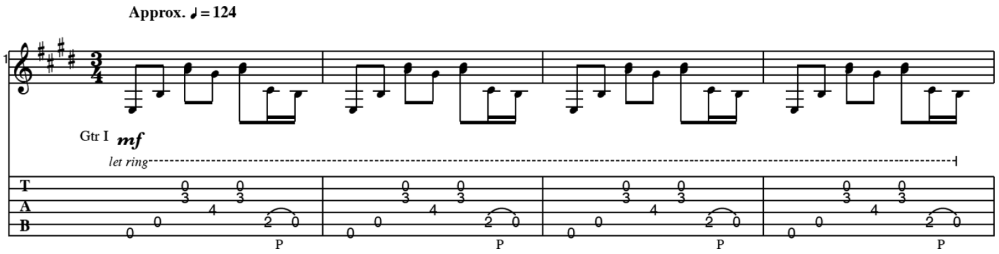
12

Predictability



Optimal temporal propagation.
EBEF♯BE, Turning, Alex de Grassi

All of us are familiar with day-to-day weather phenomena. The typical time scale of the processes associated with the weather is a few days. Most of us also know that weather predictions with a lead time longer than about ten days are not very reliable because of the chaotic nature of the atmospheric flow on these time scales. As Chuck Leith, referring to the Lorenz 'butterfly effect', put it: 'It is not all butterflies, even talking about the weather can change the weather . . .' (see e.g., http://www.archives.ucar.edu/exhibits/washington/science/early_atmos_sci).

As we have seen in Chapter 6, the state vector of the climate system involves not only the atmospheric flow but also the ocean circulation, land and sea ice and, for example, land-surface properties. Climate change refers to the long-term (50–100 years) development of the climate state due to internal variability and the variations in external (e.g., solar) forcing. Over the past several decades, the trajectory of this system is also influenced by the release of CO_2 due to human activities, which can be seen as an external radiative forcing on these time scales.

Regardless of these changes in the radiative forcing conditions, it is of central importance to know the predictability horizons of climate variability on different time scales. We address this issue in this final chapter, again from a stochastic dynamical

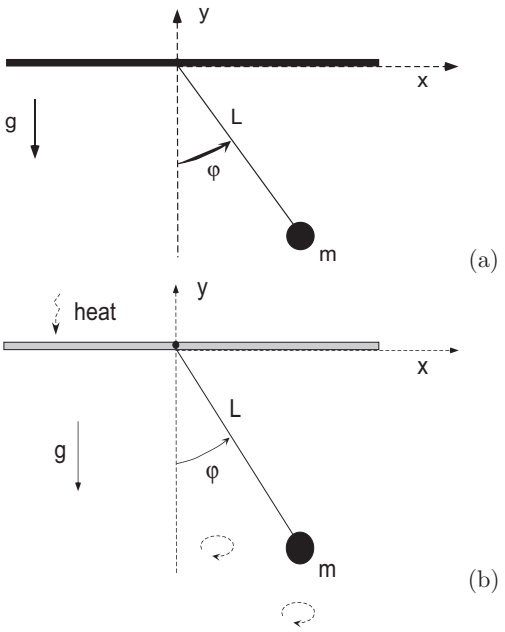


Figure 12.1 (a) Sketch of a single pendulum in vacuum, consisting of a mass m at the end of a rope with length L which is connected to a fixed surface. (b) The same pendulum except that it now moves in air that is heated at the upper surface.

systems point of view. In Section 12.1, the essence of the prediction problem is illustrated using a simple example. Then in Section 12.2, measures of predictability, such as predictive power and relative entropy, are presented. This is followed by the methodology (Section 12.3) to study the behaviour of nearby trajectories needed to assess error propagation. In the next section (12.4), methods to incorporate observations into models to 'steer' trajectories are discussed. As we see, many dynamical systems concepts and methods are used in the study of the predictability of phenomena in the climate system.

12.1 The prediction problem

It is often questioned in the popular media: how can we ever aim to predict the climate state for 2050 with some confidence if we cannot predict the weather more than a few days ahead with a reasonable skill? This question is at the heart of the prediction problem, and we address it in the next section.

12.1.1 A simple example

Consider, first, a single pendulum (Fig. 12.1) where a mass m is connected to the end of a wire of length L, which is connected to a fixed surface. The position of the mass can

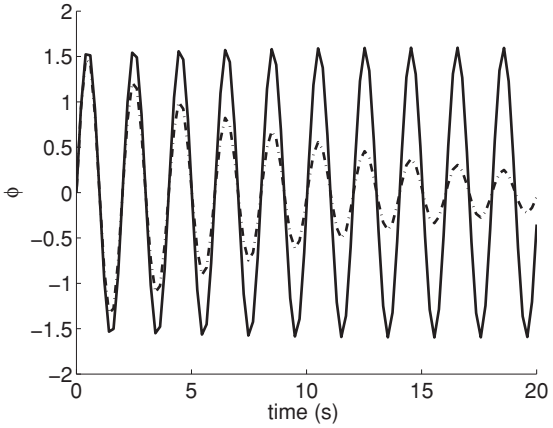


Figure 12.2 Plot of the amplitude of the angle $\phi(t)$ of the mass of a simple pendulum for $L = 1$ m, $g = 9.8$ ms^{-2} versus time (in s) without (drawn) and with (dash-dotted) the presence of air friction ($\gamma = 0.1$ s^{-1}). The initial conditions are $\phi(0) = 0$ and $d\phi/dt = 5$.

be completely specified by the angle ϕ. The prediction problem can be loosely formulated as follows: if at a certain time we release the mass of the pendulum at a specific angle (and with a specific speed), what is the position of the mass at a later time?

As a first approach, we idealise the situation as being in a vacuum (Fig. 12.1a). We can use Newton's second law with gravity as the only force to give the equation for the angle position as

$$\frac{d^2\phi}{dt^2} + \frac{g}{L}\sin\phi = 0, \tag{12.1}$$

where g is the gravitational acceleration. If we know both ϕ and $d\phi/dt$ at the initial time t_0, then (12.1) provides a model to calculate $\phi(t_1)$ from $\phi(t_0)$. For small angles ϕ, we can approximate $\sin\phi \approx \phi$ and (12.1) can be solved as

$$\phi(t) = A\sin\omega_0 t + B\cos\omega_0 t;\ \omega_0 = \sqrt{\frac{g}{L}},$$

where A and B are determined from the initial conditions. A typical solution of $\phi(t)$ is shown as the drawn curve in Fig. 12.2; once the initial conditions are known accurately, we can make an accurate prediction of the position and velocity of the mass at a later time.

If, instead, the pendulum motion is considered to occur in air, the problem is suddenly much more complicated (Fig. 12.1b). The air undergoes a complex motion, interacts with the mass of the pendulum and introduces rapid fluctuations in the position of the mass of the pendulum. We can call these fast motions of the pendulum the 'weather' of the pendulum and the slow motion of the mass of the pendulum its 'climate'.

For an exact prediction of the pendulum's weather, one would need to represent the interactions of the air with the pendulum mass. We can divide, for example, the air volume into grid boxes and compute the motion of the air parcels using the Navier-Stokes equations. As the motion of the air is chaotic on the small time scale (much smaller than the time scale of the period $2\pi/\omega_0$ of the pendulum), we are certainly not able to predict the detailed position of the pendulum mass on this time scale.

However, we may predict the effect of the air on the pendulum's 'climate' relatively easily. For example, the effect of the collisions of the air molecules with the pendulum mass can be represented as a simple net frictional force, giving a modified equation (12.1), again for small ϕ as

$$\frac{d^2\phi}{dt^2} + \gamma\frac{d\phi}{dt} + \frac{g}{L}\phi = 0. \tag{12.2}$$

In this case, we use a so-called parameterisation and do not need to compute the positions of the air parcels. However, the use of this parameterisation has its toll. There is now a friction coefficient γ in the model that is unknown, and it needs to be determined empirically.

The behaviour of the position of the mass of the pendulum (for a specific value of γ) is shown as the dash-dotted curve in Fig. 12.2, and it is clear that it approaches zero for $t \to \infty$ due to the presence of the friction. When the air around the pendulum is slowly heated at the boundary of the container (Fig. 12.1b), the motion of the air parcels will be affected (the viscosity of the air is temperature dependent) and hence also the slow variations in the position of the pendulum. This effect could be taken into account by a temperature-dependent friction coefficient γ.

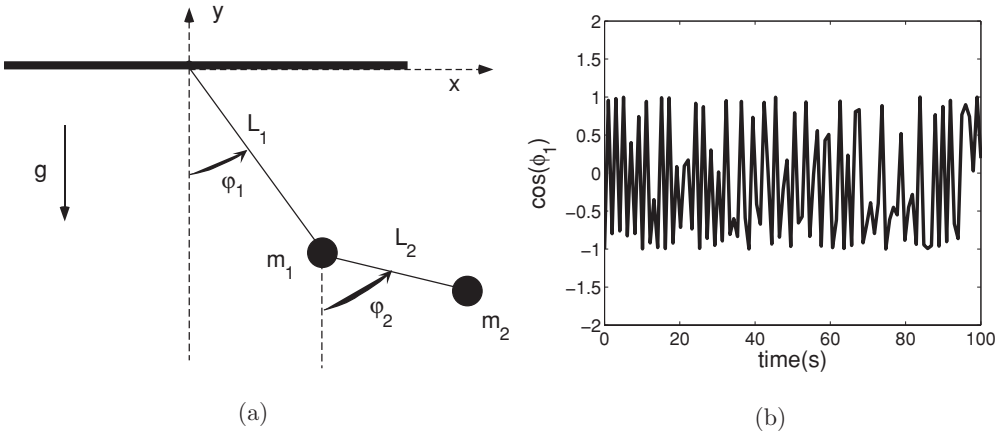
In summary, because the 'climate' of the pendulum is determined by very different processes than its 'weather', detailed predictions of the position of the pendulum on the long time scale (the 'climate') still may be made despite the fact that the air motion (and the pendulum's 'weather') is unpredictable on the short time scale.

However, the processes responsible for the 'climate' of the pendulum also may lead to unpredictable behaviour. An example is the slightly more complicated system of the double pendulum (Fig. 12.3a). In vacuum, Newton's law now leads to a coupled set of nonlinear equations given by

$$\ddot{\phi}_1 + \frac{g}{L_1}\sin\phi_1 + \frac{m_1}{m_1+m_2}\frac{L_2}{L_1}\left[\cos(\phi_2-\phi_1)\ddot{\phi}_2 - \sin(\phi_2-\phi_1)\dot{\phi}_2^2\right] = 0,$$

$$\ddot{\phi}_2 + \frac{g}{L_2}\sin\phi_2 + \frac{L_1}{L_2}\left[\cos(\phi_2-\phi_1)\ddot{\phi}_1 + \sin(\phi_2-\phi_1)\dot{\phi}_1^2\right] = 0,$$

where a dot indicates differentiation to t. A plot of the variable $\cos\phi_1(t)$ is shown in Fig. 12.3b, displaying irregular oscillations. It is well known that this double pendulum system displays chaotic behaviour, with the largest Lyapunov exponent being positive



(a) (b)

Figure 12.3 (a) The double pendulum consisting of two masses m_1 and m_2 connected by strings of length L_1 and L_2. (b) Solution $\cos\phi_1$ for the case $m_1 = m_2 = 1$ kg, $L_2 = 2$ m, $L_1 = 1$ m and $g = 9.8\ \mathrm{ms}^{-2}$. The initial conditions are $\phi_1(0) = \phi_2(0) = \pi$ and $d\phi_1/dt = 0$, $d\phi_2/dt = 5$.

(Stachowiak and Okada, 2006). The new element here is that this pendulum system already displays chaotic motion on the 'climate' time scale, limiting predictability of the system, even if no air molecules are considered.

A pendulum system having even more complexity is the double pendulum placed in air that is heated from the boundaries. Here both the 'weather' and the 'climate' of the pendulum have very different limits of predictability and interact with each other, whereas there are also slowly varying boundary influences affecting the motion of the air and the pendulum.

12.1.2 Prediction using climate models

As seen in Chapter 6, a climate model is the recipe that relates future states to past states, similarly to the equation of the pendulum in the previous section. Such models necessarily have many approximations to the real world as a spatially and temporally discrete representation of the state vector is used and parameterisations are incorporated (with uncertain parameters). With a given initial condition, the model can be used to compute future states; this is usually referred to as a simulation.

There are several reasons why it is not so useful for predictions to perform a single simulation with a particular climate model over a certain time period: (i) the initial conditions are poorly known, (ii) there are many uncertainties in the representation of physical and chemical processes (Section 6.4) and (iii) the presence of internal variability may introduce a random element in the model solutions, as in the double pendulum motion (Section 12.1.1).

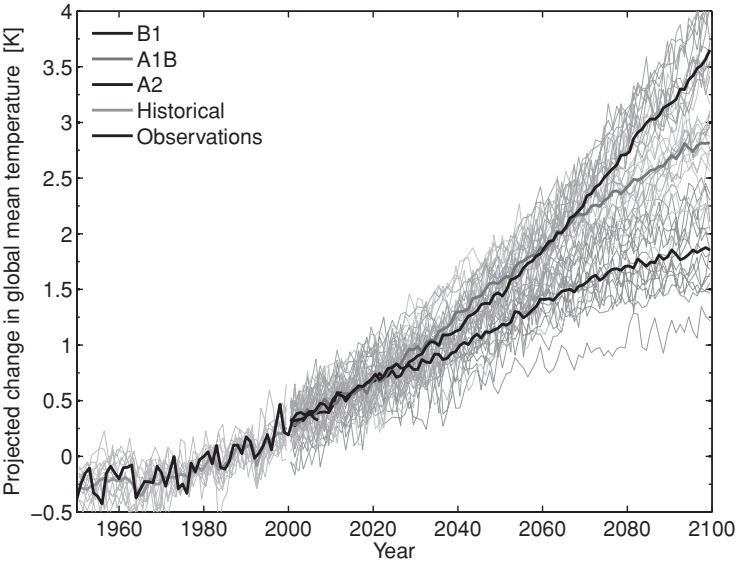


Figure 12.4 Global mean, annual mean and surface air temperature projections from fifteen different global climate models under three different greenhouse emission scenarios from 2000 to 2100 (thin lines): SRES A2 (red), A1B (green), and B1 (blue), designated as high-, medium-, and low-emission paths, respectively. The same models forced with historical forcings are shown as the thin grey lines, and the observed global mean temperatures from 1950 to 2007 are shown as the thick black line. The multimodel mean for each emission scenario is shown with thick coloured lines (figure from Hawkins and Sutton, 2009). (See Colour Plate.)

This has motivated the use of the following types of multiple simulations, or ensembles:

- The standard ensemble: in these simulations, one considers the sensitivity of the climate model solutions to the initial conditions and provides an ensemble of 'equally likely climate development'.
- The perturbed-parameter ensemble: here one investigates the uncertainties due to the representation of physical and chemical processes by varying parameters in the climate model; prominent examples can be found on http://www.climateprediction.net.
- The multimodel ensemble: here one investigates the uncertainties due to a different representation of physical processes in different models. The different models serve as a perturbed parameter ensemble and the different simulations with a single model as a standard ensemble.

The final aim with these ensemble methods is to estimate a range of possible trajectories from an uncertain initial condition with an imperfect model.

An example of the use of ensemble simulations is given in Fig. 12.4 from simulation efforts carried out for the IPCC-AR4 assessment (Hawkins and Sutton, 2009), where one is interested in, for example, the global mean surface temperature up to 2100.

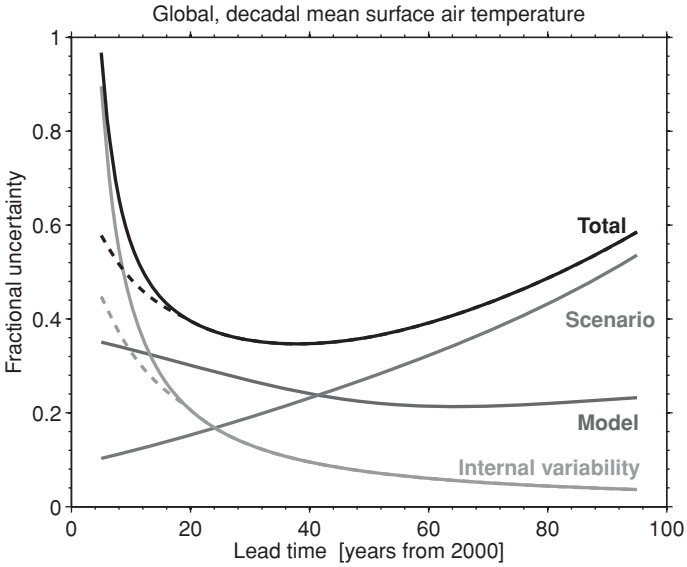


Figure 12.5 The relative importance of each source of uncertainty in decadal mean surface air temperature predictions is shown by the fractional uncertainty (the 90% confidence level divided by the mean prediction), for the global mean, relative to the warming since the year 2000 (i.e., a lead of zero years) (figure from Hawkins and Sutton, 2009).

Here, an additional element of uncertainty comes from the future forcing conditions, for example, future emissions, and hence simulations for different so-called emission scenarios are performed. Hence, predictions are usually referred to as projections. In Fig. 12.4, the annual mean global mean surface temperature from fifteen different Global Climate Models is plotted for three different emission scenarios (low: B1; medium: A1b; and high: A2). The spread around the multimodel mean provides here an estimate of effect of model uncertainty on future climate under different emission scenarios.

The different contributions to the uncertainty (due to internal variability, model uncertainty and forcing scenario uncertainty) for decadal scale global mean temperature predictions is shown in Fig. 12.5; the initial condition uncertainty is not important on the longer time scales. On the shorter lead times, the internal variability is the most important component, whereas the contribution of the forcing scenario uncertainty is relatively small. Indeed, in Fig. 12.4, the model results for the different scenarios are very similar up to 2040. On the longer time scale, however, scenario uncertainty becomes dominant.

12.2 Concepts of predictability

Objective measures of predictability have been developed since the work of Lorenz (1969). Predictability studies of the first kind address how uncertainties in the initial

state affect the prediction at a later stage. Indeed, initial uncertainties can amplify as the prediction lead time increases, thus limiting the predictability of the first kind. Examples are weather prediction and El Niño prediction (which is a climate prediction of the first kind). For these studies, a standard ensemble is suited from which it can be seen how nearby trajectories behave.

Predictability studies of the second kind address the predictability of the response of the system to changes in boundary conditions. Examples are the response of atmospheric flows due to changes in sea-surface temperature and the response of climate to changes in orbital insolation variations. For these kinds of studies, a parameter ensemble is suited as it shows how trajectories for nearby parameter values behave.

In this section, after a more abstract discussion on the density of trajectories, a measure of predictability is presented for predictability of the first kind, that is, the concept of predictive power.

12.2.1 The Liouville equation

For a one-dimensional stochastic dynamical system, as discussed in Section 3.5, the probability density function $p(x,t)$ of the random variable X_t, evolving according to the Itô SDE,

$$X_t = X_0 + \int_0^t a(X_s, s)ds + \int_0^t b(X_s, s)\, dW_s, \tag{12.4}$$

is determined by the Fokker-Planck equation (3.73), that is,

$$\frac{\partial p}{\partial t} + \frac{\partial (ap)}{\partial x} - \frac{1}{2}\frac{\partial^2 (pb^2)}{\partial x^2} = 0. \tag{12.5}$$

For a deterministic system (with $b = 0$), this reduces to the so-called Liouville equation,

$$\frac{\partial p}{\partial t} = -\frac{\partial (ap)}{\partial x}. \tag{12.6}$$

Given an uncertainty in the initial conditions, represented by an initial probability density function $p(x,0) = f(x)$, the Liouville equation provides the development of the probability density function in time. For an N-dimensional deterministic system

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}), \tag{12.7}$$

with initial condition $\mathbf{x}(0) = \mathbf{x}_0$, the probability density function is usually indicated by a density $\rho(\mathbf{x}, t)$, and the Liouville equation (12.6) generalises to

$$\frac{\partial \rho}{\partial t} + \sum_{k=1}^{N} \frac{\partial (\rho \mathbf{F}_k)}{\partial x_k} = 0. \tag{12.8}$$

Equation (12.8) can be formally solved as (Ehrendorfer, 1994)

$$\rho(\mathbf{x}, t) = f(\mathbf{x}_0)\, e^{-\int_0^t \psi(\mathbf{x}(\mathbf{x}_0,s))ds}, \tag{12.9}$$

where $f(\mathbf{x})$ is again the initial density and ψ is given by

$$\psi(\mathbf{x}) = \sum_{k=1}^{N} \frac{\partial \mathbf{F}_k}{\partial x_k}, \tag{12.10}$$

which is the trace of the Jacobian matrix.

Example 12.1 Solution of the 1D Liouville equation In Ehrendorfer (1994), the one-dimensional example

$$\frac{dx}{dt} = ax^2 + bx + c\ ,\ \Delta = \frac{b^2}{4} - ac > 0, \tag{12.11}$$

is considered. With $x(0) = \eta$, the solution of (12.11) can be determined as

$$x(\eta, t) = \frac{r_1(a\eta + r_2)e^{\gamma t} - r_2(a\eta + r_1)}{-a(a\eta + r_2)e^{\gamma t} + a(a\eta + r_1)}, \tag{12.12}$$

with $r_1 = b/2 + \sqrt{\Delta}$, $r_2 = b/2 - \sqrt{\Delta}$ and $\gamma = r_1 - r_2$, as can be verified by direct substitution. To obtain the density $\rho(x, t)$, we determine $\psi = 2ax + b$ and use it to obtain

$$\rho(x, t) = f(\eta)e^{-\int_0^t (2ax(\eta,s)+b)ds}, \tag{12.13}$$

which eventually provides an expression of ρ in terms of η. The latter is implicitly a function of x and t through inversion of (12.12), giving

$$\eta = \eta(x, t) = \frac{1}{a}\left[\frac{ax(r_2e^{\gamma t} - r_1) + r_1r_2(e^{\gamma t} - 1)}{ax(1 - e^{\gamma t}) - r_1e^{\gamma t} + r_2}\right] \tag{12.14}$$

to finally give

$$\rho(x, t) = \frac{f(\eta)}{\gamma^2} \exp(bt + \frac{2r_1}{\gamma} \ln[a\eta(e^{-\gamma t} - 1) - r_2 + r_1e^{-\gamma t}] - \frac{2r_2}{\gamma} \ln[a\eta(1 - e^{\gamma t}) + r_1 - r_2e^{\gamma t}]). \tag{12.15}$$

In Fig. 12.6, the development of the density ρ over time (six values in the interval $t \in [0, 0.3]$) is shown for $a = -1$, $b = 1$ and $c = 2$. The initial density is Gaussian with a mean $\bar{x} = -2$ and variance 0.1. In this case, the fixed points of the dynamical system (12.11) are determined by $x^2 - x - 2 = 0$, of which $\bar{x} = -1$ is unstable and $\bar{x} = 2$ is stable. In time, the density moves away from the unstable fixed point to give a substantial spread. ■

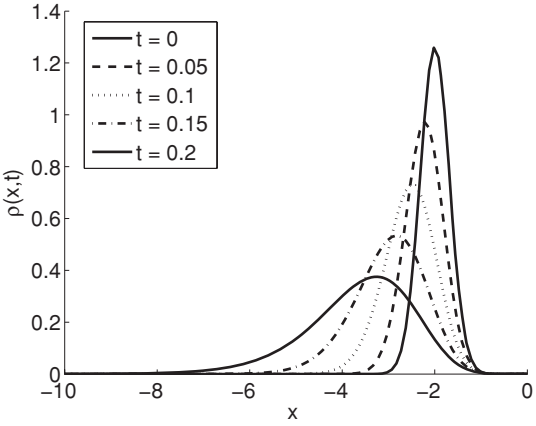


Figure 12.6 Development of the probability density function $\rho(x, t)$ in time for the problem (12.11) with $a = -1$, $b = 1$ and $c = 2$ (redrawn from Ehrendorfer, 1994).

12.2.2 Predictive power

Several measures have been developed of the predictability of nonlinear systems, such as predictive power (Schneider and Griffies, 1999) and prediction utility (Kleeman, 2002). Of those, we discuss only the predictive power next.

Consider the random n-dimensional state vector of a system $\mathbf{X}_\nu$ where ν indicates a time index. Indicate a particular realisation by $\mathbf{x}_\nu$ and denote its prediction by $\hat{\mathbf{x}}_\nu$. Because of the stochastic nature of the system, there is a prediction error

$$\mathbf{e}_\nu = \mathbf{x}_\nu - \hat{\mathbf{x}}_\nu \rightarrow \mathbf{x}_\nu = \hat{\mathbf{x}}_\nu + \mathbf{e}_\nu, \tag{12.16}$$

and hence in terms of random variables, this can be written as

$$\mathbf{X}_\nu = \hat{\mathbf{X}}_\nu + \mathbf{E}_\nu, \tag{12.17}$$

where the predictor $\hat{\mathbf{X}}_\nu$ is the random vector of which $\hat{\mathbf{x}}_\nu$ is a realisation.

The probability distribution of the state $\mathbf{X}_\nu$ is the climatological distribution, which reflects the prior (before any predictive information is available besides the climatological mean) uncertainty. The statistical properties of $\mathbf{X}_\nu$ can, for example, be obtained by analysing a long data set of observations or a long control simulation of a particular model. The probability distribution of the prediction error $\mathbf{E}_\nu$ reflects the a posteriori uncertainty that remains after the prediction has become available. Intuitively, in particular for univariate processes, one would expect that the predictability of the system is low if the variance of the prediction error is as large (or larger) than the climatological variance.

For multivariate processes, measures based on information theory can be used to quantify predictability. The degree of uncertainty associated with a probability density

function $p_{\mathbf{x}}(\mathbf{x})$ of a random variable $\mathbf{x}$ is given by the entropy (Cover and Thomas, 2006)

$$S_{\mathbf{x}} = -k \int p_{\mathbf{x}}(\mathbf{x}) \ln p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}, \tag{12.18}$$

where k is a constant. The quantity $S_{\mathbf{x}}$ can be seen as the additional information, say, the average number of bits, which is necessary to completely specify $\mathbf{x}$.

The prior entropy $S_{\mathbf{X}_\nu}$ is the average missing information when only the climatological distribution is known. The posterior entropy $S_{\mathbf{E}_\nu}$ is the average missing information after the prediction has become available. The predictive information R_ν, given by

$$R_\nu = S_{\mathbf{X}_\nu} - S_{\mathbf{E}_\nu}, \tag{12.19}$$

is the average information about the state contained in the prediction; it should always be positive. The predictive power $\alpha_\nu \in [0, 1]$ is then defined as (Schneider and Griffies, 1999)

$$\alpha_\nu = 1 - e^{-R_\nu}. \tag{12.20}$$

When the probability distribution of an n-dimensional state vector $\mathbf{X}$ is Gaussian (see Section 3.2), with covariance matrix Σ, then it can be shown (Schneider and Griffies, 1999) that

$$S_{\mathbf{x}} = \frac{k}{2}(n + n \log 2\pi + \ln \det(\Sigma)), \tag{12.21}$$

which can be easily verified for $n = 1$ for which $\Sigma = \sigma^2$ and p_X given by (3.8). When the climatological covariance matrix is indicated by Σ_ν and the covariance matrix of the prediction error by C_ν, the predictive information is given by

$$R_\nu = -\frac{k}{2} \ln \frac{\det C_\nu}{\det \Sigma_\nu}, \tag{12.22}$$

and with the choice $k = 1/n$, the predictive power α_ν is finally determined as (using the product rule of determinants)

$$\alpha_\nu = 1 - e^{-R_\nu} = 1 - (\det C_\nu \, {\Sigma_\nu}^{-1})^{\frac{1}{2n}}. \tag{12.23}$$

For univariate processes with climatological variance σ_c^2 and prediction error variance σ^2, α_ν reduces to

$$\alpha_\nu = 1 - \frac{\sigma^2}{\sigma_c^2}. \tag{12.24}$$

When the prediction error variance is equal to the climatological variance, the predictive power is indeed zero. In this case, the prediction does not add any additional information than already available through the climatology.

In practice, the matrices C_ν and Σ_ν can be determined by performing a long control and a standard ensemble simulation. From the long control integration, we obtain a trajectory $\mathbf{x}^c_\nu$, $\nu = 1, \ldots N$, where the superscript c refers to the control simulation. From this time series, the mean and (time-independent) covariance matrix of the climatological distribution can be estimated as

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{\nu=1}^{N} \mathbf{x}^c_\nu, \tag{12.25a}$$

$$\hat{\Sigma}_\nu = \frac{1}{N-1}\sum_{\nu=1}^{N} (\mathbf{x}^c_\nu - \bar{\mathbf{x}})(\mathbf{x}^c_\nu - \bar{\mathbf{x}})^T. \tag{12.25b}$$

Next, consider a standard ensemble of the same model in which M ($N \gg M$) initial conditions $\mathbf{x}^1_0, \ldots, \mathbf{x}^M_0$ are integrated forward in time. At time t_ν, these initial conditions have evolved to states $\mathbf{x}^1_\nu, \ldots, \mathbf{x}^M_\nu$. The mean of the M-member ensemble

$$\hat{\mathbf{x}}_\nu = \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}^i_\nu \tag{12.26}$$

is a prediction of the model state $\mathbf{x}_\nu$ at lead time t_ν.

The residuals

$$\mathbf{e}^i_\nu = \mathbf{x}^i_\nu - \hat{\mathbf{x}}_\nu \tag{12.27}$$

form a sample of the prediction error distribution and are unbiased as they have zero (ensemble) mean. The sample covariance matrix of the residuals, given by

$$\begin{aligned}\hat{\mathrm{C}}_\nu &= \frac{1}{M-1}\sum_{i=1}^{M} \mathbf{e}^i_\nu (\mathbf{e}^i_\nu)^T, \\ &= \frac{1}{M-1}\sum_{i=1}^{M} (\mathbf{x}^i_\nu - \hat{\mathbf{x}}_\nu)(\mathbf{x}^i_\nu - \hat{\mathbf{x}}_\nu)^T,\end{aligned} \tag{12.28}$$

is then an estimate of the prediction error covariance matrix.

Example 12.2 North Atlantic multidecadal predictability An example of the use of the predictive power was given in Schneider and Griffies (1999) using the ensemble simulation results of Griffies and Bryan (1997) regarding North Atlantic multidecadal variability. The data consist of twelve ensemble simulations over a period of 30 years in addition to the 200-year control simulation with the GFDL-R15 model. A principal component analysis of the dynamic topography of the control simulation provides the first two EOF patterns as shown in Fig. 12.7; the EOFs together explain about 35% of the variance of the dynamic topography field. EOF1represents variations in the

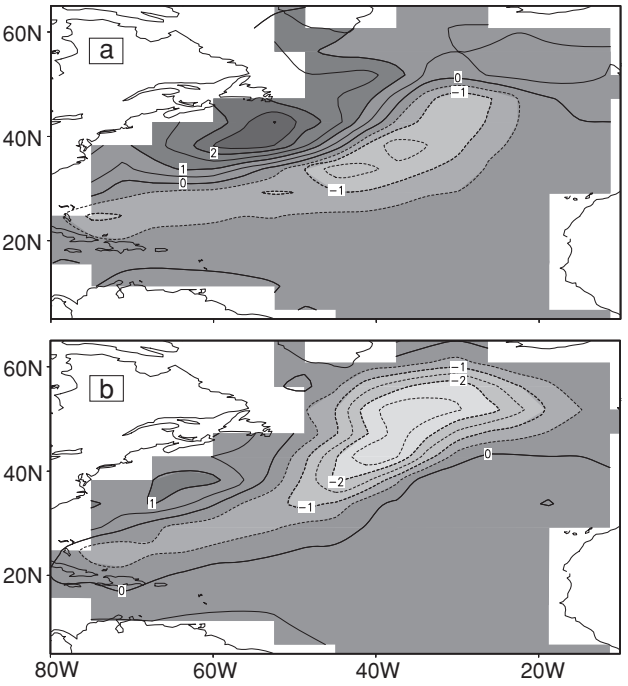


Figure 12.7 First (a) and second (b) EOF of North Atlantic dynamic topography (cm). The amplitude is scaled by the standard deviations of the associated principal components (figure from Schneider and Griffies, 1999).

strength of the North Atlantic Current's northeastward drift, whereas EOF2 displays gyre-shaped variations in the North Atlantic circulation, with the strongest current variations located in the central portion of the basin.

The data from the ensemble integrations are projected onto the space formed by the two EOFs, and estimates of covariance matrix $\hat{C}_\nu$ are obtained for each lead time $\nu = 1, \ldots, 30$ years. Fig. 12.8 shows the predictive power (here indicated as PP) as a function of lead time ν with 95% confidence intervals obtained by Monte Carlo simulation of 1,000 samples (cf. Section 5.2). The bias of the PP estimate is small enough that the PP can be considered significantly greater than zero when the 95% confidence interval does not include zero.

The null hypothesis is that the climatological covariance matrix Σ_ν and the prediction error covariance matrix $\hat{C}_\nu$ are equal, in which case there are no predictable components in the state space of the two EOFs. This null hypothesis is rejected at the 95% significance level for PPs greater than 0.28 (which is the dash-dotted line in Fig. 12.8). The overall PP decays rapidly over the first ten years of the forecasting lead time, remains marginally significant up to about year 17 and becomes insignificant beyond year 17. ■

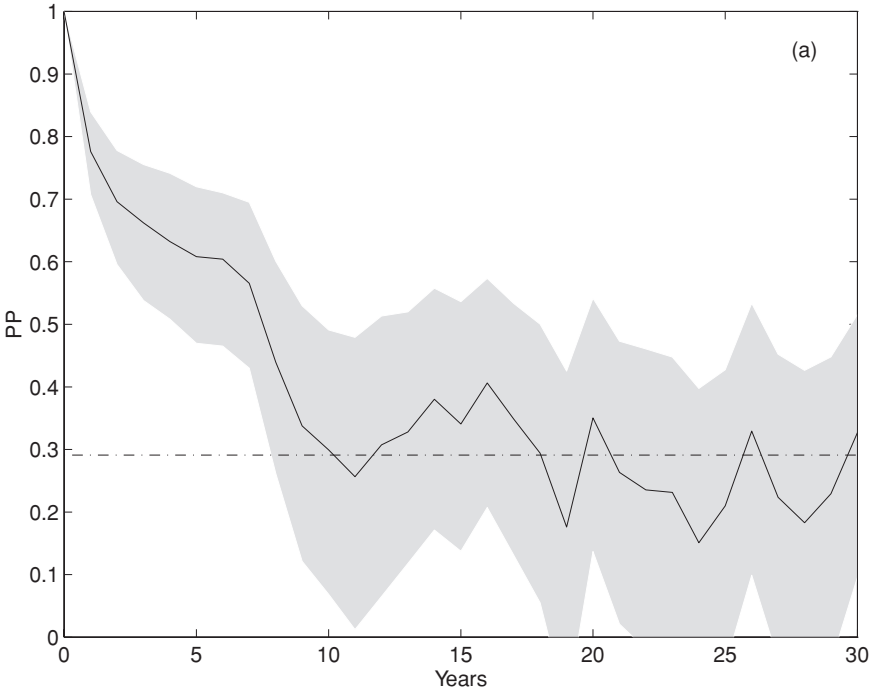


Figure 12.8 Overall predictive power (PP) for the first two EOFs for North Atlantic dynamic topography as a function of forecast lead time (figure from Schneider and Griffies, 1999).

12.3 Behaviour of nearby trajectories

From the previous sections, it is clear that the spread of trajectories for different initial conditions is crucial for predictability of the first kind. The predictive power measure is based on this spread. One of the central measures of the spread of trajectories is the Lyapunov exponent of the model (cf. Section 2.5). However, these exponents are not easy to determine for complicated climate models. What is usually only available is a climate model with specific choices for parameters for which we have to specify initial conditions and then integrate these in time. In this section, we discuss methods addressing the spread of trajectories in such climate models.

12.3.1 Linear view: singular vectors

We write the model equations of an autonomous dynamical system as

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}), \tag{12.29}$$

and indicate the solution $\mathbf{x}$ at time t of a trajectory starting at $\mathbf{x}(t_0)$ for time t_0 as (Kalnay, 2003)

$$\mathbf{x}(t) = \mathcal{M}(\mathbf{x}(t_0)), \tag{12.30}$$

where $\mathcal{M}$ is often referred to as the propagator.

A nearby trajectory $\mathbf{y}$ can be determined by looking at the evolution of the initial conditions $\mathbf{x}(t_0) + \mathbf{y}(t_0)$, where, for example, $\mathbf{y}(t_0) = \epsilon\mathbf{v}$ with $||\mathbf{v}|| = 1$ and $\epsilon \ll 1$. By Taylor expansion, we find

$$\mathcal{M}(\mathbf{x}(t_0) + \mathbf{y}(t_0)) = \mathcal{M}(\mathbf{x}(t_0)) + \frac{\partial \mathcal{M}}{\partial \mathbf{x}}\mathbf{y}(t_0) + \mathcal{O}(\epsilon^2) \approx \mathbf{x}(t) + \mathbf{y}(t), \tag{12.31}$$

where $\mathbf{y}$ satisfies

$$\frac{d\mathbf{y}}{dt} = \mathbf{J}\,\mathbf{y}, \tag{12.32}$$

with $\mathbf{J}$ is the Jacobian matrix of $\mathbf{F}$. The linear system of equations (12.32) is called the tangent linear model, and its solutions are indicated by (using the notation $\mathcal{L}(t_0, t) = \partial\mathcal{M}/\partial\mathbf{x}$)

$$\mathbf{y}(t) = \mathcal{L}(t_0, t)\mathbf{y}(t_0), \tag{12.33}$$

indicating the propagation of the initial perturbation $\mathbf{y}(t_0)$ to the perturbation $\mathbf{y}(t)$ at time t.

When the interval (t_0, t) is split into two intervals (t_0, t_1) and (t_1, t), the evolution of the perturbation is given by (integrating first up to t_1 and then to t)

$$\mathbf{y}(t) = \mathcal{L}(t_1, t)\mathcal{L}(t_0, t_1)\mathbf{y}(t_0) \rightarrow \mathcal{L}(t_0, t) = \mathcal{L}(t_1, t)\mathcal{L}(t_0, t_1). \tag{12.34}$$

With the standard inner product $\langle , \rangle$ on $\mathbb{R}^n$, the adjoint of the tangent linear model $\mathcal{L}^T$ is defined as $\langle \mathcal{L}\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, \mathcal{L}^T\mathbf{w}\rangle$, and it follows that

$$\mathcal{L}^T(t_0, t) = (\mathcal{L}(t_1, t)\mathcal{L}(t_0, t_1))^T = \mathcal{L}^T(t_0, t_1)\mathcal{L}^T(t_1, t). \tag{12.35}$$

This shows that the adjoint tangent linear model can be viewed as starting at time t and following the trajectory backwards in time.

To determine the action of the (adjoint) tangent linear model on vectors, we use the notation $\mathbf{L} \equiv \mathcal{L}(t_0, t_1)$. For any matrix $n \times n$ matrix $\mathbf{L}$, there exist unitary matrices $\mathbf{U}$ and $\mathbf{V}$ (with $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$, the identity matrix) such that

$$\mathbf{U}^T\mathbf{L}\mathbf{V} = \mathbf{S}, \tag{12.36}$$

where $\mathbf{S}$ is a diagonal matrix containing the singular values $\sigma_i, i = 1, \ldots n$ of $\mathbf{L}$.

Multiplying (12.36) from the left with $\mathbf{U}$ gives

$$\mathbf{L}\mathbf{V} = \mathbf{U}\mathbf{S} \rightarrow \mathbf{L}\mathbf{v}_i = \sigma_i\mathbf{u}_i \tag{12.37}$$

and multiplying (12.36) from the right with $\mathbf{V}^T$ gives, similarly,

$$\mathbf{U}^T\mathbf{L} = \mathbf{S}\mathbf{V}^T \rightarrow \mathbf{L}^T\mathbf{U} = \mathbf{V}\mathbf{S}^T = \mathbf{V}\mathbf{S} \rightarrow \mathbf{L}^T\mathbf{u}_i = \sigma_i\mathbf{v}_i. \tag{12.38}$$

The $\mathbf{v}_i$ are called the right (or initial) singular vectors and the $\mathbf{u}_i$ the left (or final) singular vectors. Combining both (12.38) and (12.37) gives

$$\mathbf{L}^T\mathbf{L}\mathbf{v}_i = \sigma_i\mathbf{L}^T\mathbf{u}_i = \sigma_i^2\mathbf{v}_i, \tag{12.39}$$

and hence the initial singular vectors can be determined as eigenvectors of $\mathbf{L}^T\mathbf{L}$.

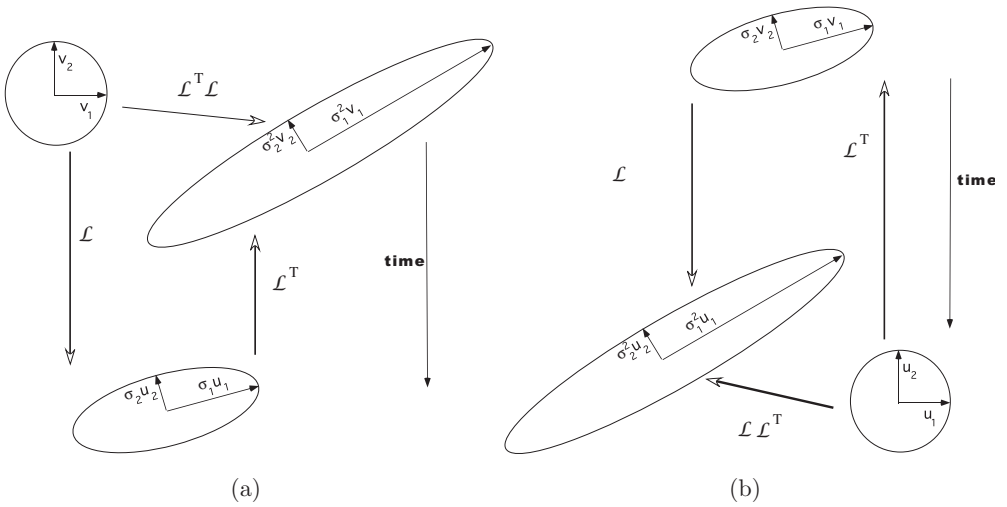


Figure 12.9 (a) Application of the tangent linear model forward in time and the adjoint tangent linear model backward in time to a two-dimensional sphere at initial time. (b) Application of the adjoint tangent linear model backward in time and the tangent linear model forward in time to a two-dimensional sphere at final time (figure based on Kalnay, 2003).

By applying $\mathbf{L}$ to each initial singular vector $\mathbf{v}_i$, the norm of this vector is changed by a factor σ_i, and its direction will be rotated to that of the vector $\mathbf{u}_i$. Similarly, by applying $\mathbf{L}^T$ to each $\mathbf{u}_i$, the norm of this vector is changed by a factor σ_i, and its direction will be rotated to that of the vector $\mathbf{v}_i$ (Fig. 12.9). If $\mathbf{L}$ is applied first to $\mathbf{v}_i$ and then $\mathbf{L}^T$ to $\mathbf{u}_i$, the norm of the vector $\mathbf{v}_i$ is changed by a factor σ_i^2. As the σ_i are real and ordered as $\sigma_1 > \sigma_2 > \ldots > \sigma_n$, the norm of the vector $\mathbf{v}_1$ is affected most.

Having understood the action of $\mathbf{L}$ and $\mathbf{L}^T$, we can now consider the behaviour of nearby trajectories associated with the initial perturbation vector $\mathbf{y}(t_0)$. These can behave very differently, and for the prediction problem, we are in particular interested in which directions; the distance between $\mathbf{x}(t)$ and the trajectory starting at $\mathbf{x}(t_0) + \mathbf{y}(t_0)$ is maximal, say in the norm $\|,\|$ associated with standard inner product, at time $t = t_1$. Using the tangent linear model, with $\mathbf{y}(t_1) = \mathbf{L}\mathbf{y}(t_0)$, we find

$$\|\mathbf{y}(t_1)\|^2 = \langle \mathbf{L}\mathbf{y}(t_0), \mathbf{L}\mathbf{y}(t_0)\rangle = \langle \mathbf{L}^T\mathbf{L}\mathbf{y}(t_0), \mathbf{y}(t_0)\rangle, \tag{12.40}$$

which suggests that this optimisation problem is solved by the singular vectors, and hence the first singular vector is also called the first optimal vector.

In Xue et al. (1997), the singular values and vectors are determined from the tangent linear model of the Zebiak-Cane ENSO model (Section 8.2). The method of constructing the tangent linear model makes use of principle component reduction and a large ensemble of two-year simulations. In Fig. 12.10, the first singular vector optimised over a time interval of six months from January of model year 26 is shown. The thermocline is deeper over much of the western part of the basin, and the westerly

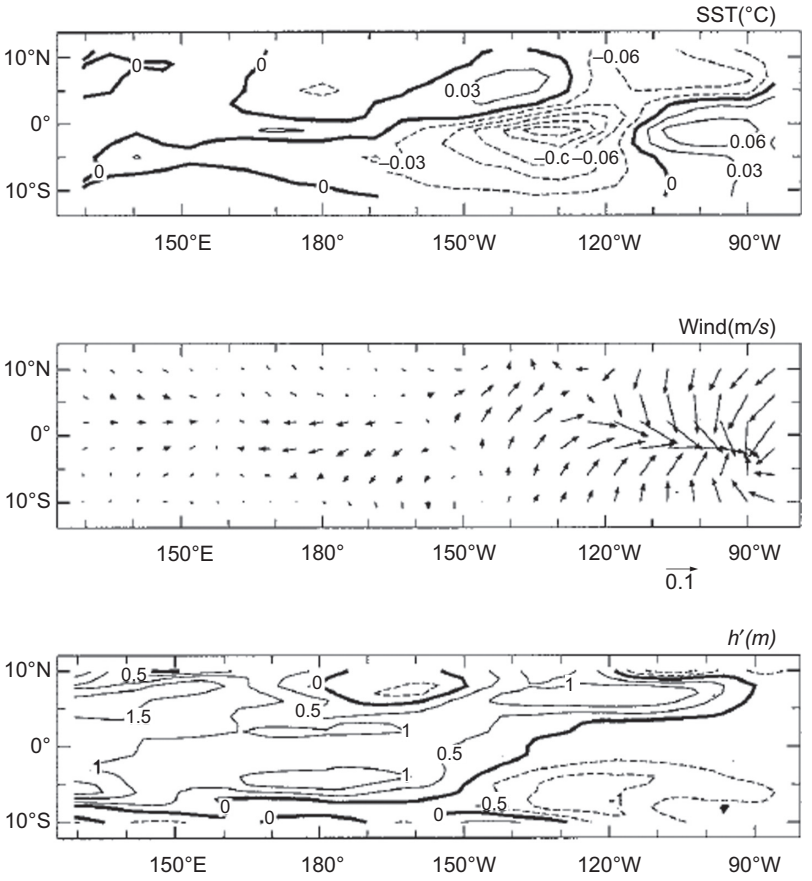


Figure 12.10 The first singular vector (SST, wind and thermocline depth) of the local TLM of the Zebiak-Cane model of ENSO optimised at 6 months from January of year 26 (figure from Xue et al., 1997).

wind anomalies of the singular vector are dominant in the eastern Pacific. With this perturbation on the model state, it develops into a substantial El Niño–like anomaly pattern after six months, with SST anomalies reaching up to 2°C (Xue et al., 1997).

12.3.2 Nonlinear view: CNOPs

When it is aimed to take nonlinear aspects of the spread of trajectories into account, one has to resort to nonlinear optimisation methods, maximising optimal growth of some norm of the solution. To study nonlinear mechanisms of trajectory divergence, Mu (2000) proposed the concept of nonlinear singular vectors and nonlinear singular values, and it was applied in Mu and Wang (2001) to shallow-water flows. In Mu and

Duan (2003), the concept of the conditional nonlinear optimal perturbation (CNOP) was introduced, which we discuss next.

In general, assume that the equations governing the evolution of perturbations can be written as

$$\begin{cases} \dfrac{\partial \boldsymbol{x}}{\partial t} + F(\boldsymbol{x}; \bar{\boldsymbol{x}}) = 0, & \text{in} \quad \Omega \times [0, t_e], \\ \boldsymbol{x}|_{t=0} = \boldsymbol{x}_0, \end{cases} \tag{12.41}$$

where t is time, $\boldsymbol{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))$ is the perturbation state vector on a basic state $\bar{\mathbf{x}}$ and F is a nonlinear differentiable operator. Furthermore, $\boldsymbol{x}_0$ is the initial perturbation, $(\boldsymbol{x}, t) \in \Omega \times [0, t_e]$ with Ω a domain in R^n, and $t_e < +\infty$.

Suppose the initial value problem (12.41) is well posed, and the nonlinear propagator $\mathcal{M}$ is defined as the evolution operator of (12.41), which determines a trajectory from the initial time $t = 0$ to time t_e. Hence, for fixed $t_e > 0$, the solution $\boldsymbol{x}(t_e) = \mathcal{M}(\boldsymbol{x}_0; \bar{\boldsymbol{x}})(t_e)$ is well defined, that is,

$$\boldsymbol{x}(t_e) = \mathcal{M}(\boldsymbol{x}_0; \bar{\boldsymbol{x}})(t_e). \tag{12.42}$$

For a chosen norm $\|\cdot\|$, the perturbation $\boldsymbol{x}_{0\delta}$ is called the conditional nonlinear optimal perturbation (CNOP) with constraint condition $\|\boldsymbol{x}_0\| \le \delta$ if and only if

$$J(\boldsymbol{x}_{0\delta}) = \max_{\|\boldsymbol{x}_0\| \le \delta} J(\boldsymbol{x}_0), \tag{12.43}$$

where the 'cost function' J is given by

$$J(\boldsymbol{x}_0) = \|\mathcal{M}(\boldsymbol{x}_0; \bar{\boldsymbol{x}})(t_e)\|. \tag{12.44}$$

The CNOP is the initial perturbation whose nonlinear evolution attains the maximal value of the functional J at time t_e with the constraint conditions. The CNOP can be regarded as the most (nonlinearly) unstable initial perturbation superposed on the basic state.

In Mu et al. (2004), the CNOP approach was applied to the two-box model (Stommel, 1961) of the thermohaline circulation described by the dimensionless equations (cf. Example 6.1)

$$\frac{dT}{dt} = \eta_1 - T(1 + \mid T - S \mid), \tag{12.45a}$$

$$\frac{dS}{dt} = \eta_2 - S(\eta_3 + \mid T - S \mid), \tag{12.45b}$$

where $T = T_e - T_p$, $S = S_e - S_p$ are the dimensionless temperature and salinity difference between the equatorial and polar box and $\Psi = T - S$ is the dimensionless meridional overturning streamfunction. The initial perturbation is written as $\boldsymbol{x}_0 = (T_0', S_0') = (\delta \cos\theta, \delta \sin\theta)$, where δ indicates the magnitude and θ the direction of the vector.

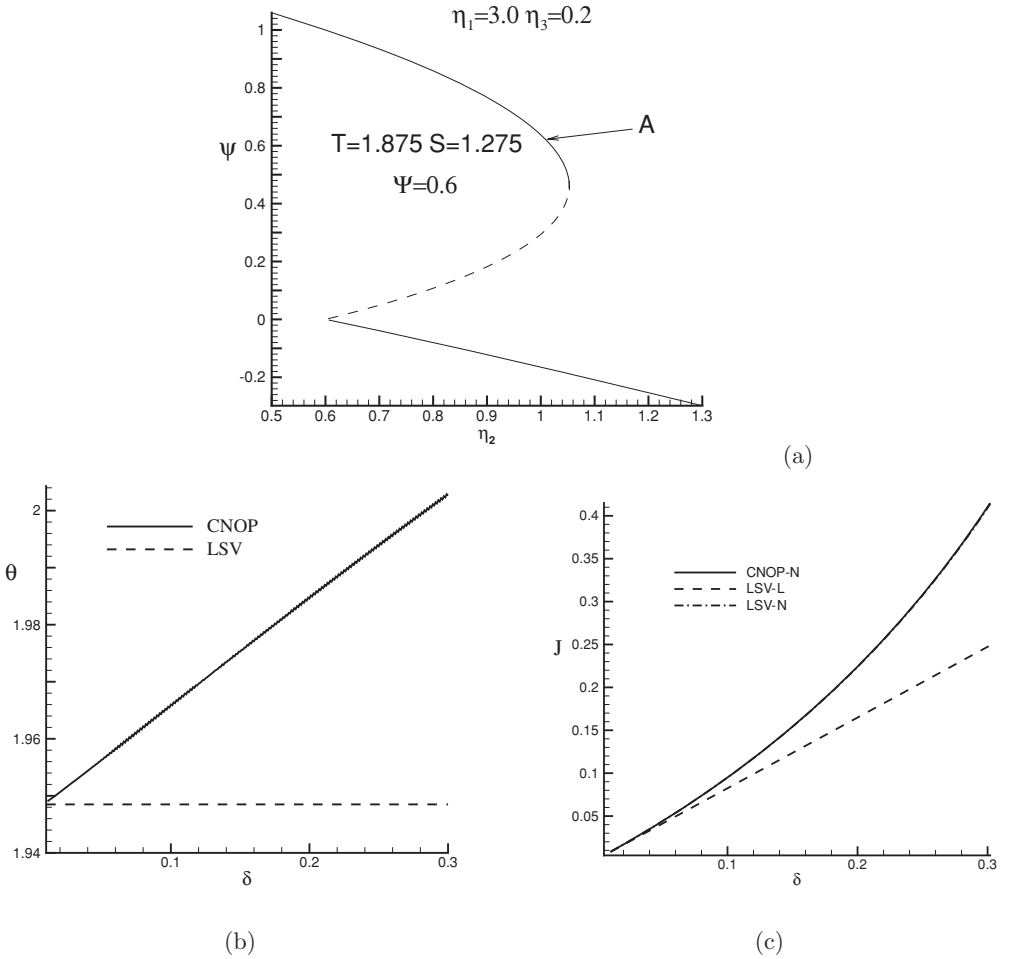


Figure 12.11 (a) Bifurcation diagram for the two-box model (12.45) with η_2 as control parameter and $\eta_1 = 3.0$, $\eta_3 = 0.2$. (b) Values of θ the CNOP and the first linear singular vector (LSV) for the two-box model. (c) Evolution of the cost function J of the time-dependent solution (figure from Mu et al., 2004).

For a thermally dominant (stable) steady state, the state $\bar{T} = 1.875$, $\bar{S} = 1.275$, $\bar{\Psi} = 0.6$ for $\eta_2 = 1.02$ as indicated as point A in Fig. 12.11a is chosen. Furthermore, $\delta = 0.3$ is used as a maximum amplitude of the perturbations. The time $t_e = 2.5$ is about half the time the solution takes to equilibrate to steady state from a particular initial perturbation. The amplitude $\delta = 0.3$ is about 10% of the typical amplitude of the steady state of temperature and salinity $(\bar{T}, \bar{S})$. For θ in the range $\pi/4 < \theta < 5\pi/4$, the initial perturbation flow has $\Psi'(0) < 0$ and hence weakens the thermally dominated flow. For other values of θ, the initial perturbation flow has $\Psi'(0) > 0$, which strengthens the thermally dominated flow.

The results for the CNOP and the first linear singular vectors (LSV) are shown in Fig. 12.11b. The directions of the LSVs, being independent of δ, have constant values of $\theta_1 = 1.948$ (dashed line) and $\theta_2 = 5.089$ (not shown). The directions of the CNOPs (solid curve) increase monotonously, with δ varying from 0.01 to 0.3. The difference between the CNOP and (first) LSV is relatively small when δ is small. Integrating the model with CNOPs and LSVs as initial conditions, respectively, we obtain their value at time t_e, which are denoted as CNOP-N and LSV-N in Fig. 12.11c. To make a comparison, the linear evolution (by the linearised model) of LSVs is also shown (LSV-L) in Fig. 12.11c. It is clear that CNOPs increase nonlinearly as the initial perturbation constraint increases, whereas LSV-Ls only increases linearly. The line of LSV-N is between CNOP-N and LSV-L, but the difference between LSV-N and CNOP-N is hardly distinguishable in Fig. 12.11c.

12.3.3 Nonlinear view: Lyapunov techniques

In the previous sections, all the measures of growth or decay of vectors along a trajectory were dependent on the norm chosen, and hence the results can only be coupled to the physics of the problem when combined with other information of the model system or observations. As we saw in Section 2.4.3, the Lyapunov exponents of an attractor indicate whether exponential divergence of trajectories occurs. When there is a positive Lyapunov exponent, there is sensitivity to initial conditions. These Lyapunov exponents describe the long-term exponential rate of stretching or contraction in the attractor and are norm independent. They reduce to the Floquet multipliers when the trajectory is periodic and to the growth factors of the normal modes when the trajectory is steady. Lyapunov vectors $\boldsymbol{\phi}_i$ are the generalisation of normal mode vectors of a stable steady state and Floquet vectors for a stable periodic orbit to a general time-dependent trajectory. In this section, we discuss methods to obtain information on the Lyapunov vectors directly from the model simulations.

Direct computation

From the Oseledec theorem (see Section 5.5.2), it follows that the Lyapunov exponents $\lambda^{\pm}$, which characterise the asymptotic evolution of linear disturbances to bounded trajectories of arbitrary time dependence, can be calculated as $\lambda_i^{\pm} = \ln s_{\pm}$, where $s_{\pm}$ are the eigenvalues of the matrices

$$\mathcal{S}_{\pm} = \lim_{t \to \pm\infty} (\mathcal{L}^T(t_0, t)\mathcal{L}(t_0, t))^{\frac{1}{2(t-t_0)}}, \tag{12.46}$$

where $\mathcal{L}(t_0, t)$ is again the tangent linear model. The eigenvalues are norm independent, independent of the initial time t_0 and $\lambda^- = -\lambda^+$. A norm-independent set of Lyapunov vectors $\boldsymbol{\phi}_i$, such that $\boldsymbol{\phi}_i$ grows at a rate $\lambda_i^{\pm}$ as $t \to \pm\infty$, can also be defined using the Oseledec theorem (Eckmann and Ruelle, 1985) using nested subspaces and

can be shown to reduce to Floquet vectors and normal modes when the trajectory is periodic and steady, respectively.

As can be anticipated from the discussion on singular vectors in the Section 12.3.1, the set of singular vectors with large optimisation intervals $t - t_0$ are orthogonalisations of the Lyapunov vectors. Let the evolution of the initial singular vectors $\mathbf{v}_j(t_1, t_2)$ using an optimisation interval $t_2 - t_1$ be indicated by

$$\boldsymbol{\xi}_j(t; t_1, t_2) = \mathcal{L}(t_1, t)\mathbf{v}_j(t_1, t_2), \tag{12.47}$$

then the backward singular vectors $\hat{\boldsymbol{\eta}}_j(t)$ and forward singular vectors $\hat{\hat{\boldsymbol{\xi}}}_j(t)$ are defined as

$$\hat{\boldsymbol{\eta}}_j(t) = \lim_{t_1 \to -\infty} \boldsymbol{\xi}_j(t; t_1, t), \tag{12.48a}$$

$$\hat{\hat{\boldsymbol{\xi}}}_j(t) = \lim_{t_2 \to \infty} \boldsymbol{\xi}_j(t; t, t_2). \tag{12.48b}$$

To determine a backwards singular vector, the evolved singular vector that was initialised in the distant past is determined at time t. Similarly, the forward singular vector is determined by the initial singular vector with an optimisation time far into the future. The relation between these singular vectors and the Lyapunov vectors is given by

$$\hat{\boldsymbol{\eta}}_j(t) = \sum_{i=1}^{j} \hat{p}_{ji}\boldsymbol{\phi}_i(t), \tag{12.49a}$$

$$\hat{\hat{\boldsymbol{\xi}}}_j(t) = \sum_{i=j}^{N} \hat{q}_{ji}\boldsymbol{\phi}_i(t), \tag{12.49b}$$

with coefficients $\hat{p}_{ji}$ and $\hat{q}_{ji}$ and where the Lyapunov vectors are ordered with respect to the magnitude of the Lyapunov exponents.

The relations (12.49) provide a method to compute the Lyapunov vectors (Wolfe and Samelson, 2007). It follows that $\langle \boldsymbol{\phi}_i, \hat{\hat{\boldsymbol{\xi}}}_j \rangle = 0$ for $i < j$ and $\langle \boldsymbol{\phi}_i, \hat{\boldsymbol{\eta}}_j \rangle = 0$ for $i > j$, and hence the $\boldsymbol{\phi}_n$ can be expressed into the vectors $\hat{\hat{\boldsymbol{\xi}}}_j$ and $\hat{\boldsymbol{\eta}}_j$ through

$$\boldsymbol{\phi}_n = \sum_{i=n}^{N} \langle \boldsymbol{\phi}_n, \hat{\hat{\boldsymbol{\xi}}}_i \rangle \hat{\hat{\boldsymbol{\xi}}}_i, \tag{12.50a}$$

$$\boldsymbol{\phi}_n = \sum_{j=1}^{n} \langle \boldsymbol{\phi}_n, \hat{\boldsymbol{\eta}}_j \rangle \hat{\boldsymbol{\eta}}_j, \tag{12.50b}$$

where the dependence on time has been suppressed. Taking the inner product of (12.50a) with $\hat{\boldsymbol{\eta}}_k$ and of (12.50b) with $\hat{\hat{\boldsymbol{\xi}}}_k$ and eliminating the $\langle \hat{\hat{\boldsymbol{\xi}}}_k, \boldsymbol{\phi}_n \rangle$ term leads to

the linear system of equations

$$\langle \hat{\boldsymbol{\eta}}_k, \boldsymbol{\phi}_n \rangle = \sum_{j=1}^{n} \left[\sum_{i=n}^{N} \langle \hat{\boldsymbol{\eta}}_k, \hat{\boldsymbol{\xi}}_i \rangle \langle \hat{\boldsymbol{\xi}}_i, \hat{\boldsymbol{\eta}}_j \rangle \right] \langle \hat{\boldsymbol{\eta}}_j, \boldsymbol{\phi}_n \rangle \ , \ k \leq n. \tag{12.51}$$

The solution of this system of equations provides the expansion coefficients of the Lyapunov vectors in terms of the backwards singular vectors.

In Wolfe and Samelson (2007), it was realised that because the $\hat{\boldsymbol{\eta}}_i$ and $\hat{\boldsymbol{\xi}}_j$ are orthonormal sets of vectors, it holds that

$$\sum_{i=1}^{N} \langle \hat{\boldsymbol{\eta}}_k, \hat{\boldsymbol{\xi}}_i \rangle \langle \hat{\boldsymbol{\xi}}_i, \hat{\boldsymbol{\eta}}_j \rangle = \delta_{kj}, \tag{12.52}$$

and hence using this in (12.51), it follows that

$$\sum_{j=1}^{n} \left[\sum_{i=1}^{n-1} \langle \hat{\boldsymbol{\eta}}_k, \hat{\boldsymbol{\xi}}_i \rangle \langle \hat{\boldsymbol{\xi}}_i, \hat{\boldsymbol{\eta}}_j \rangle \right] \langle \hat{\boldsymbol{\eta}}_j, \boldsymbol{\phi}_n \rangle = 0 \ , \ k \leq n. \tag{12.53}$$

With the notation $y_k^n = \langle \hat{\boldsymbol{\eta}}_k, \boldsymbol{\phi}_n \rangle, k = 1, \ldots, n$ and

$$D_{kj}^n = \sum_{i=1}^{n-1} \langle \hat{\boldsymbol{\eta}}_k, \hat{\boldsymbol{\xi}}_i \rangle \langle \hat{\boldsymbol{\xi}}_i, \hat{\boldsymbol{\eta}}_j \rangle, \ k, j \leq n \tag{12.54}$$

(12.53) can be written as (with D indicating the matrix with elements D_{kj})

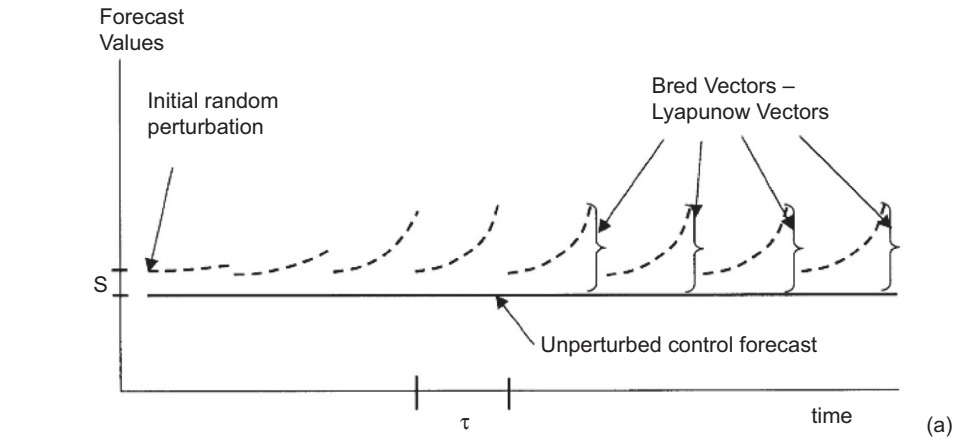
$$\mathrm{D}^n \mathbf{y}^n = 0. \tag{12.55}$$

The Lyapunov vectors can hence be determined from only the first $n-1$ forward singular vectors and n backward singular vectors. As we saw in Section 12.2.3, these singular vectors can be obtained from a singular value decomposition for large optimisation times.

Breeding techniques

A method to determine the leading Lyapunov vector that is often applied in operational practice is the calculation of so-called Bred vectors. The original idea (Toth and Kalnay, 1997) was to construct an ensemble of optimal perturbations to carry out an ensemble forecast by selecting the most important growing error. This essentially nonlinear method consists of the following steps (Fig. 12.12a):

(i) Add to the initial state $\mathbf{x}(t_0)$ a small random perturbation $\delta\mathbf{x}(t_0)$ to obtain a perturbed initial state $\tilde{\mathbf{x}}(t_0)$.
(ii) Integrate the model from both the unperturbed and the perturbed initial state for a given time $T = t_1 - t_0$.
(iii) Measure the distance $||\delta\mathbf{x}(t_1)||$ between the two trajectories at time t_1 and rescale this distance to have the same size as the initial one $||\delta\mathbf{x}(t_0)||$.



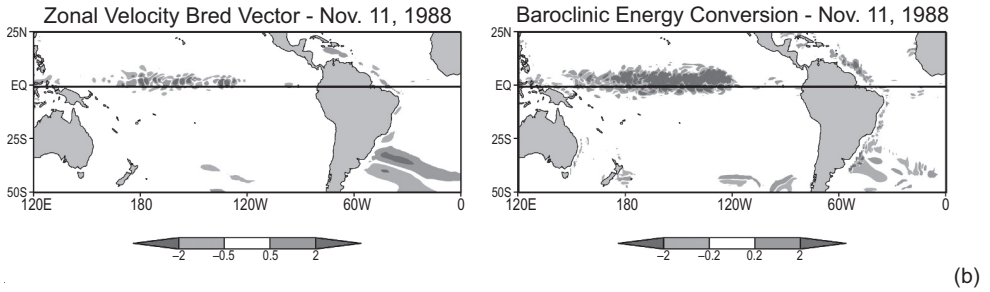


Figure 12.12 (a) Sketch of the procedure to determine Bred vectors. (b) Bred vector (*left panel*) of the zonal velocity starting at 11 November 1988 from the GFDL MOM2 model using a ten-day breeding interval and corresponding baroclinic energy conversion (figure from Hoffman et al., 2009). (See Colour Plate.)

(iv) Add this rescaled perturbation to the 'control simulation' $\mathbf{x}(t_1)$ and repeat from step (ii) until convergence.

After several rescaling steps, the perturbation evolves towards the leading Lyapunov vector (Toth and Kalnay, 1997), and hence the method 'breeds' the nonlinear perturbation that grows fastest. An example of a Bred vector determined from simulations with a global version of the GFDL MOM2 model ($1^\circ \times 1^\circ$ horizontal resolution in midlatitudes reducing to $1^\circ \times 1/2^\circ$ near the equator) with twenty vertical levels and forced by the NCEP reanalysis data was presented in Hoffman et al. (2009). The zonal velocity of the Bred vector for a ten-day breeding interval (Fig. 12.12b) starting on 11 November 1988 shows a dipole pattern near South America and a wave pattern in the Tropical Pacific; the latter could be identified with tropical instability waves (Hoffman et al., 2009).

12.4 Data assimilation

The path of the trajectories of climate models can be 'improved' by combining the model outcomes with available observations in so-called data-assimilation methods. Central to this improvement is an adequate state estimation. Here a distinction is made among a so-called filter, a smoother and a predictor. Suppose that discrete measurements $x_1, \ldots, x_n$ are available, and we are interested in a state estimation $\hat{x}_k$. For a predictor we use only the measurements up to x_{k-1}, for a filter we use measurements up to x_k and for a smoother measurements in the future ($x_l, l > k$) are also used.

There are basically two classes of data-assimilation methods: (i) those based on optimisation methods using variational techniques and (ii) those based on ensemble approaches. The central ideas of all data assimilation methods, however, go back to those behind the Kalman filter, which we therefore discuss in detail in the next subsection.

12.4.1 The Kalman Filter

To present the Kalman filter in its simplest form, a one-dimensional linear stochastic discrete dynamical system of the form, with $x_j = x(t_j)$,

$$x_j = cx_{j-1} + w_j, \tag{12.56}$$

is considered, with $c \in \mathbb{R}$ and w_j representing Gaussian noise having a $N(0, \sigma_w)$ distribution. Suppose that we measure the state of the system through the output (Fig. 12.13a)

$$y_j = hx_j + v_j, \tag{12.57}$$

where $h \in \mathbb{R}$ and v_j again represent Gaussian noise with a $N(0, \sigma_v)$ distribution; v_j and w_j are independent.

The question that is addressed by the Kalman filter is the following: can we use the y_j to optimally (to minimise the effect of the noise) determine an estimate $\hat{x}_j$ of the state x_j?

Assuming that an estimate $\hat{x}_{j-1}$ is available (Fig. 12.13b), the Kalman filter at time t_j proceeds as follows. First, an a priori estimate $\hat{x}_j^-$ is taken as

$$\hat{x}_j^- = c\hat{x}_{j-1} \tag{12.58}$$

and used to estimate the output $\hat{y}_j = h\hat{x}_j^-$. The difference between this estimate and the measured y_j is, in a second step, used to correct the a priori estimate $\hat{x}_{j-1}^-$ according to

$$\hat{x}_j = \hat{x}_j^- + k_j(y_j - \hat{y}_j) = \hat{x}_j^- + k_j(y_j - h\hat{x}_j^-), \tag{12.59}$$

where k_j is the Kalman gain at time t_j.

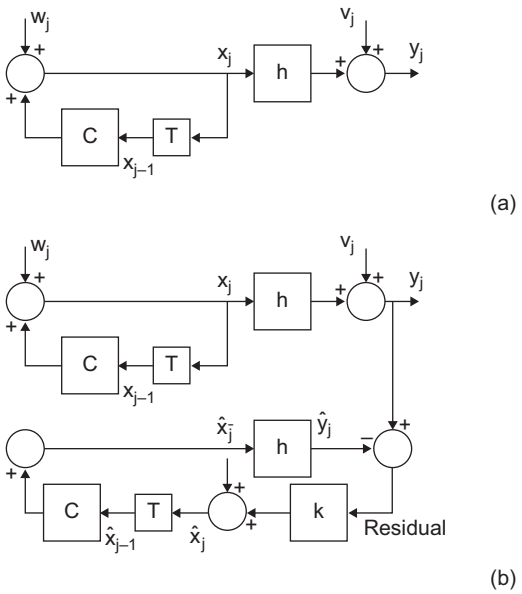


Figure 12.13 (a) Block diagram of the original system where the block 'T' indicates the time delay (slightly modified from http://www.swarthmore.edu/NatSci/echeeve1/Ref/Kalman/ScalarKalman.html). (b) Block diagram of the Kalman filter construction.

To determine k_j, a priori e_j^- and a posteriori e_j errors and variances p_j^- and p_j are defined as

$$e_j^- = x_j - \hat{x}_j^- \;\; ; \;\; p_j^- = E[(e_j^-)^2], \tag{12.60a}$$

$$e_j = x_j - \hat{x}_j \;\; ; \;\; p_j = E[(e_j)^2], \tag{12.60b}$$

where, as in Chapter 3, $E[x]$ denotes the expectation value of x. The a priori error variance can be written as (using 12.56)

$$p_j^- = E[(x_j - \hat{x}_j^-)^2] = E[(cx_{j-1} + w_j - \hat{x}_j^-)^2] = \\ E[(cx_{j-1} + w_j - c\hat{x}_{j-1})^2] = c^2 p_{j-1} + \sigma_w^2, \tag{12.61}$$

where the last equality arises because w_j is uncorrelated to both output and the a priori estimate. The variance p_{j-1} (computed at t_{j-1}) is known at time t_j.

In the Kalman filter, the gain k_j is determined such that p_j is minimised, hence, using (12.59),

$$\frac{\partial p_j}{\partial k_j} = \frac{\partial E[(x_j - \hat{x}_j)^2]}{\partial k_j} = \frac{\partial E[(x_j - \hat{x}_j^- + k_j(y_j - h\hat{x}_j^-))^2]}{\partial k_j} = 0, \tag{12.62}$$

from which it follows that

$$k_j = \frac{E[(y_j - h\hat{x}_j^-)(x_j - \hat{x}_j^-)]}{E[(y_j - h\hat{x}_j^-)^2]}. \tag{12.63}$$

This expression can be simplified using (12.57) according to

$$E[(y_j - h\hat{x}_j^-)(x_j - \hat{x}_j^-)] = E[h(x_j - \hat{x}_j^-)^2] + E[v_j(x_j - \hat{x}_j^-)] = hE[(e_j^-)^2] = hp_j^-,$$
$$E[(y_j - h\hat{x}_j^-)^2] = E[(h(x_j - \hat{x}_j^-) + v_j)^2] = h^2 p_j^- + \sigma_v^2,$$

because v_j is uncorrelated to both output and the a priori estimate. As a consequence, (12.63) reduces to

$$k_j = \frac{hp_j^-}{h^2 p_j^- + \sigma_v^2} = \frac{h(c^2 p_{j-1} + \sigma_w^2)}{h^2(c^2 p_{j-1} + \sigma_w^2) + \sigma_v^2}, \tag{12.64}$$

where in the last equation the expression (12.61) for p_j^- was used.

Finally, the a posteriori error variance p_j can be determined from

$$\begin{aligned} p_j &= E[(x_j - \hat{x}_j)^2] = E[(x_j - (\hat{x}_j^- + k_j(hx_j + v_j - h\hat{x}_j^-)))^2] \\ &= E[((x_j - \hat{x}_j^-)(1 - hk_j) - k_j v_j)^2] \\ &= (1 - hk_j)^2 p_j^- + k_j^2 \sigma_v^2. \end{aligned}$$

Expressing σ_v^2 into p_j^- using (12.64), that is,

$$\sigma_v^2 = p_j^- \frac{h(1 - hk_j)}{k_j}, \tag{12.65}$$

then eventually gives

$$p_j = (1 - hk_j)p_j^- = (1 - hk_j)(c^2 p_{j-1} + \sigma_w^2). \tag{12.66}$$

We now summarise the two-step process of the Kalman filter and simultaneously provide its multidimensional extension. In this case, the system and measurement is described by

$$\mathbf{x}_j = \mathbf{C}\mathbf{x}_{j-1} + \mathbf{W}_j, \tag{12.67a}$$
$$\mathbf{y}_j = \mathbf{H}\mathbf{x}_j + \mathbf{V}_j, \tag{12.67b}$$

with obvious extensions from the preceding one-dimensional case, and let Σ_W and Σ_V indicate the covariance matrices of the noise $\mathbf{W}$ and $\mathbf{V}$, respectively.

The predictor and corrector steps in the Kalman filter are then given by

$$\hat{\mathbf{x}}_j^- = \mathbf{C}\hat{\mathbf{x}}_{j-1}, \tag{12.68a}$$
$$\hat{\mathbf{x}}_j = \hat{\mathbf{x}}_j^- + \mathbf{K}_j(\mathbf{y}_j - \mathbf{H}\hat{\mathbf{x}}_j^-). \tag{12.68b}$$

The a priori and a posteriori error covariances are given by (compare with [12.60])

$$\mathbf{P}_j^- = E[(\mathbf{x} - \hat{\mathbf{x}}_j^-)(\mathbf{x} - \hat{\mathbf{x}}_j^-)^T], \tag{12.69a}$$

$$\mathbf{P}_j = E[(\mathbf{x} - \hat{\mathbf{x}}_j)(\mathbf{x} - \hat{\mathbf{x}}_j)^T], \tag{12.69b}$$

and $\mathbf{K}_j$ is determined by minimising $\mathbf{P}_j$ which results in (compare with [12.64])

$$\mathbf{K}_j = \frac{\mathbf{P}_j^- \mathbf{H}^T}{\mathbf{H}\mathbf{P}_j^- \mathbf{H}^T + \Sigma_V}, \tag{12.70}$$

where the a priori covariance matrix is found from (compare with [12.61])

$$\mathbf{P}_j^- = \mathbf{C}\mathbf{P}_{j-1}\mathbf{C}^T + \Sigma_W, \tag{12.71}$$

and finally the a posteriori covariance matrix is given by (compare with [12.66])

$$\mathbf{P}_j = (\mathbf{I} - \mathbf{K}_j\mathbf{H})\mathbf{P}_j^-. \tag{12.72}$$

In the climate data assimilation literature, one often uses a slightly different terminology. Here a background model state $\mathbf{x}^b$ (instead of $\hat{\mathbf{x}}^-$) serves as the prior estimate with covariance error matrix $\mathbf{B}$ (instead of $\mathbf{P}^-$). The estimated state is called the analysis $\mathbf{x}^a$ (instead of $\hat{\mathbf{x}}$), and the a posteriori covariance matrix is usually indicated by $\mathbf{A}$, the analysis errors (instead of $\mathbf{P}$). The covariance matrix of the observational errors is usually indicated by $\mathbf{R}$ (instead of Σ_V), and often the model error covariance Σ_W is neglected (perfect model). From the Kalman filter, the analysis $\mathbf{x}^a$ then follows from

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b), \tag{12.73}$$

and the Kalman-gain matrix $\mathbf{K}$ and the analysis errors follow from

$$\mathbf{K} = \frac{\mathbf{B}\mathbf{H}^T}{\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}}, \tag{12.74a}$$

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}. \tag{12.74b}$$

Although the Kalman filter is a basis for many of the data-assimilation methods in oceanography and meteorology (and hence its basic technique was presented in this section), the multidimensional version becomes prohibitively computationally expensive for large-dimensional dynamical systems, and hence many approximative Kalman filter methods are in use (Evensen, 2009).

12.4.2 Ensemble-based methods

As mentioned in Section 12.1, ensemble methods are often used in climate modelling. Here, many simulations are performed with different initial conditions, and using data assimilation techniques, observations can be used to limit the spread of the ensemble.

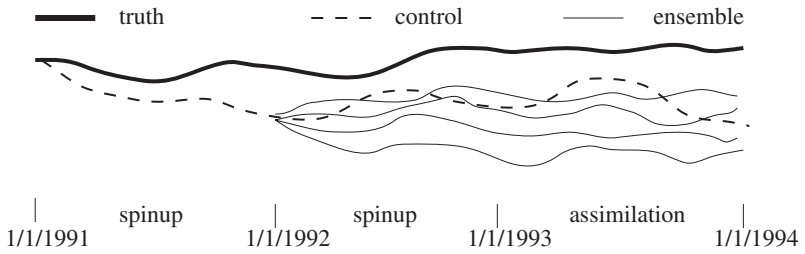


Figure 12.14 Schematic of the experiment set-up in Leeuwenburgh (2005).

Next, the state vector of ensemble member i is indicated by $\mathbf{x}_i$, and we indicate the ensemble mean operator with an overbar.

Ensemble Kalman Filter

In the Ensemble Kalman filter (EnKF) method with given observations $\mathbf{y}$ at a time t, these observations first are perturbed (Burgers et al., 1998) to obtain

$$\mathbf{y}_i = \mathbf{y} + \mathbf{e}_i \rightarrow \mathbf{R}_e = \overline{\mathbf{e}\mathbf{e}^T}, \tag{12.75}$$

and the analysis for each ensemble member follows from (12.73) as

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \mathbf{K}_e\left(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^b\right). \tag{12.76}$$

The Kalman gain matrix $\mathbf{K}_e$ is now computed using ensemble mean quantities, similar to (12.74),

$$\mathbf{K}_e = \frac{\mathbf{B}_e\mathbf{H}^T}{\mathbf{H}\mathbf{B}_e\mathbf{H}^T + \mathbf{R}_e}, \tag{12.77}$$

where

$$\mathbf{B}_e = \overline{(\mathbf{x}^b - \overline{\mathbf{x}^b})(\mathbf{x}^b - \overline{\mathbf{x}^b})^T}, \tag{12.78}$$

and the analysis error is computed from

$$\mathbf{A} = \overline{(\mathbf{x}^a - \overline{\mathbf{x}^a})(\mathbf{x}^a - \overline{\mathbf{x}^a})^T} = (\mathbf{I} - \mathbf{K}_e\mathbf{H})\mathbf{B}_e + \mathcal{O}(N^{-1/2}), \tag{12.79}$$

if N is the number of ensemble members. The EnKF is used in many prediction and data-assimilation studies in climate dynamics (Evensen, 2009), and a website with up-to-date information and codes can be found at http://enkf.nersc.no/.

An example of the use of the EnKF is provided in the identical twin study of Leeuwenburgh (2005), where SSH data are assimilated into a global ocean model. The 'truth' was created by forcing the ocean model with NCEP reanalysis fields, whereas during the control and assimilation runs, the model was forced by ERA40 fields (Fig. 12.14). SSH 'observations' from the truth were assimilated into along

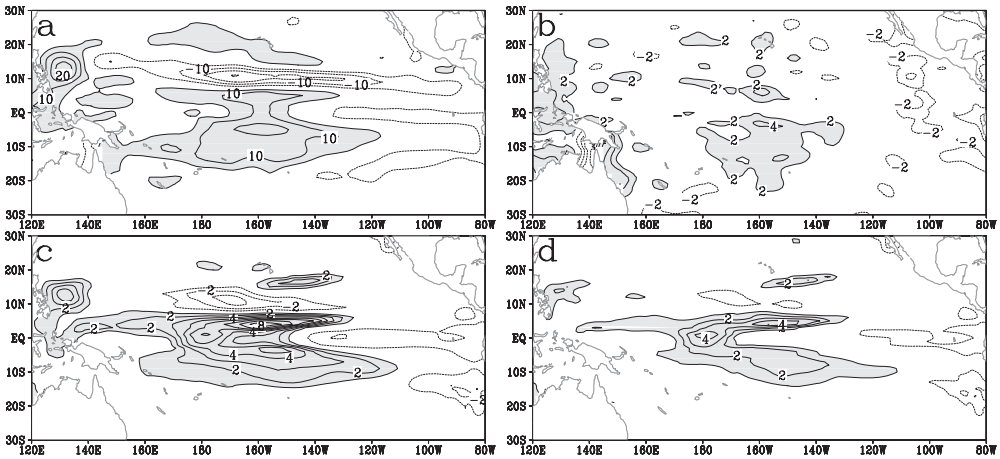


Figure 12.15 Errors in the (a–c) forecast (without data assimilation) and (b–d) analysis at 8 January 1993, for sea-level height (cm) (a–b) and temperature (°C) (c–d) at 50 depth (figure from Leeuwenburgh, 2005).

TOPEX/POSEIDON tracks over the equatorial Pacific every ten days. The results indicate that the assimilation of SSH leads to a significant improvement along the equator in all subsurface fields relative to the unconstrained control simulation (Fig. 12.15).

Particle filtering

With an ensemble of trajectories (or particles), we obtain at a certain time a sampling of the probability density function of the background (or prior estimate), and through the Kalman filter, we obtain an estimate of the probability density function of the analysis (or posterior estimate). When the model is nonlinear and/or the prior probability density function is non-Gaussian, the EnKF methodology is no longer appropriate. One of the possibilities is to use Bayes's theory to determine the posterior probability density function from the prior probability density function using the extra information provided by additional observations. This requires that the probability density functions are properly sampled and ensemble methods are ideally suited to do so; it leads to so-called particle filter methods (Van Leeuwen, 2009).

To illustrate the basics of a particle filter method, assume that the model is represented by the equations

$$\mathbf{x}_{k+1} = \mathcal{M}_k(\mathbf{x}_k, \mathbf{W}_k), \tag{12.80}$$

where $\mathbf{x}_k$ is the state vector at time t_k, $\mathcal{M}$ is the propagator and $\mathbf{W}_k$ is the noise in the model. At time t_k, the observation vector is indicated by $\mathbf{y}_k$. To determine the probability density function of the state vector $\mathbf{x}_k$, N independent particles or trajectories are available (from a model simulation over the interval $[t_{k-1}, t_k]$), which provide states $\mathbf{x}_k^i$, $i = 1, \ldots, N$. The probability density function of the model, $p_N(\mathbf{x}_k)$, can

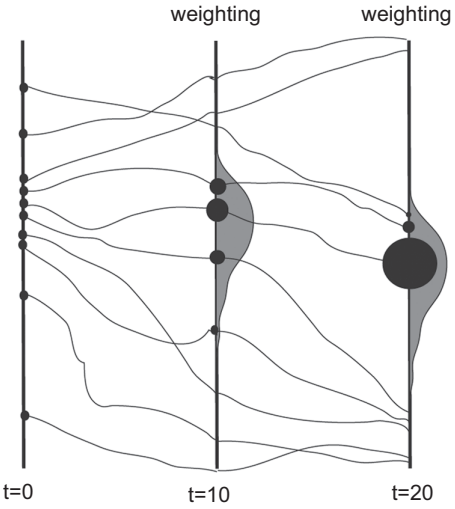


Figure 12.16 Schematic of the standard particle filter method with importance sampling (figure from Van Leeuwen, 2009).

then be represented by

$$p_N(\mathbf{x}_k) = \frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{x}_k - \mathbf{x}_k^i), \tag{12.81}$$

where δ is the Dirac function.

Suppose now that at an initial time $t = t_0$, N trajectories are computed. As observations are not yet taken into account, the particles are given an equal weight $w_0^i = 1/N$, as any of these trajectories could be close to later observations and $p_N(\mathbf{x}_0)$ is chosen (Fig. 12.16). After time $\Delta t = t_1 - t_0$, the prior probability density function of the model $p_N(\mathbf{x}_1)$ can be computed through (12.81), which can be written as

$$p_N(\mathbf{x}_1) = \sum_{i=1}^{N} w_0^i \delta\left(\mathbf{x}_1 - \mathbf{x}_1^i\right). \tag{12.82}$$

Next, the observations $\mathbf{y}_1$ are taken into account to compute a posterior probability density function. The joint probability of two events a and b can be written in two ways using conditional probabilities

$$P(a \in A, b \in B) = P(a \in A | b \in B)P(b \in B), \tag{12.83a}$$

$$P(a \in A, b \in B) = P(b \in B | a \in A)P(a \in A), \tag{12.83b}$$

and, combining these two relations, we find Bayes's theorem

$$P(a \in A | b \in B) = \frac{P(b \in B | a \in A)P(a \in A)}{P(b \in B)}. \tag{12.84}$$

In the particle filter methodology, Bayes's theorem is used in the form

$$p_N(\mathbf{x}_1|\mathbf{y}_1) = \frac{p_N(\mathbf{y}_1|\mathbf{x}_1)p_N(\mathbf{x}_1)}{p_N(\mathbf{y}_1)} = \sum_{i=1}^{N} w_1^i \delta\left(\mathbf{x}_1 - \mathbf{x}_1^i\right), \tag{12.85}$$

where the new weights are found from (using [12.82])

$$w_1^i = \frac{p_N(\mathbf{y}_1|\mathbf{x}_1)}{p_N(\mathbf{y}_1)} w_0^i. \tag{12.86}$$

The probability $p_N(\mathbf{y}_1|\mathbf{x}_1)$ of the observation $\mathbf{y}_1$ given the model state $\mathbf{x}_1$, is directly linked to the observational error. For example, with a univariate measurement y_1 having a Gaussian distribution with a variance σ_{obs}^2, the probability follows from

$$p_N(y_1|\mathbf{x}_1) \sim e^{-\frac{(H(\mathbf{x}_1)-y_1)^2}{2\sigma_{obs}^2}}, \tag{12.87}$$

where $H(\mathbf{x}_1)$ is the model equivalent of the observation calculated using the observational operator H. The probability $p_N(\mathbf{y}_1)$ serves as a normalisation that can be calculated by the constraint

$$\sum_{i=1}^{N} w_1^i = \sum_{i=1}^{N} \frac{p_N(\mathbf{y}_1|\mathbf{x}_1)}{p_N(\mathbf{y}_1)} w_0^i = 1. \tag{12.88}$$

Once the new weights have been determined, the posterior probability density function is known. In the particle filter procedure, the weights of the different particles increase when the trajectories are closer to observations (see Fig. 12.16).

A typical example, where horizontal mixing coefficients in a global ocean model are estimated from sea-level observations using a particle filter method, can be found in Vossepoel and Van Leeuwen (2007). Let the horizontal diffusion coefficient for heat and salt be indicated by K_H and the horizontal viscosity by A_H. In a 128-member ensemble starting from the same initial conditions, the ocean model is integrated forward in time. For each ensemble member, the diffusivity and viscosity are taken as $A_H = cA_H^0$ and $K_H = cK_H^0$, where c is chosen from a uniform distribution on the interval (0, 1) and A_H^0 and K_H^0 are standard values.

From the simulations, the global sea-surface height field is determined, and the spread of this quantity in the ensemble is shown in Fig. 12.17a. In the regions where the spread of the ensemble is largest, the sea level is most sensitive to differences in lateral mixing parameterization. Synthetic sea-level observations are next taken from a simulation of the model with a value $c = 0.2$ (and also from $c = 0.8$) and random uncorrelated noise has been added with a standard deviation of 5 cm. These 'observations' are assimilated using the particle filter, and the weights w_i from (12.86) are shown for both cases ($c = 0.2$ and $c = 0.8$) in Fig. 12.17c–d. The narrow weights for the case $c = 0.2$, compared with those for $c = 0.8$, indicate a high sensitivity of sea level to horizontal mixing when the global horizontal mixing (the 'truth' from a

low mixing model case) is low. The shape of the weights for $c = 0.8$ demonstrates the strong nonlinear dependence of sea level on the horizontal mixing coefficients. The resulting sea level in the inverse estimation for the case $c = 0.8$ is close to the synthetic truth (Fig. 12.17b).

12.4.3 Variational data assimilation

The Kalman filter result (12.74) is optimal, and it can be shown that the analysis $\mathbf{x}^a$ in (12.73) is a solution of the optimisation problem (Lorenc, 1988)

$$\mathbf{x}^a = \min_{\mathbf{x}} J(\mathbf{x}), \tag{12.89}$$

where the cost function J is given by

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}). \tag{12.90}$$

The variational data assimilation methods used differ in (i) the cost function J that is minimised, (ii) which control variable is used in the minimisation and (iii) how errors are taken into account. As an example, we provide a brief description of 4D-Var. Here the control variable is the initial state $\mathbf{x}(t_0)$ at time $t = t_0$. Suppose the initial condition $\mathbf{w}^b(t_0)$ of the background model is given. The analysis is the model trajectory that simultaneously minimises the distance to the initial background $\mathbf{w}^b(t_0)$ and the observations $\{\mathbf{y}_i : i = 0, \cdots, N-1\}$. This is an optimisation problem, which in the incremental 4D-Var formulation (Courtier et al., 1994) is stated as

$$\delta \mathbf{w}^a = \min_{\delta \mathbf{w}} J(\delta \mathbf{w}), \tag{12.91a}$$

$$J(\delta \mathbf{w}) = \delta \mathbf{w}^T \mathbf{B}^{-1} \delta \mathbf{w} + \sum_{i=0}^{N-1} \mathbf{d}_i^T \mathbf{R}_i^{-1} \mathbf{d}_i, \tag{12.91b}$$

$$\mathbf{d}_i = \mathbf{y}_i - H_i \mathcal{M}(t_0, t_i)(\mathbf{w}^b(t_0)) - \mathbf{H}_i \mathbf{L}(t_0, t_i) \delta \mathbf{w}. \tag{12.91c}$$

In the preceding equations, J is the cost function that measures the distance to the observations and the initial conditions, $\delta \mathbf{w}^a$ is the optimal increment on the initial background $\mathbf{w}^b(t_0)$ state and $\mathbf{d}_i$ is the departure of the model trajectory from observation $\mathbf{y}_i$ (Fig. 12.18). The operators $\mathcal{M}$ and H_i are the evolution operator and the observation operator with $\mathbf{L}$ and $\mathbf{H}_i$ their linearisations around the background trajectory $\mathbf{w}^b(t_i)$. The matrices $\mathbf{B}$ and $\mathbf{R}_i$ are the covariance matrices for the background errors and observational errors. Given an optimum $\delta \mathbf{w}^a$ of (12.91), the analysis $\mathbf{w}^a(t_i)$ is given by

$$\mathbf{w}^a(t_i) = \mathcal{M}(t_0, t_i)(\mathbf{w}^b(t_0) + \delta \mathbf{w}^a). \tag{12.92}$$

A minimum of the cost function J is computed, for example, with a quasi-Newton conjugate gradient method in which the gradient ∇J is needed. For (12.91), the

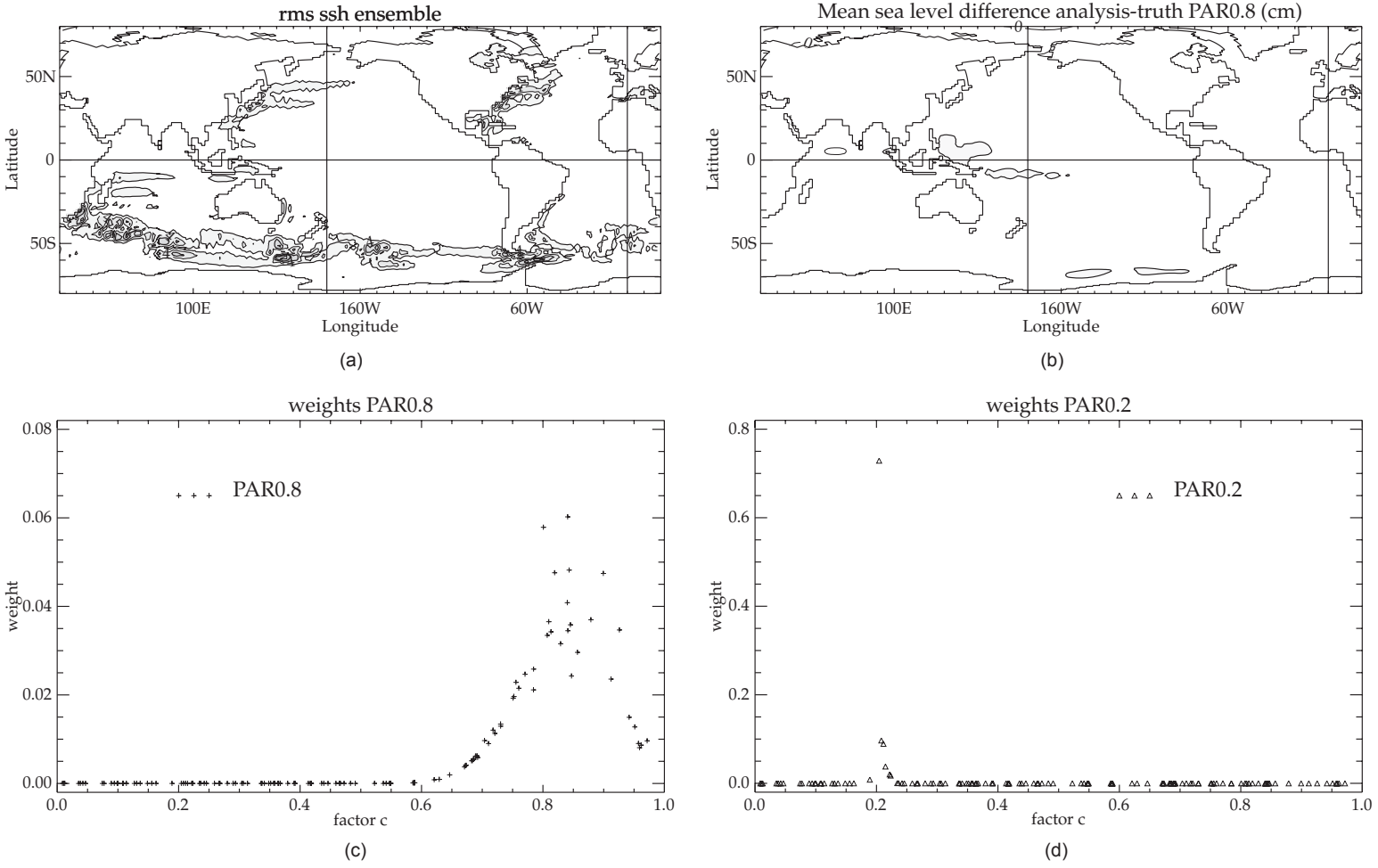


Figure 12.17 (a) Root mean square sea level for the 128-member ensemble of the global ocean model. The contour interval is 2 cm. (b) Difference between synthetic observations and inverse estimate for $c = 0.8$. (c) Values of w_i for each of the ensemble members as a function of c, with 'observations' for the case $c = 0.2$. (d) Same as (c) with 'observations' for the case $c = 0.8$ (figure from Vossepoel and Van Leeuwen, 2007).

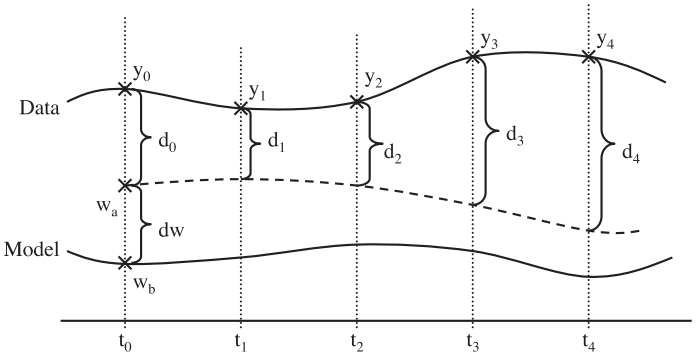


Figure 12.18 Sketch of the 4D-Var method, showing an assimilation interval with 4 points and indicating the model trajectory, the observations and the analysis.

gradient is given by

$$\nabla J(\boldsymbol{\delta}\mathbf{w}) = 2\mathbf{B}^{-1}\boldsymbol{\delta}\mathbf{w} - 2\sum_{i=0}^{N-1}\mathbf{L}(t_0, t_i)^T\mathbf{H}_i^T\mathbf{R}_i^{-1}\mathbf{d}_i. \tag{12.93}$$

For explicit time integration models, the procedure is to evaluate the cost function by forward time stepping, whereas the gradient is evaluated by integrating backward in time using the adjoint model $\mathbf{L}(t_0, t_i)^T$. For 4D-Var with implicit time stepping, only forward time stepping is used to evaluate the gradient (Terwisscha van Scheltinga and Dijkstra, 2005).

The 4D-Var method is used in many applications in the geosciences, such as numerical weather prediction (Lorenc, 2007), atmospheric reanalysis (Kalnay et al., 1996; Compo et al., 2011) and ocean state estimation (Wunsch and Heimbach, 2007; Moore et al., 2011). Comparisons between the use of 4D-Var and EnKF methods on several atmospheric models are provided in Kalnay et al. (2007) and Buehner et al. (2010).

12.5 Outlook

The aim of this chapter was to show the role of concepts of (stochastic) dynamical systems theory in climate prediction. As we have seen, one of the central elements in prediction is the amplification of errors (perturbations) along trajectories. Many systems, even on climate time scales, are nonlinear and display non-Gaussian statistics, and concepts of CNOPs, Lyapunov vectors and Bred vectors are better suited in real applications than those (e.g., singular vectors) based on linear systems and Gaussian statistics. The concepts of dynamical systems are now also widely and successfully applied in prediction systems where, by incorporating observations through data

assimilation, adequate state estimations and useful predictions can be obtained for highly nonlinear systems.

The skill of numerical weather prediction has improved over the years, and forecasts a few days ahead over many regions over the globe have reasonable skill. The growth of perturbations associated with instabilities in the atmospheric flow field limits the predictability horizon to about ten days. On the seasonal time scale, skill is improving, and regions of high-potential predictability (Rowell, 2010) are being identified. El Niño prediction is hampered by the existence of the spring predictability barrier, with a large growth of errors during the northern hemispheric spring due to a particular sensitivity of the equatorial ocean-atmosphere system (Duan et al., 2009). There is much activity in the area of decadal predictability (Keenlyside et al., 2008), but the processes controlling the predictability horizons are not very clear yet. In the IPCC-AR4 and also for the next IPCC assessment, attempts are made to determine the range of possible future climate states in 2100 from ensembles of simulations with a suite of global climate models (Meehl et al., 2007). As mentioned in Section 12.1, the additional uncertainty here is the forcing of the system, in particular the concentration of atmospheric greenhouse gas concentrations. Hence, in the IPCC-AR4 assessment, one only talks about projections instead of predictions.

Important issues in climate research where the potential of dynamical systems has not been fully explored are the detection of critical behaviour and the a priori determination of extremes. How do we sense that the climate state vector is close to a transition point such as a Hopf bifurcation or a saddle-node bifurcation? Can we predict the approximation of these points from the methods used in the previous sections with a useful skill? In recent years, there has been substantial effort to detect transition behaviour from time series (Held and Kleinen, 2004; Livina and Lenton, 2007; Scheffer et al., 2009; Lenton, 2011; Thompson and Sieber, 2011), although in many cases, the available data series are too short to provide convincing answers. Extreme behaviour is centrally important in climate change issues: eventually, it is the extremes in temperature (heat waves) and precipitation (flooding) that provide most impact and damage to nature and society. The analysis of extreme behaviour in dynamical systems is currently an active research area (Sterl et al., 2008; Holland et al., 2012).

The main challenge of the application of all the techniques in this chapter to problems in climate predictability is related to the 'curse of dimensionality'. The spatial resolution and the number of processes represented in climate models will increase with time (cf. Chapter 6), and a model in which all dynamically relevant scales of motion are represented will have an estimated 10^{12} degrees of freedom. With the speed of processors being stagnant and the computer hardware switching to graphical processors units (GPUs) and multicore platforms, enormous challenges lie ahead to be able to apply data-assimilation techniques to future climate models on the algorithm aspects, the data-handling aspects and the high-performance computing

aspects. Definitely, major technical hurdles must be cleared before reliable forecasts of future climate conditions can be produced using the best climate models available.

Finally, although this may not be realised by the general public and policy makers, climate prediction provides a historic opportunity for humankind to determine at least one aspect of life relatively far into the future with a specified quantitative uncertainty. The impact of such predictions cannot be underestimated, as is now already the case for numerical weather predictions a few days ahead.