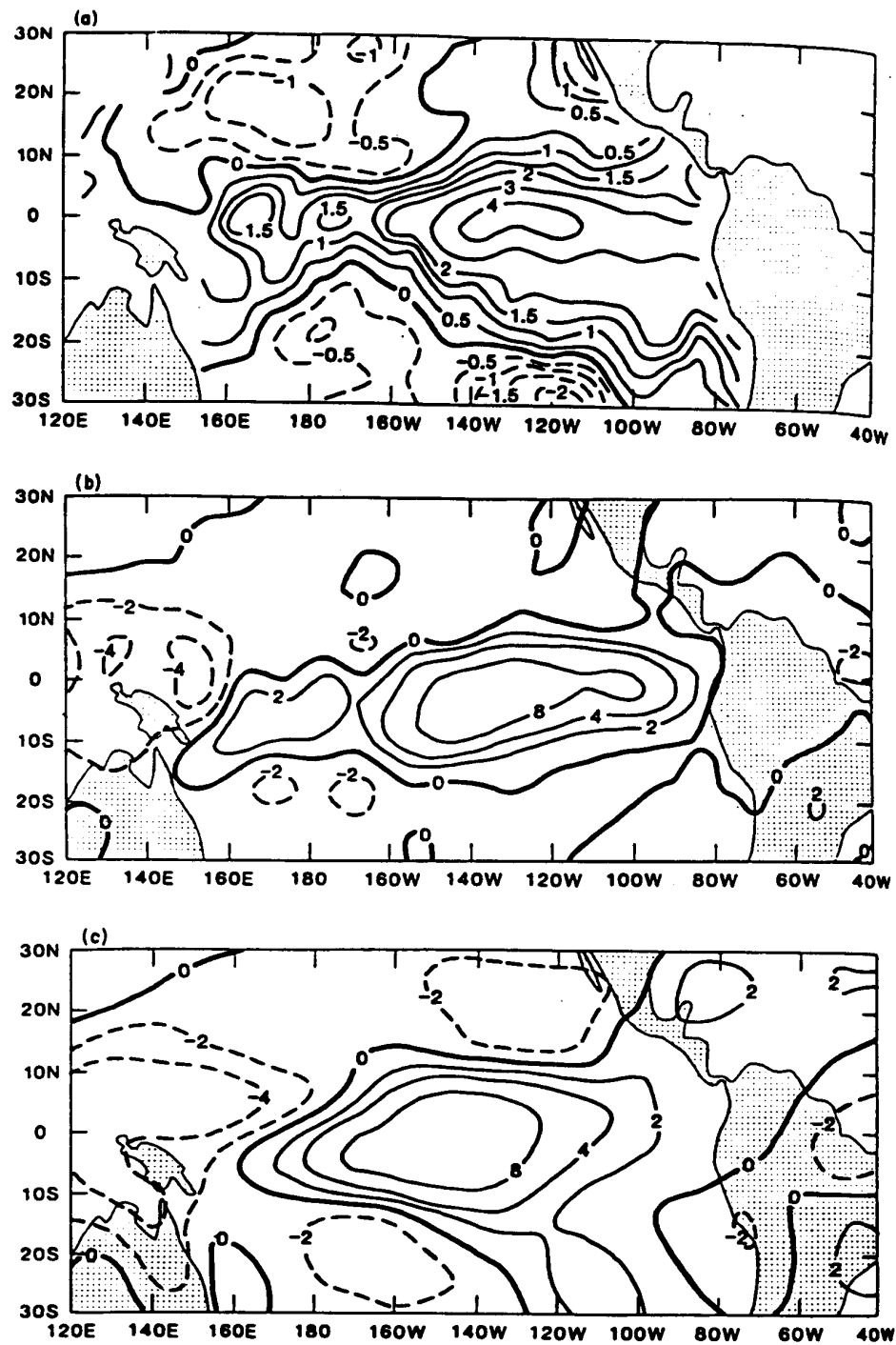


Figure 10.25:(a) Observed sea-surface temperature anomaly (K) for January 1983. (b) Simulated rainfall anomaly (mm day⁻¹). (c) Observed rainfall anomaly (mm day⁻¹) as inferred from outgoing long-wave radiation during winter 1983. From Shukla (1985).

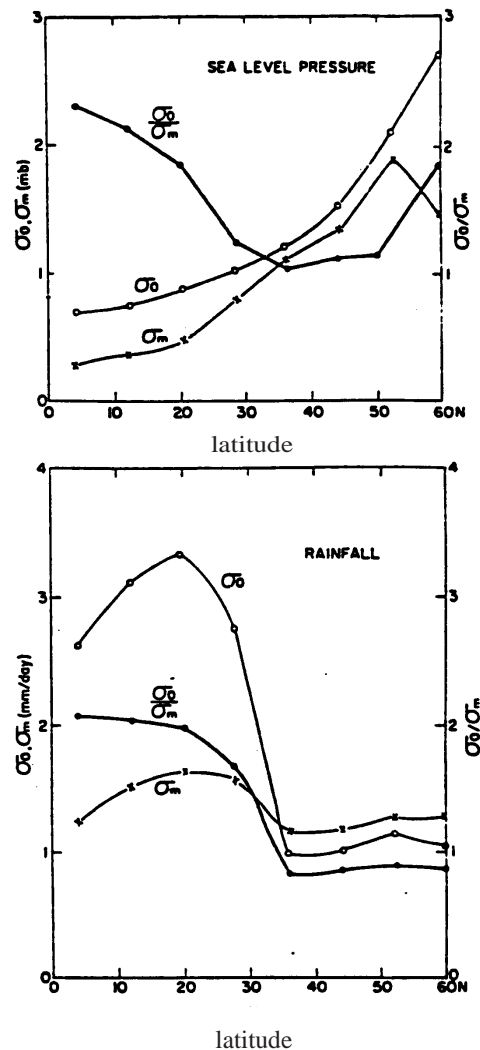


Figure 10.26: Model and observed zonally-averaged standard deviations, σ_m and σ_o , as functions of latitude, and their ratio, for: (a) mean July sea-level pressure; and (b) rainfall. Observed values are for land stations and model values are for grid-points over land. From Charney and Shukla (1981).

controlled by the year-to-year variations of the sea surface temperature, and were predictable far beyond the limit of deterministic predictability for individual weather events.

10.11 The World's Simplest GCM

Further studies of atmospheric predictability have been performed by using highly simplified models. The "World's Simplest GCM" (Lorenz, 1975, 1984, 1990) is described by the following equations:

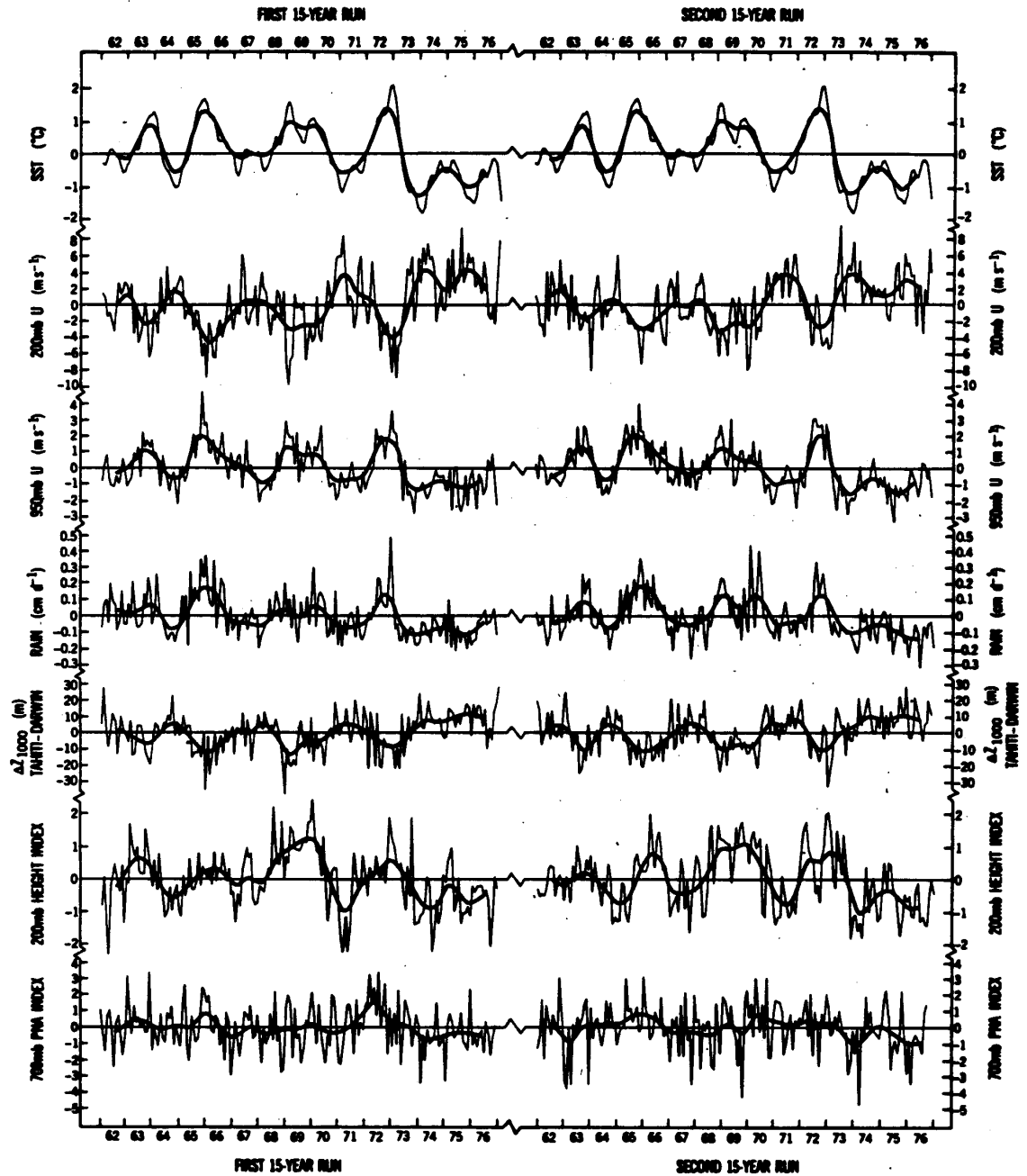


Figure 10.27: Temporal variations of monthly indices of sea surface temperature, zonal wind at 200 and 950 mb, precipitation, east-west sea level pressure gradient across the South Pacific, 200 mb height and the Pacific-North American pattern, for the first (left half of figure) and second (right half) 15-year model runs. The smooth curves superimposed on these time series are obtained using a running mean. From Lau (1985).

$$\begin{aligned}
\dot{X} &= -Y^2 - Z^2 - aX + aF \\
\dot{Y} &= XY - bXZ - Y + G \\
\dot{Z} &= bXY + XZ - Z.
\end{aligned}
\tag{10.82}$$

Here X represents the strength of the westerlies, Y and Z are the sine and cosine components of a planetary wave train, F represents the meridional heating contrast (a “forcing”), and G represents land-sea contrast (another “forcing”). For $F = G = 0$, the model has the trivial steady solution $X = Y = Z = 0$. For $G = 0$, we can have a steady solution with $Y = Z = 0$ even though $X \neq 0$; in this case, X has the steady solution $X = F$. This eddy-free solution can be unstable, however. Instability can lead to the growth of Y and Z . We interpret “large” F (~ 8) as “winter,” and “small” F (~ 6) as “summer.” Compare (10.82) with (10.74).

For fixed F , there can be one, two, or three steady solutions, depending on the value of G , as shown in Fig. 10.28. Here $F = 2$, so this is “super summer.” The steady solutions

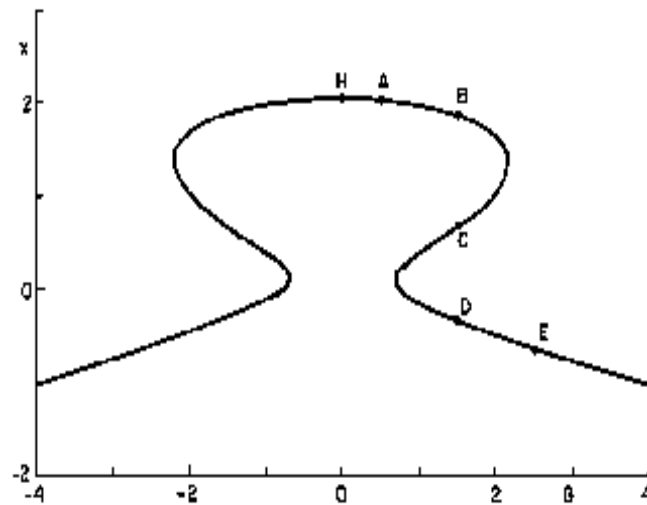


Figure 10.28: Variations of X with G , for steady solutions of (10.82) with $a = 0.25$, $b = 4.0$, and $F = 2.0$. H represents the Hadley solution. From Lorenz (1984).

shown in Fig. 10.28 are found most easily by fixing X and solving for G , Y , and Z .

Fig. 10.29 demonstrates that the model exhibits sensitive dependence on initial conditions. The three solutions shown are for the same values of the external parameters, but have slightly different initial conditions. The solutions clearly diverge after some time.

Fig. 10.30 shows two “summer” solutions started from different initial conditions. Note that these two summers are both fairly regular in appearance, but nevertheless appear quite different from each other, indicating that the model is capable of producing two “kinds”

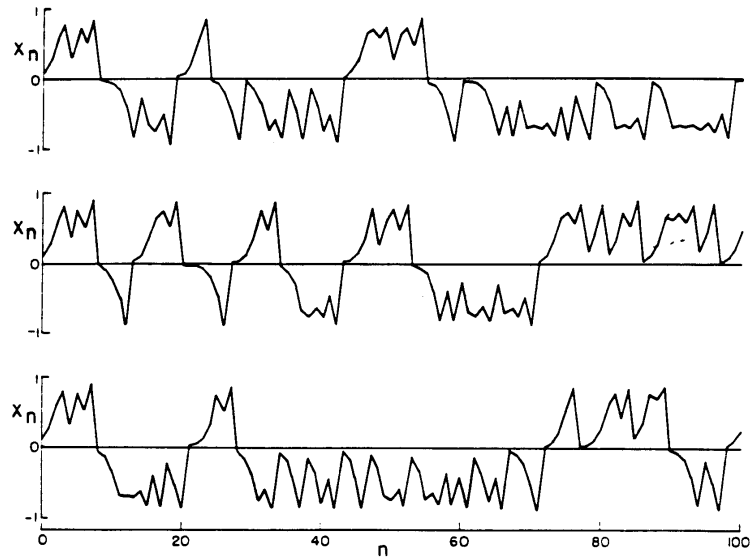


Figure 10.29: Solutions of (10.82) extending for 100 time steps, starting from initial values of 0.0999 (upper), 0.1000 (middle), and 0.1001 (lower). The straight-line segments joining consecutive points are solely for the purpose of making the chronological order easier to see. This figure illustrates sensitive dependence on initial conditions. From Lorenz (1975).

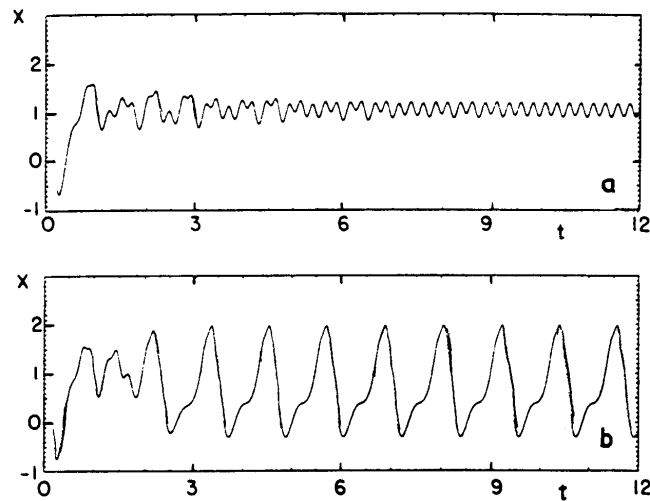


Figure 10.30: The variations of X (dimensionless) with t (months) in a numerical solution of (10.82), with $a = 0.25$, $b = 4.0$, $F = 6.0$, and $G = 1.0$ (summer conditions). The initial state is $(2.4, 1.0, 1)$. (b) The same as panel a, except that the initial state is $(2.5, 1.0, 0)$. From Lorenz (1990).

of summers: active summers (like the one in the lower panel) and inactive summers (like the one in the upper panel). Similarly, Fig. 10.31 shows two “winter” solutions started from different initial conditions. The two winters are highly irregular and look much the same,

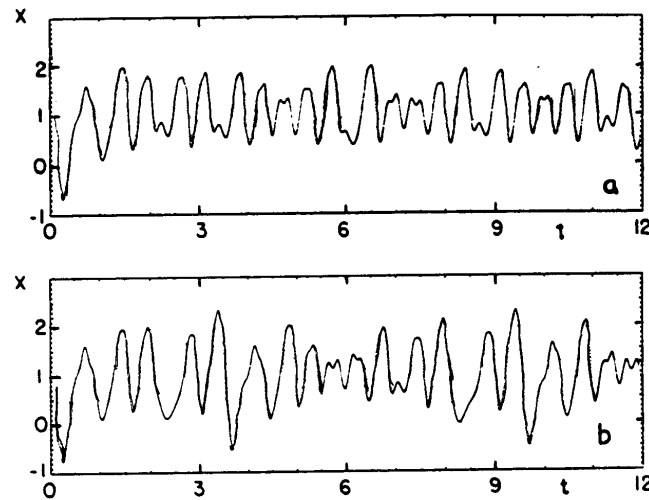


Figure 10.31:(a) The same as Fig. 8.19a, except that $F = 8.0$ (winter conditions). (b) The same as panel b, except that the initial state is $(2.5, 1.0, 0)$. From Lorenz (1990).

suggesting that all model winters are essentially equivalent; the model makes only one “kind” of winter.

Fig. 10.32 shows results from a six-year run, in which two different kinds of summers occur. The interpretation is very simple and interesting. Winters are chaotic. The model locks into either an active summer or an inactive summer, based on the “initial conditions” at the end of the winter, which are essentially random. When winter returns, all information about the previous summer is obliterated by nonlinear scrambling. The dice are rolled again at the beginning of the next summer.

Define σ as the (nondimensional) standard deviation of X within the period July through September. Active summers have large σ , and inactive summers have small σ . We can calculate one value of σ for each summer, or in other words one for each year. Fig. 10.33 shows the variations of σ (dimensionless) with time in years, in a 100-year numerical solution of (10.82), for the conditions of Fig. 10.32. Active and inactive summers alternate, at irregular intervals.

10.12 Pushing the attractors around

Palmer (1993, 1999) has argued that the climate system can be viewed as occupying a collection of attractors. For example, one attractor might represent an El Niño state, while another represents an La Niña state. The climate itself is a set of statistics which tells what the attractors look like and how frequently each is visited. Palmer argues that slowly varying external forcings can alter the frequency with which each attractor is visited, and that this is how climate change manifests itself. For example, some climate states may have frequent and/or persistent El Niños, while others have very few El Niños.

To illustrate these ideas, Palmer used a modified version of the model given by (10.74):

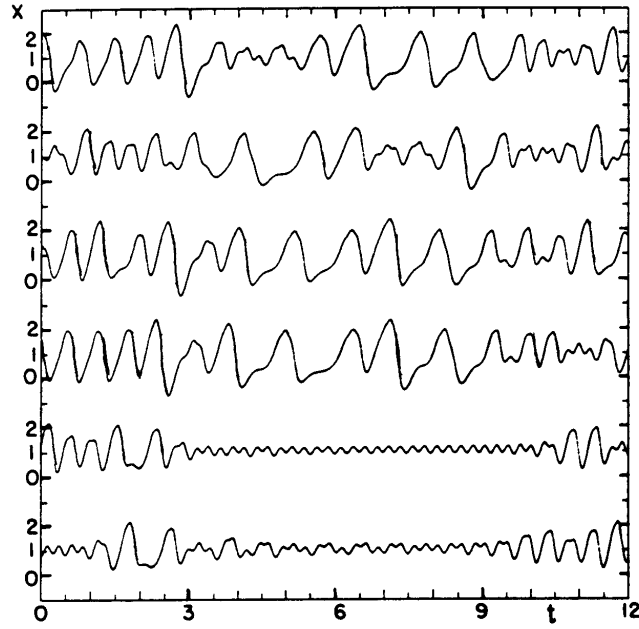


Figure 10.32: The variations of X (dimensionless) with t (months) in a 6-year numerical solution of (10.82), with $A = 0.25$, $b = 4.0$, $F = 7.0 + 2.0 \cos(2\pi t/\tau)$, and $G = 1.0$, where $\tau = 12$ months. Each row begins on 1 January, and, except for the first, each row is a continuation of the previous one. From Lorenz (1990).

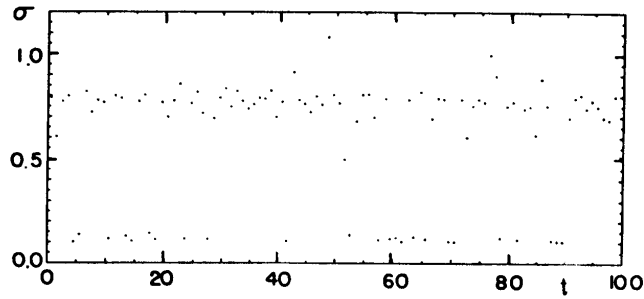


Figure 10.33: The variations of σ (dimensionless) with t (years) in a 100-year numerical solution of (10.82), for the conditions of Fig. 6, where σ is the standard deviation of X within the period July through September. From Lorenz (1990).

$$\begin{aligned}
 \dot{X} &= -\sigma X + \sigma Y + f_0 \cos \theta, \\
 \dot{Y} &= -XZ + rX - Y + f_0 \sin \theta, \\
 \dot{Z} &= XY - bZ.
 \end{aligned}
 \tag{10.83}$$

Here f_0 is a forcing which tries to push X and Y in the direction of the angle θ in the (X, Y) plane. In (10.74), $f_0 = 0$. Fig. 10.34 shows how different choices of θ affect the

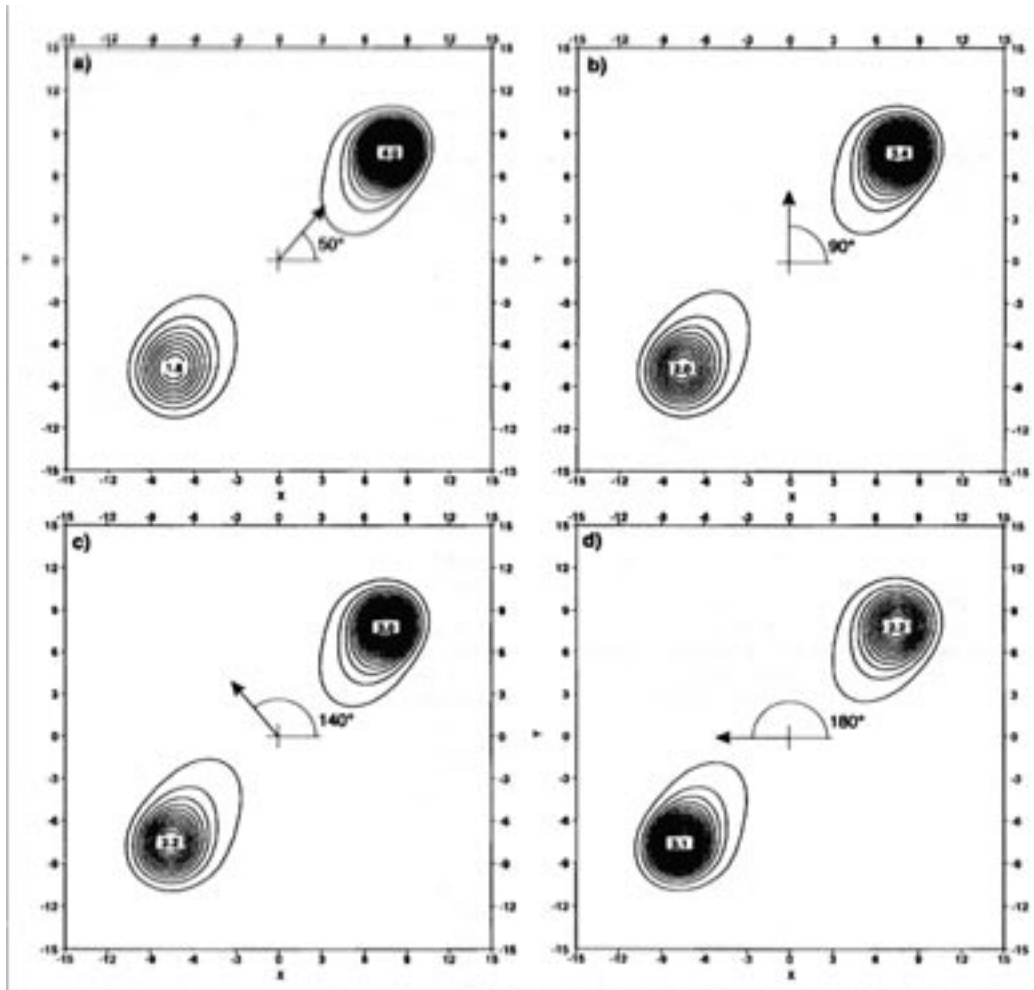


Figure 10.34: The impacts of various imposed forcings on the PDF of the butterfly model. state vector with running time mean, in the (X, Y) plane. From Palmer (1999).

probability density function (PDF) of the solution in the (X, Y) plane. The maxima of the PDF are the attractors of the model. As θ changes, the locations of the maxima of the PDFs do not change much. This means that the attractors of the model are insensitive to θ . Note, however, that the maxima become stronger or weaker as θ changes. This means that by varying the forcing we can cause some attractors to be visited more often, and others to be visited less often. Also notice that there is some tendency for the solution to be pushed in the direction in which the forcing acts.

10.13 Summary

A circulation is turbulent to the extent that its predictability time is shorter than the

time scale of interest to us. The fundamental cause of turbulence is shearing instability. Vortex stretching leads to a kinetic-energy cascade in three-dimensional turbulence. In contrast, enstrophy cascades in two-dimensional turbulence, while kinetic energy anti-cascades. The Rhines scale limits the meridional growth of the turbulent vortices, which become elongated into pairs of zonal jets.

Deterministic weather prediction is impossible beyond about two weeks (some nuts to the contrary notwithstanding). It may be possible to predict some statistics of the weather for a month or a season ahead, because of the external forcing due to persistent boundary anomalies. This appears to be most feasible in the tropics, because of the favorable signal-to-noise ratio. Longer term climate anomalies are also predictable provided that they are driven by an external forcing which is itself predictable, and provided that the response to this forcing is large enough to detect against the climatic noise due to natural variability and competing external forcings.

The ocean circulation is also chaotic but has a much longer predictability time than the atmosphere; therefore, to the extent that the sea surface temperature is predictable, and to the extent that the statistics of the weather are influenced by the sea surface temperature, the statistics of the weather are predictable beyond the predictability limit of the atmosphere itself.

Even a chaotic system responds in a statistically predictable way to sufficiently strong external forcing; this is why summers are predictably warm and winters are predictably cold. The climate responds predictably to a sufficiently strong external forcing, provided that the forcing itself is predictable.

Problems

1. Summarize the basic dynamical differences between two-dimensional and three-dimensional turbulence. In your answer, be sure to discuss:
 - a) vortex stretching and twisting;
 - b) vorticity conservation;
 - c) the shape of the energy spectrum;
 - d) the existence of an energy cascade.
2. Analyze the effects of buoyancy on a spherical bubble of warm air, with uniform potential temperature θ , sitting in an infinite environment of uniform potential temperature $\theta - \Delta\theta$, where $\Delta\theta > 0$. Draw a sketch indicating:
 - i) Where vorticity generation occurs.
 - ii) The sign of the vorticity generation.
 - iii) Where vorticity generation does not occur.
3. Show that in two-dimensional flow the variance of a passive scalar, per unit wave

number, varies as k^{-1} .

4. Show that in two-dimensional flow the geopotential variance, per unit wave number, varies as k^{-5} .
5. Consider Lorenz's butterfly model:

$$\dot{X} = -\sigma(X - Y)$$

$$\dot{Y} = -XZ + rX - Y$$

$$\dot{Z} = XY - bZ$$

Here σ , b and r are constants. Find the stationary solutions for X , Y , and Z , and investigate the stability of these solutions as functions of σ , b , and r . Assume $r \geq 1$, $b \geq 0$.

6. A Fortran code for the "World's Simplest GCM," as discussed by Lorenz (1984), can be obtained from the instructor. The equations of the model are:

$$\dot{X} = -Y^2 - Z^2 - aX + aF, \quad (10.84)$$

$$\dot{Y} = XY - bXZ - Y + G, \quad (10.85)$$

$$\dot{Z} = bXY + XZ - Z. \quad (10.86)$$

Time derivatives have been approximated using the fourth-order Runge-Kutta time-differencing scheme, which is discussed in many numerical analysis books, with a time step of $\Delta t = 1/30$. Following Lorenz, we interpret a time unit as corresponding to 4 hours, so that six time units correspond to one day. The time step is thus 4/30 hours.

We use two slightly different versions of the model. The first, which we think of as the "real world," uses $a = 0.25$, $b = 4$, $F = 8$, and $G = 1$. The second, which we think of as the "model," is identical to the first, except that $b = 4.01$. The code is set up in "real world" mode.

The "standard" initial conditions are $X = 2.5$, $Y = 1$, $Z = 0$.

- a) Demonstrate that two identical runs of the "real world" give identical results.
- b) Do a test to show that the "real world" exhibits sensitive dependence on initial conditions.

c) Make a 100-day run of the “real world” and save the results for X , Y , and Z once per simulated day. (The code is already set up to do this.)

Starting from each of the first 90 days of this “real world” weather record, make a ten-day forecast by running the “model.” Save the results from each of these forecasts, once per simulated day.

Produce a plot similar to Fig. 10.22, with at least the curves $E_{0,k}$ and $E_{k-1,k}$. (You may want to do more than just these two.) Here use X as the variable you study, just as Lorenz (1982) used the 500 mb height.

d) Still using X as your variable, plot $\frac{dE}{dk}$ versus E , as in Fig. 10.24, for both the “perfect model” and the “imperfect model.” Estimate the value of a , the error e-folding rate.

7. Explain as clearly as you can why it is not possible to make skillful deterministic weather forecasts of indefinite range.
8. Describe the methods used by Lorenz to estimate the limits of predictability.
9. Find the equilibrium solutions of Lorenz’s “World’s Simplest GCM,” and analyze their stability.

