Chapter 22

Data Assimilation

ABSTRACT
This chapter outlines the problem of predictability and the methods used to blend, in an optimal way, observations with model computations in order to guide the latter and to produce improved simulations of geophysical fluid phenomena. The methods invoke physical as well as statistical reasoning and rely on certain approximations that facilitate their implementation in operational forecast models.

22.1 NEED FOR DATA ASSIMILATION

Personal experience teaches us that a weather forecast is reliable for only a few days from the moment when it is issued. The time up to which the prediction is made is called *lead time*, and a reliable forecast can only be had if the lead time is not too long, typically no more than a week for midlatitude weather. Predictions further into the future are so imprecise that very simple prediction methods, such as the use of climatological values for the day concerned or persistence of today's weather, may perform as well as sophisticated weather forecast systems. Before discussing the reasons why forecasting errors increase with lead time, we anticipate that any forecast system will need periodic reinitialization if predictions are needed on a regular basis. Such a reinitialization must certainly take into account the most recent observations in order to infer a correct present state of the system, an operation known as *field estimation* in forecast jargon. From this better estimate, a forecast can be restarted on a better basis.

For weather forecasts, field estimates are *sequential* in the sense that they only use existing data, that is, from the past up to the day on which the forecast begins. For other applications, the best field estimate of a past situation is constructed, in which case data from moments after the forecast has begun can also be incorporated, and a nonsequential method is used. A typical example in which all available data are used is a so-called *re-analysis*, which incorporates the best fields over a given time period together with the physics; these provide the best picture of reality at any moment.

The melding of physical laws and observations, be it in a sequential or nonsequential way, is carried out through so-called *data assimilation*. Data

Introduction to Geophysical Fluid Dynamics, Volume 101, ISSN 0074-6142, DOI: 10.1016/B978-0-12-088759-0.00022-5

assimilation may be performed intermittently, for example every day using data from the previous day, or continuously, as data become available.

Since data assimilation exploits observational data, it is also possible to quantify the forecast errors up to a point once new data have arrived. Forecast errors can then be used to assess the skill of the forecast system, a measure of its predictive capability. Here, it is customary to compare the error of the forecast with the error of an elementary forecast. Rudimentary forecasts are one of the following: Persistence (tomorrow the weather will be the same as today), climatology (next week's weather will be the average weather of the last twenty years at the same time of year), or random forecast (e.g., one of the two preceding methods with an added random noise of zero average and prescribed variance). One measure of forecasting skill is the Brier skill score, S, which is calculated from an error measure ϵ^f of the actual forecast system and the corresponding error measure ϵ^r of a rudimentary (or reference) forecast system (Brier, 1950; see also Wilks, 2005):

$$S = 1 - \frac{\epsilon^f}{\epsilon^r}, \tag{22.1}$$

Clearly, if the forecasting system has an error (ϵ^f) equal to the rudimentary approach error (ϵ^r), its skill is nil. On this scale, the unattainable perfect forecasting skill corresponds to a score of unity (100%). Should the skill of the forecast system falls below zero, the forecast system should be considered as no better than the most basic forecast system, although there might still be some useful information in the forecast. The skill score is often used to quantify the improvement of a new forecast system over an older version, in which case ϵ^r is taken as the error of that older system.

Clearly the skill score depends on the nature of the chosen error norm ϵ, which is, for example, the root-mean-square (rms) error between two fields, the error on the maximum temperature, the error on the hours of sunshine etc., but more importantly the skill varies with lead time. The further the forecast extents into the future, the lower the skill score tends to be, and we naturally come back to the question of why it is so difficult to make accurate long-range forecasts. The previous chapters might have biased our perception of geophysical fluid dynamics toward a system governed by equations, the solutions of which uniquely follow from an adequate set of initial and boundary conditions. This is the case in theory, but we ought to accept the idea that with imperfect models and inaccurate boundary and, especially, initial conditions, errors can have a tendency to accumulate over the course of a long-range forecast and thus to reduce skill with lead time. As a matter of fact, the situation is more dramatic than that. Even if we could control the initial errors below any arbitrary value, which obviously we would never be able to do, some dynamical equations, by their very nature, lead to solutions that diverge rapidly for even extremely small changes in the initial conditions. The famous *Lorenz equations*

(Lorenz, 1963, page 135) provide an archetype of a system of equations that exhibit such a behavior:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y-x), \tag{22.2a}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = rx - y - xz, \tag{22.2b}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - bz, \tag{22.2c}$$

in which σ, r, and b are fixed parameters, and x, y, and z are temporal variables. The equations form a low-order truncation of a spectral model of atmospheric motions and, despite their innocent look, are known to generate chaotic trajectories such that two very close initial conditions will lead to completely different solutions after some time (Fig. 22.1). Clearly, there is a *predictability limit*.

More generally, the accumulation of errors, even when starting with arbitrarily small errors in the initial conditions, can result in the existence of a predictability limit in strongly nonlinear systems. This limit is estimated to be one to two weeks for the global atmosphere and on the order of one month for midlatitude ocean eddies. It is therefore not surprising that forecast skill decreases with lead time approaching the predictability limit of the system (Fig. 22.2). An idea of the predictability limit can be obtained by considering the autocorrelation of the solution for time delay $\Delta t > 0$:

$$\rho(\Delta t) = \frac{\frac{1}{T}\int_{\Delta t}^{T} u(t)u(t-\Delta t)\,\mathrm{d}t}{\sqrt{\frac{1}{T}\int_{\Delta t}^{T} u(t)^2\,\mathrm{d}t}\sqrt{\frac{1}{T}\int_{\Delta t}^{T} u(t-\Delta t)^2\,\mathrm{d}t}} \tag{22.3}$$

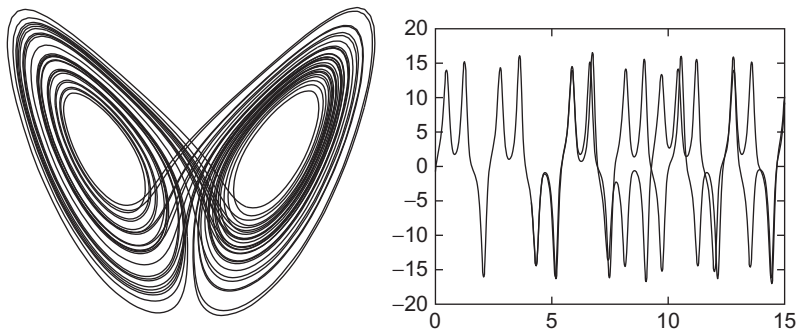

FIGURE 22.1 Trajectory in (x, z) space (left panel) showing the vacillation of the solution between two cycles. For two slightly different initial conditions, the corresponding two solutions $x(t)$ track each other for some time and then diverge (right panel). Graph obtained by using `chaos.m` that solves the Lorenz equations (22.2) with $\sigma = 10$, $r = 28$ and $b = 8/3$.

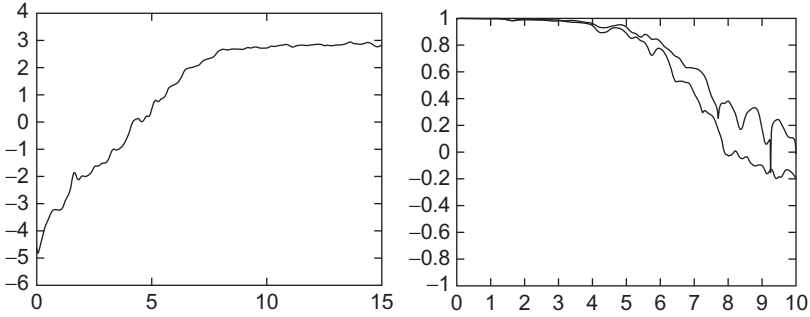

FIGURE 22.2 For incorrect initial conditions, the logarithm of the forecast error increases as a function of lead time (left panel). Skill score decreases as a function of lead time for two different base forecasts (right panel). For each curve an ensemble of 200 simulations with the Lorenz equations was averaged.

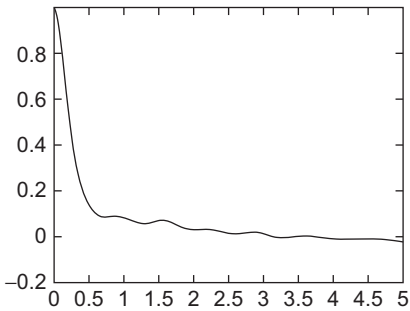

FIGURE 22.3 Autocorrelation as a function of time delay Δt for a solution $x(t)$ of the Lorenz equations (22.2). The curve, obtained by averaging 200 different trajectories, shows that, on average, after a few time units, the solution is not well correlated with its own previous values.

with time $T \to \infty$. This function measures the extent to which the solution at a given moment is on average close to the solution at a moment Δt earlier. In this sense, the delay Δt for which the autocorrelation value ρ approaches zero defines the time interval over which the solution no longer resembles its own past, i.e., the time over which it "decorrelates" from itself. Put another way, the value of Δt for which ρ falls close to zero is the memory time of the dynamical process. For a purely random function, one without any past memory, the autocorrelation is zero for any delay Δt. For the solution to the Lorenz equations, the predictability time can inferred from Fig. 22.3. It is important to note, however, that the system may still be considered as deterministic as every initial condition determines a unique evolution. The gradual loss of predictability means that, as time goes on, we become less able to identify this unique trajectory even with the finest numerical surgery and measurement tools.

In geophysical fluid dynamics, motions are not only controlled by initial conditions but also by boundary conditions, and the predictability limit depends

on the relative importance of boundary conditions to initial conditions. If the system mostly responds to forcing applied at a boundary, such as an semi-enclosed shallow sea with a strong ocean tide at its opening, the future behavior can be predicted very far ahead, and the model's skill is essentially measured by the accuracy of the forcing. Since forcings and boundary conditions are typically well known, skill remains high for long periods. The predictability issue manifests itself most severely with systems essentially controlled by initial conditions. The global atmosphere, which has no lateral boundary, is the quintessential example, and its initial conditions must be set with utmost care, and, nonetheless, the prediction skill may start very high but degrade quite quickly. Between those two extremes, the predictability of a system depends on the relative importance of initial over boundary conditions (Fig. 22.4).

Despite the inherent problem of predictability, we can increase forecast skills by reducing to a maximum the uncertainties in the model and its initial and boundary conditions, so as to push further away the predictability limit of the forecasting system. We can try to keep some of the errors under control, especially those already classified as modeling errors (see Section 4.8). Within the pages of this book, we encountered various levels of model simplifications, such as the hydrostatic approximations, the quasi-geostrophic approximation, and the shallow-water model. For all of these, numerical discretization in space and time added further sources of errors. For initial and boundary conditions based on observed fields, we can also distinguish several other kinds of error, with the most obvious one being instrumental error (of generally known and relatively low standard deviation). In Section 1.8, we also encountered the representativity error due to the fact that a point measurement (e.g., a temperature measurement in a city) is not the variable we actually would like to observe (the temperature at the 50 km scale). Synopticity errors (i.e., lack of simultaneity among several measurements) can be a concern when observations are binned into time slots for analysis and assimilation (e.g., assembling into a single "snapshot" view of the ocean the data gathered during the span of a cruise) because waves propagate, and temporal shifts can lead to severe Doppler effects (Rixen, Beckers & Allen, 2001). In general, any data treatment before assimilation into the model, such as interpolation, must be taken into account when assessing the errors associated with the "observation."

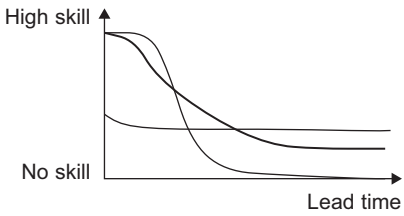

FIGURE 22.4 Predictability behavior for boundary-condition dependent systems (flat line), initial-condition dependent systems (steep line) and mixed situations (intermediate line).

The distinction between modeling and observational errors is not always clear, and the discrete sampling necessitated by the numerical grid can be considered either a modeling error (truncation of continuous operators) or an observational error (inaccurate because incomplete data). In any case, we face the problem that some of the information, both from model and observations, is incomplete and corrupted by errors. The main objective of data assimilation is the reduction of the influence of those errors on the simulation by utilizing data to guide the model in the best possible way. This improved analysis will generally extend the lead time for which predictions are reliable. This, however, is not the sole advantage of data assimilation; others will be discussed at the end of the chapter.

The methods described below were for the most part developed within the context of atmospheric modeling (e.g., Bengtsson, Ghil & Kallen, 1981; Ghil & Childress, 1987; see also Navon, 2009 for a review) and later adapted by the oceanographic community (e.g., Evensen, 1994; Ghil, 1989; Ghil & Malanotte-Rizzoli, 1991). The presentation uses the unified notation proposed by Ide, Courtier, Ghil and Lorenc (1997) also adopted in the reference book of Kalnay (2003), with a few exceptions to stay consistent with the notation used elsewhere in this book. Further seminal textbooks are those by Bennett (1992), Malanotte-Rizzoli (1996), Robinson, Lermusiaux and Sloan (1998), and Wunsch (1996).

22.2 NUDGING

Among the first methods used to guide numerical simulations by data injection is the nudging method. It starts from the governing equations of the state vector $\mathbf{x}$, which is the collection of variables in the forecasting system,

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathcal{Q}(\mathbf{x}, t),$$

where the operator $\mathcal{Q}$ bearing on the set of variables stands for the model equations. Assuming the observations are distributed in exactly the same way as the state-vector components (i.e., observations are collected on the same grid as the numerical model), we group them into a vector $\mathbf{y}$ ($\mathbf{y}$ being what $\mathbf{x}$ should have been if the model had been correct). The nudging method then simply adds a correction term proportional to the difference between model results and observations:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathcal{Q}(\mathbf{x}, t) + \mathbf{K}(\mathbf{y} - \mathbf{x}). \tag{22.4}$$

The additional term is the product of a matrix $\mathbf{K}$ with the model-observation misfit. For the nudging method, this matrix is the diagonal matrix formed with

a set of so-called time scales of *relaxation* noted τ_i, which may differ from variable to variable: $K_{ii} = 1/\tau_i$ and $K_{ij} = 0$ for $i \neq j$. Since the difference between the model simulation $\mathbf{x}$ and observations $\mathbf{y}$ is zero where there is no error, the additional term only acts as a correction wherever necessary, "nudging" (i.e., pulling) the solution towards the observations.

The "strength" of the nudging depends intimately on the values ascribed to the relaxation time scales τ_i. If the time scale is long compared with the time scale of evolution of the corresponding variable, the correction is kept small and qualified as background relaxation. Such background relaxation very often uses climatological values in place of "observations." When no observation is present for a variable, the corresponding relaxation time is simply set to infinity. When observations are only available at certain moments, the relaxation time scale is made to decrease from a large to a small value when approaching the moment t^o at which data are available to ensure a smooth temporal incorporation of data. An example of time-dependent function is

$$\frac{1}{\tau} = K \exp\left[-(t - t^o)^2 / T^2\right] \tag{22.5}$$

with T being an appropriate time scale over which the observation can nudge the simulation. The time relaxation can also be made spatially dependent when different dynamical regimes can be identified for physical reasons. One particular form of nudging is surface relaxation in ocean models, where the simulated sea surface is nudged towards the observed surface fields. Such nudging is often maintained with a low intensity even when full atmospheric fluxes are applied onto the ocean in order to avoid any drift (e.g., Pinardi et al., 2003). At the other extreme, when the relaxation time is taken very short and comparable with the time step, the nudging method basically replaces the simulated value with the corresponding observation, a process called *direct insertion* when carried out in a single time step.

Nudging was widely used in the early days of data assimilation but has now been superseded by more sophisticated techniques (see below). Nonetheless nudging remains popular near boundaries, where it can be interpreted as a boundary condition corrected by observations. In this case, the time-continuous nature of the correction is beneficial by avoiding sudden shocks in the model.

22.3 OPTIMAL INTERPOLATION

The previous method, though robust and found useful in the past, is rather an ad-hoc approach, and we will now present a method that is based on sound statistical optimization. In particular, relationships between different variables will be exploited in order to enhance the assimilation. For this, we will use a notation that is slightly different from the discretization notation used until now. Because sequential assimilation cycles perform an analysis at not every time step of the

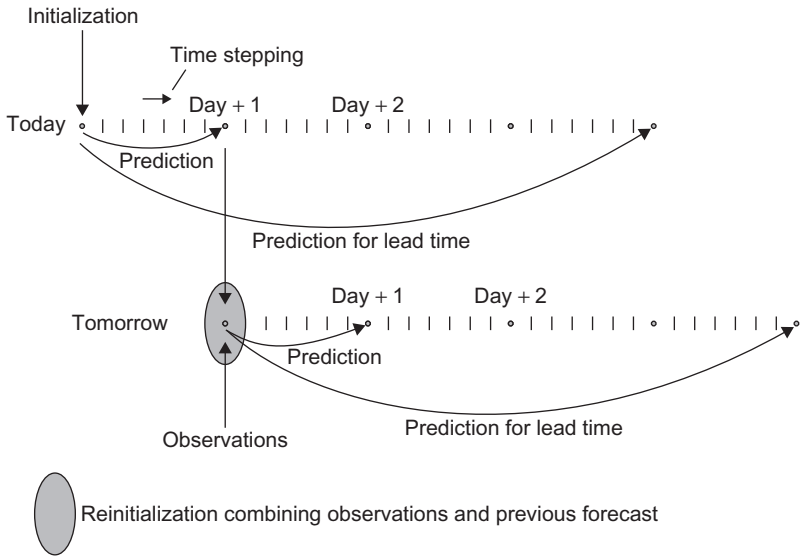

FIGURE 22.5 Schematic representation of reinitialization through sequential data assimilation on a daily basis. On a given day, the model provides a forecast for several days ahead. The next day, two pieces of information become available, the forecast for the day and the daily observation. Combining the two generates a new initial condition for the model, and an improved forecast can be provided for the following days.

model (Fig. 22.5), we shall denote by $\mathbf{x}_n$ a particular cycle of the assimilation.[1] We also should keep in mind that in practical applications, it is advantageous to use a state vector defined by anomalies (i.e., departure from a reference state), normalized so that each element of the state vector should be comparable with the others. The state vector then does not gather velocity and temperature but normalized versions of them. Because it contains variables of different types, we are heading for a so-called *multivariate* approach. Forecast values will be referenced by superscript ^{f} and the "analyzed" variables, obtained after combining forecast with observations by the superscript ^{a}.

For the purpose of illustration, we start with the elementary problem of having at our disposal at a given moment two pieces of information about a temperature of unknown true state T^t. The information can originate from either measurement or model and can include an error ϵ. For the two given values T_1 and T_2, we therefore write

$$T_1 = T^t + \epsilon_1, \quad \langle \epsilon_1 \rangle = 0, \quad T_2 = T^t + \epsilon_2, \quad \langle \epsilon_2 \rangle = 0, \tag{22.6a}$$

[1] Note the difference with the notation $\mathbf{x}^n$ previously used to refer to the variable at time step n.

where we assume that, on average, denoted by brackets $\langle \quad \rangle$, errors vanish. In other words, we suppose the values to be unbiased. We can estimate the unknown, true temperature by a linear combination of the two available values:

$$T = w_1 T_1 + w_2 T_2 = (w_1 + w_2) T^t + (w_1 \epsilon_1 + w_2 \epsilon_2) \tag{22.7}$$

and on average, this estimate will take the value

$$\langle T \rangle = (w_1 + w_2) T^t, \tag{22.8}$$

so that we obtain an unbiased estimate of the true state if we take $w_1 + w_2 = 1$. In this case, we perform a de-facto weighted average among the two pieces of available information, an intuitive approach. An unbiased estimate, or *analysis*, T^a of the true state is therefore

$$T^a = (1 - w_2)\, T_1 + w_2\, T_2 = T_1 + w_2\, (T_2 - T_1) \tag{22.9}$$

while in reality there is an error

$$T^a - T^t = (1 - w_2)\, \epsilon_1 + w_2\, \epsilon_2. \tag{22.10}$$

This error is zero on average, but its variance is not zero:

$$\left\langle (T^a - T^t)^2 \right\rangle = (1 - w_2)^2 \left\langle \epsilon_1^2 \right\rangle + w_2^2 \left\langle \epsilon_2^2 \right\rangle + 2(1 - w_2) w_2\, \langle \epsilon_1 \epsilon_2 \rangle. \tag{22.11}$$

The actual errors ϵ_1 and ϵ_2 are not known, otherwise we would have had ready access to T^t. However, given some basic information on the source of the errors, we can assess the so-called *error variance* $\left\langle \epsilon_1^2 \right\rangle$ or, equivalently, the standard deviation $\sqrt{\left\langle \epsilon_1^2 \right\rangle}$. If the two pieces of information (observed and/or modeled) leading to T_1 and T_2 are independent of each other, we may reasonably suppose that the errors ϵ_1 and ϵ_2 are uncorrelated, which in statistical terms means $\langle \epsilon_1 \epsilon_2 \rangle = 0$. Hence, the error variance $\left\langle \epsilon^2 \right\rangle$ of the analysis is

$$\left\langle \epsilon^2 \right\rangle = (1 - w_2)^2 \left\langle \epsilon_1^2 \right\rangle + w_2^2 \left\langle \epsilon_2^2 \right\rangle. \tag{22.12}$$

Naturally, the best estimate of T^t is the one with the lowest error variance. Finding the value of w_2 that minimizes the right-hand side of the preceding equation,

we obtain

$$w_2 = \frac{\langle \epsilon_1^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} \tag{22.13}$$

and obtain the minimal error variance:

$$\langle \epsilon^2 \rangle = \frac{\langle \epsilon_1^2 \rangle \langle \epsilon_2^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} = \left(1 - \frac{\langle \epsilon_1^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} \right) \langle \epsilon_1^2 \rangle, \tag{22.14}$$

while the temperature estimate is

$$T^a = T_1 + \left(\frac{\langle \epsilon_1^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} \right) (T_2 - T_1). \tag{22.15}$$

We observe that the error variance of the combination of T_1 and T_2 is smaller than each of $\langle \epsilon_1^2 \rangle$ and $\langle \epsilon_2^2 \rangle$. Using information from two sources, even if one of the two has a relatively large error, reduces on average the uncertainty. This is the idea underlying data assimilation using error analysis to reduce overall uncertainty. If optimization such as the minimization of Eq. (22.12) is used, the process can be quite effective at decreasing the forecast error.

We can reach the same optimal estimate by finding the T value that minimizes a weighted measure of the differences between the analysis and the available information, with weights inversely proportional to the error variance of the information:

$$\min_T J = \frac{(T - T_1)^2}{2\langle \epsilon_1^2 \rangle} + \frac{(T - T_2)^2}{2\langle \epsilon_2^2 \rangle}. \tag{22.16}$$

In other words, we do not mind that the analysis departs from the relatively uncertain observation, but we do require that the analysis falls closer to the observation if the latter happens to be more accurate. The minimum of Eq. (22.16) is reached when T takes the value T^a of Eq. (22.15).

The optimal reduction in error can be used to blend observations with a model forecast through the technique called *Optimal Interpolation*.[2] Typically, there are many more model data than available observations[3] so that the size M of $\mathbf{x}$ is much larger than the size P of the vector $\mathbf{y}$ containing observations.

---

[2] Sometimes the term *objective analysis* is used instead. This is a poor choice of words, however, as the latter is generally no more than a mathematical interpolation that stands opposed to the historical *subjective analysis* of weather patterns by pencil and paper.

[3] In 2006, The European Centre for Medium-Range Weather Forecasts, ECMWF, used $M = 3 \times 10^7$ state variables for its operational ensemble T256 weather forecast model and assimilated $P = 3 \times 10^6$ observations every cycle of 12 hours. The Mercator ocean model PSY3v1 operates with $M = 10^8$ but "only" $P = 0.25 \times 10^6$ data were assimilated once a week.

To blend observations into a model simulation at a certain point in time, we construct the $\mathbf{x}^a$, called the *analysis*, as a linear combination of the forecast $\mathbf{x}^f$ up to that point in time with the observations $\mathbf{y}$ that have become available:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^f\right). \tag{22.17}$$

This procedure uses a linear *observation operator* $\mathbf{H}$ that selects or interpolates the simulated state variables to the points of observation in order to quantify the model-observation misfit $(\mathbf{y} - \mathbf{H}\mathbf{x}^f)$ (e.g., the temperature forecast is interpolated to the position in which a meteorological station measures temperature). In some cases, the matrix $\mathbf{H}$ may contain mathematical operations that relate multiple forecast fields to observed parameters not directly predicted. For example, the so-called Gelbstoff (literally, yellow matter) measured in the ocean by satellites is an aggregate value of organic compounds in the water column, but the corresponding dispersion model typically calculates the three-dimensional structure of the organic matter, and the matrix $\mathbf{H}$ has to add the values at the various grid points spanning the height of the water column at the point of observation. Occasionally the relationship between model variables and observations is nonlinear, and the correction term, called *innovation* vector and noted $\mathbf{d}$, should be replaced by

$$\mathbf{d} = \mathbf{y} - \mathcal{H}\left(\mathbf{x}^f\right) \tag{22.18}$$

where $\mathcal{H}$ stands for a nonlinear function. Here, we only consider a matrix $\mathbf{H}$ assuming a linear relationship between observed and modeled values.

We can also mention that the interpretation of observation errors depends on how the data are prepared, and how the observation operator is constructed. Altimetric data can, for example, be assimilated along tracks, and the observational error estimated as a combination of instrumental, representativity (mix of spatial scales), synopticity (time-slot binning), and interpolation errors when sampling the model along the tracks using $\mathbf{H}$. If, for convenience, the tracks are gridded beforehand on the same mesh as the model grid, this interpolation, itself performed, for example, with a spatial Optimal Interpolation version (Numerical Exercise 22.2), has an associated error covariance, which must be taken into account in prescribing the observational error. Generally, interpolation from the model to data location (through $\mathbf{H}$) introduces fewer errors because there are typically many more model grid points than observation points.

The matrix $\mathbf{K}$ of size $M \times P$, where M is the number of model variables and P the number of observations, is determined in order to obtain the "best" analysis, that is, an optimal blending of data into model. This analysis depends on the error fields of both forecast and observations. The forecast error is

$$\boldsymbol{\epsilon} = \mathbf{x} - \mathbf{x}^t, \tag{22.19}$$

that is, the difference between the calculated field $\mathbf{x}$ and the (unknown) true state $\mathbf{x}^t$. Likewise, the observational error is

$$\boldsymbol{\epsilon}^o = \mathbf{y} - \mathbf{y}^t. \tag{22.20}$$

From these errors, even though we do not know their actual values because the true states are not known, we can define statistical averages and variances. Obviously, for unbiased models and observations—which we both assume—the averages (first-order moments) $\langle \boldsymbol{\epsilon} \rangle$ and $\langle \boldsymbol{\epsilon}^o \rangle$ are zero. The error-covariance matrix for observations

$$\mathbf{R} = \left\langle \boldsymbol{\epsilon}^o \boldsymbol{\epsilon}^{o\mathsf{T}} \right\rangle \tag{22.21}$$

is square matrix with nonzero elements on the diagonal if each observation is subject to its own error. On the diagonal, we therefore find the error variance of each observation. Off-diagonal terms are nonzero whenever two separate observations are correlated, as it occasionally happens with satellite observations. Note that off-diagonal terms are symmetric and that for any vector $\mathbf{z}$, the quadratic form $\mathbf{z}^\mathsf{T}\mathbf{R}\mathbf{z} = \left\langle \left(\mathbf{z}^\mathsf{T}\boldsymbol{\epsilon}^o\right)^2 \right\rangle$ is never negative so that a covariance matrix is always semi-positively defined. If errors are random and span the whole state-vector space, the covariance matrix is strictly positively defined.

The analysis step (22.17) can be expressed as

$$\mathbf{x}^t + \boldsymbol{\epsilon}^a = \mathbf{x}^t + \boldsymbol{\epsilon}^f + \mathbf{K}\left(\boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^f\right) + \mathbf{K}\underbrace{\left(\mathbf{y}^t - \mathbf{H}\mathbf{x}^t\right)}_{=0}. \tag{22.22}$$

The last term is zero because, by definition, a perfect model would perfectly match the observations of the true state. It follows that the error on the analysis field is

$$\boldsymbol{\epsilon}^a = \boldsymbol{\epsilon}^f + \mathbf{K}\left(\boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^f\right), \tag{22.23}$$

from which we can proceed to construct the error covariance $\left\langle \boldsymbol{\epsilon}^a \boldsymbol{\epsilon}^{a\mathsf{T}} \right\rangle$ of the analysis by multiplying (22.23) by its own transposed and take the average:

$$\begin{aligned}\left\langle \boldsymbol{\epsilon}^a \boldsymbol{\epsilon}^{a\mathsf{T}} \right\rangle = &\left\langle \boldsymbol{\epsilon}^f \boldsymbol{\epsilon}^{f\mathsf{T}} \right\rangle + \mathbf{K}\left\langle (\boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^f)\boldsymbol{\epsilon}^{f\mathsf{T}} \right\rangle \\ &+ \left\langle \boldsymbol{\epsilon}^f \left(\boldsymbol{\epsilon}^{o\mathsf{T}} - \boldsymbol{\epsilon}^{f\mathsf{T}}\mathbf{H}^\mathsf{T}\right) \right\rangle \mathbf{K}^\mathsf{T} \\ &+ \mathbf{K}\left\langle \left(\boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^f\right)\left(\boldsymbol{\epsilon}^{o\mathsf{T}} - \boldsymbol{\epsilon}^{f\mathsf{T}}\mathbf{H}^\mathsf{T}\right) \right\rangle \mathbf{K}^\mathsf{T}.\end{aligned} \tag{22.24}$$

Defining covariance matrices

$$\mathbf{P} = \left\langle \boldsymbol{\epsilon}\boldsymbol{\epsilon}^\mathsf{T} \right\rangle \tag{22.25}$$

for both forecast and analysis error fields (with superscripts ^{f} and ^{a}, respectively) and making the reasonable assumption that there is no correlation[4] between observational error and forecast error, i.e., $\langle \boldsymbol{\epsilon}^o \boldsymbol{\epsilon}^{f\mathsf{T}} \rangle = 0$, we can rewrite the error-covariance matrix after analysis as

$$\begin{aligned}\mathbf{P}^a &= \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f - \mathbf{P}^f\mathbf{H}^\mathsf{T}\mathbf{K}^\mathsf{T} + \mathbf{K}\left(\mathbf{R} + \mathbf{H}\mathbf{P}^f\mathbf{H}^\mathsf{T}\right)\mathbf{K}^\mathsf{T} \\ &= \mathbf{P}^f - \mathbf{P}^f\mathbf{H}^\mathsf{T}\mathbf{A}^{-1}\mathbf{H}\mathbf{P}^f + \left(\mathbf{P}^f\mathbf{H}^\mathsf{T} - \mathbf{K}\mathbf{A}\right)\mathbf{A}^{-1}\left(\mathbf{H}\mathbf{P}^f - \mathbf{A}\mathbf{K}^\mathsf{T}\right) \end{aligned} \tag{22.26}$$

in which matrix $\mathbf{A}$, defined for convenience as

$$\mathbf{A} = \mathbf{H}\mathbf{P}^f\mathbf{H}^\mathsf{T} + \mathbf{R} \tag{22.27}$$

is symmetric and can most likely be inverted.[5]

If state variables are suitably scaled to enable comparison of, say, temperature errors with velocity errors, the overall error ϵ^a of the analyzed field may be taken as the expected norm of the error vector:

$$\epsilon^a = \left\langle \boldsymbol{\epsilon}^{a\mathsf{T}} \boldsymbol{\epsilon}^a \right\rangle. \tag{22.28}$$

This, however, is nothing other than the trace of the covariance matrix $\left\langle \boldsymbol{\epsilon}^a \boldsymbol{\epsilon}^{a\mathsf{T}} \right\rangle$, and an overall measure of the analysis error is thus

$$\epsilon^a = \text{trace}\left(\mathbf{P}^a\right). \tag{22.29}$$

Since the $M \times P$ matrix $\mathbf{K}$ is still unspecified, a reasonable choice is to choose it such that it minimizes the global error. One way of proceeding is to take the trace of Eq. (22.26) and differentiate the trace with respect to all components of $\mathbf{K}$ in order to find the extremal value of the global error; because (22.26) is a quadratic form in terms of $\mathbf{K}$, and because $\mathbf{A}^{-1}$ is positive defined if it exists, the extremum is assured to be a minimum. Alternatively, we can think of the error as being a function $\epsilon^a(\mathbf{K})$ of the matrix $\mathbf{K}$ and search for an optimal $\mathbf{K}$ such that

$$\epsilon^a(\mathbf{K}+\mathbf{L}) - \epsilon^a(\mathbf{K}) = 0 \tag{22.30}$$

for an arbitrary, small departure matrix $\mathbf{L}$. After linearization with respect to $\mathbf{L}$, we thus require

$$\text{trace}\left(-\mathbf{L}\left(\mathbf{H}\mathbf{P}^f - \mathbf{A}\mathbf{K}^\mathsf{T}\right) - \left(\mathbf{P}^f\mathbf{H}^\mathsf{T} - \mathbf{K}\mathbf{A}\right)\mathbf{L}^\mathsf{T}\right) = 0. \tag{22.31}$$

[4] Decorrelation between observational error and modeling error is justified by the very different origin of the information.

[5] Because $\mathbf{P}$ and $\mathbf{R}$ are semipositive defined matrices, the chances are not low. In addition, when observations and state variables are covering a wide domain, the covariances between distant points are generally small so that the matrices have a tendency to be diagonally dominant.

Since the two terms are the transpose of each other and thus share the same trace, it is sufficient that this common trace vanishes

$$\text{trace}\left(\left(\mathbf{P}^f\mathbf{H}^\mathsf{T}-\mathbf{KA}\right)\mathbf{L}^\mathsf{T}\right)=0.$$

Since $\mathbf{L}$ is arbitrary, its matrix coefficient must be zero, and the $\mathbf{K}$ matrix that minimizes the analysis error is

$$\begin{aligned}\mathbf{K}&=\mathbf{P}^f\mathbf{H}^\mathsf{T}\mathbf{A}^{-1}\\&=\mathbf{P}^f\mathbf{H}^\mathsf{T}\left(\mathbf{HP}^f\mathbf{H}^\mathsf{T}+\mathbf{R}\right)^{-1}.\end{aligned}\tag{22.32}$$

We note that matrix $\mathbf{A}$ must be invertible in order to reach the error minimum. The $\mathbf{K}$ matrix, which combines model forecast with data, is the analog of Eq. (22.13) and is called the *Kalman gain matrix*.

The minimum error covariance of the analysis is obtained by inserting (22.32) into (22.26):

$$\mathbf{P}^a=(\mathbf{I}-\mathbf{KH})\,\mathbf{P}^f=\left(\mathbf{I}-\mathbf{P}^f\mathbf{H}^\mathsf{T}\left(\mathbf{HP}^f\mathbf{H}^\mathsf{T}+\mathbf{R}\right)^{-1}\mathbf{H}\right)\mathbf{P}^f\tag{22.33}$$

which is the analog of Eq. (22.14). We note that neither optimal Kalman gain matrix nor minimum error covariance depends on the *value* of the observations or the forecast state vector but only on their statistical error covariance. The only field that depends on the actual field values is, of course, the state vector itself, which after optimal analysis becomes

$$\mathbf{x}^a=\mathbf{x}^f+\mathbf{P}^f\mathbf{H}^\mathsf{T}\left(\mathbf{HP}^f\mathbf{H}^\mathsf{T}+\mathbf{R}\right)^{-1}\left(\mathbf{y}-\mathbf{Hx}^f\right).\tag{22.34}$$

The use of Eq. (22.32) in Eq. (22.17) to combine forecast and observations with respective error covariances $\mathbf{P}^f$ and $\mathbf{R}$ is known as *Optimal Interpolation* (OI).

With forecast and observation covariances given, an alternative derivation of Optimal Interpolation is a variational approach called *3D-Var*, in which the objective is to find the state vector that minimizes the error measure J given by

$$J(\mathbf{x})=\frac{1}{2}\left(\mathbf{x}-\mathbf{x}^f\right)^\mathsf{T}\mathbf{P}^{f\,-1}\left(\mathbf{x}-\mathbf{x}^f\right)+\frac{1}{2}(\mathbf{Hx}-\mathbf{y})^\mathsf{T}\mathbf{R}^{-1}(\mathbf{Hx}-\mathbf{y})\tag{22.35}$$

In other words, the procedure is to search for the state vector closest to both model forecast and observations and which penalizes less the more accurate information, in close analogy with Eq. (22.16). Finding the optimum state vector leads to the same analyzed field as in Eq. (22.34). The demonstration is left as an exercise (Analytical Problem 22.4).

Optimal Interpolation can also be cast in terms of the maximum likelihood estimator of the true field, i.e., the field that has the highest probability of matching reality, which is also given by Eq. (22.34) if the *probability density function*

(pdf) of each errors follows the normal (Gaussian) distribution (e.g., Lorenc, 1986). In any case, it is necessary to quantify all error variances, which is not an easy task (see for example Lermusiaux et al., 2006).

22.4 KALMAN FILTERING

In the formulation of Optimal Interpolation, it turns out that the dynamical model used to provide the forecast does not appear explicitly. Only its forecast error covariance matrix $\mathbf{P}^f$ is needed besides the forecast itself. Thus, very little of what is known about the model is actually used.

Knowing, for example, that a fluid-flow model has a tendency to propagate errors on state variables along preferential directions, such as the flow itself or a wave guide, or may be amplified by unstable modes, we ought to ask how data assimilation could take into account certain model properties. For this, we start from the fact that between times n and $n+1$ when assimilation is performed the model advances the state vector according to

$$\mathbf{x}_{n+1} = \boldsymbol{\mathcal{M}}(\mathbf{x}_n) + \mathbf{f}_n + \boldsymbol{\eta}_n, \tag{22.36}$$

where $\mathbf{f}_n$ represents the external forcing between times n and $n+1$, and $\boldsymbol{\eta}_n$ the error introduced by the model as it marches through a multiplicity of time steps from time level n of the last data assimilation to time level $n+1$ when the next data assimilation is scheduled to take place. The operator $\boldsymbol{\mathcal{M}}$ stands for the inner machinery of the model, which calculates the state vector at time level $n+1$, over multiple time steps, from its previous value at time level n. Assuming the model forecast starts at time level n with the analyzed field produced from data assimilation at that time and also assuming a linearized model (for pure convenience because it is not true!), we write with a certain level of approximation:

$$\mathbf{x}^f_{n+1} = \mathbf{M}\mathbf{x}^a_n + \mathbf{f}_n + \boldsymbol{\eta}_n, \tag{22.37}$$

where the matrix $\mathbf{M}$ replaces the nonlinear operator $\boldsymbol{\mathcal{M}}$ of Eq. (22.36). Such a matrix is actually never constructed in operational models but is only introduced here to provide an elegant presentation of the method. The unknown, true state evolves similarly but without modeling error and thus obeys

$$\mathbf{x}^t_{n+1} = \mathbf{M}\mathbf{x}^t_n + \mathbf{f}_n \tag{22.38}$$

so that the forecast error $\boldsymbol{\epsilon}^f = \mathbf{x}^f - \mathbf{x}^t$ is

$$\boldsymbol{\epsilon}^f_{n+1} = \mathbf{M}\,\boldsymbol{\epsilon}^a_n + \boldsymbol{\eta}_n. \tag{22.39}$$

Multiplying this last equation to the right by its transpose and performing the statistical average to obtain error covariance, we obtain the so-called *Lyapunov*

equation, which allows for the advancement in time of the error covariance:

$$\mathbf{P}^f_{n+1} = \mathbf{M}\mathbf{P}^a_n\mathbf{M}^\mathsf{T} + \mathbf{Q}_n = \mathbf{M}\left(\mathbf{M}\mathbf{P}^a_n\right)^\mathsf{T} + \mathbf{Q}_n \tag{22.40}$$

with the following definition of the model-error covariance matrix:

$$\mathbf{Q}_n = \langle \boldsymbol{\eta}_n \boldsymbol{\eta}_n^\mathsf{T} \rangle. \tag{22.41}$$

As earlier, errors of different origins are assumed to be uncorrelated. Since the forcing $\mathbf{f}$ disappears from the error evolution equation, we will not keep it during later developments. Note that, even if we do not write down the matrix $\mathbf{M}$ explicitly, the evolution of the error-covariance matrix $\mathbf{P}^a$ can still be calculated by using instead the actual model on each of its columns $\mathbf{c}$, as shown by operations such as $\mathbf{Mc}$ involved in the error-covariance matrix updates. In order to start the calculation of the error evolution, we need to know the initial value of $\mathbf{P}$, which is related to the error on initial conditions:

$$\mathbf{P}_0 = \left\langle \left(\mathbf{x}_0 - \mathbf{x}^t_0\right)\left(\mathbf{x}_0 - \mathbf{x}^t_0\right)^\mathsf{T} \right\rangle. \tag{22.42}$$

Now, we have a method by which we can calculate the evolution of the error-covariance, and the Kalman filter assimilation is summarized in Fig. 22.6, which includes an extension towards a nonlinear model with linearized error propagation (Extended Kalman Filter, EKF). The analysis step itself is unchanged from Optimal Interpolation, but the error covariance is updated at each assimilation step. In sum, what *Kalman Filtering* adds to Optimal Interpolation is that not only state variables but also their errors are marched in time by the forecast model.

Two extremes are noteworthy. At one extreme, the time between two consecutive assimilations is very short compared with the time scale of evolution of the dynamical process being modeled, and the state variables and their error remain virtually unchanged. The model can then be considered as persistent, with $\mathbf{M} \sim \mathbf{I}$. In this case

$$\mathbf{P}^f_{n+1} \sim \mathbf{P}^a_n + \mathbf{Q}_n. \tag{22.43}$$

In other words, the forecast error is the error of the previous analysis, without advection or any other modification, augmented by the model error introduced during the simulation between assimilations. The last error is relatively small for a well-constructed model because the time integration is short compared with the time scale of interest.

At the other extreme, when assimilation takes place only very infrequently, the model is likely to reach its limit of predictability, and its forecast becomes little more than random. Put another way, the forecast itself is the modeling error. Mathematically this amounts to write $\mathbf{M} \sim 0$ and hence

$$\mathbf{P}^f_{n+1} \sim \mathbf{Q}_n, \tag{22.44}$$

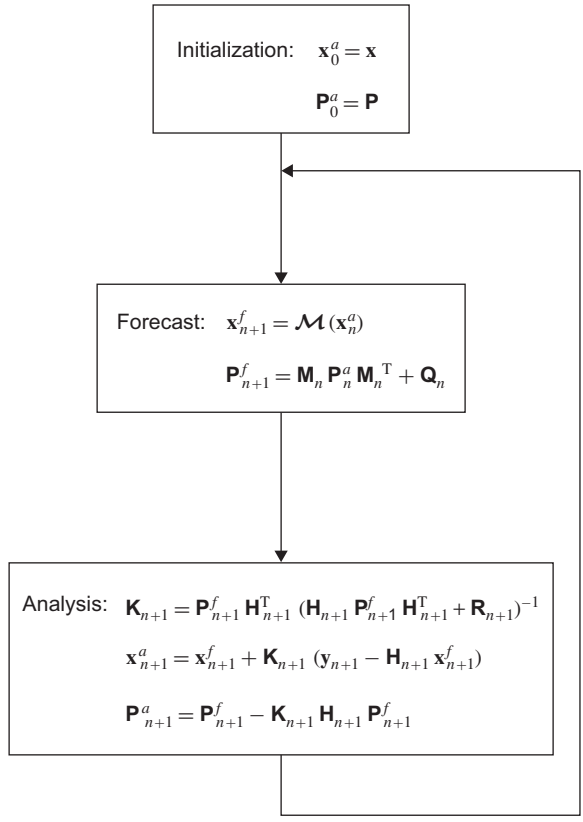

FIGURE 22.6 Sequence of steps followed in the Extended Kalman Filter assimilation scheme with changing observation network (changing matrix $\mathbf{H}$), nonlinear model forecast ($\mathcal{M}$), and model linearization for the error forecast between times of assimilation. Note that strongly nonlinear models, such as those found in ecosystem studies, lead to special problems (e.g., Robinson & Lermusiaux, 2002).

implying that the error field no longer has any memory of the previous error, a fact consistent with the trespassing of the predictability limit, but is entirely due to the simulation. Obviously, $\mathbf{Q}_n$ in this case is much larger than in the previous case because of the long-term integration and error amplifications inherent to the predictability limit.

In order to illustrate the structure of Kalman filtering in a general case, we first note that for the analysis step the error covariance matrix only appears in the combination $\mathbf{P}^f\mathbf{H}^{\mathrm{T}}$,

$$\mathbf{P}^f\mathbf{H}^{\mathrm{T}} = \left\langle \boldsymbol{\epsilon}^f \boldsymbol{\epsilon}^{f\,\mathrm{T}} \right\rangle \mathbf{H}^{\mathrm{T}} = \left\langle \left(\mathbf{x}^f - \mathbf{x}^t\right) \left(\mathbf{H}\mathbf{x}^f - \mathbf{H}\mathbf{x}^t\right)^{\mathrm{T}} \right\rangle \tag{22.45}$$

which is the covariance between the observed quantities and all others. Because it is the matrix that ultimately multiplies a vector of the size of the observational data set (P), it effectively propagates information from the data locations into the model grid. We also have

$$\begin{aligned}\mathbf{A} &= \mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R} = \left\langle \left(\mathbf{H}\mathbf{x}^f - \mathbf{H}\mathbf{x}^t\right)\left(\mathbf{H}\mathbf{x}^f - \mathbf{H}\mathbf{x}^t\right)^{\mathrm{T}} \right\rangle + \mathbf{R} \\ &= \left\langle \left(\mathbf{H}\mathbf{x}^f - \mathbf{y}\right)\left(\mathbf{H}\mathbf{x}^f - \mathbf{y}\right)^{\mathrm{T}} \right\rangle \end{aligned} \tag{22.46}$$

which can be interpreted in terms of error variance of the forecast in the observed part, combined with the corresponding observational error, reminiscent of $\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle$ in Eq. (22.13). This also shows that $\mathbf{A}^{-1}$ exists according to the statistics regarding the innovation vector $\mathbf{H}\mathbf{x}^f - \mathbf{y}$. Should a component of this innovation vector always be zero making $\mathbf{A}$ singular, this would mean that the corresponding model part would never need a correction and that it should therefore be excluded from the data analysis procedure. Also, note that the Kalman gain matrix $\mathbf{K}$ gives greater weight to observations with higher accuracy and that it transmits the corresponding information to other locations.

It is instructive to look at the assimilation of a single observation on the k^{th} component of the state vector $\mathbf{x}$. In this case

- $\mathbf{P}^f\mathbf{H}^{\mathrm{T}}$ is an $M \times 1$ matrix with components P_{ik}^f, $i = 1, \ldots, M$, and it is thus responsible for transferring the innovation learned from observation in location k to the other components of the state vector. The covariance matrix thus appears as the matrix allowing the correction of the fields using remote information. The structure of the covariance depends on the problem at hand (Fig. 22.7). In the example of Fig. 22.7, the error covariance after the analysis is reduced at the data location and then advected by the flow. Hence, downstream of the data point, error covariances are lower and reflect the remote influence of the data.
- $\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$ reduces to a scalar value $\left(P_{kk}^f + \epsilon_0^2\right)^{-1}$ in which ϵ_0^2 is the variance of the observational error and P_{kk}^f the forecast error covariance at the same location.

The error-covariance matrix thus allows the radiation of information from the data location into other regions of the domain and onto other state variables taking into account the relative error of observations and model. Thus, assimilating data from atmospheric temperature profiles might well serve to change winds elsewhere, and sea surface height measurement by an altimeter satellite may well be used to adjust density fields in deeper ocean layers. In contrast to the atmosphere for which a large network of ground-based observation systems and vertical profile soundings by means of balloons have been deployed at reasonable cost, it is far more expensive at sea to maintain a fixed network of moorings or to perform periodic ship cruises in order to sample the ocean

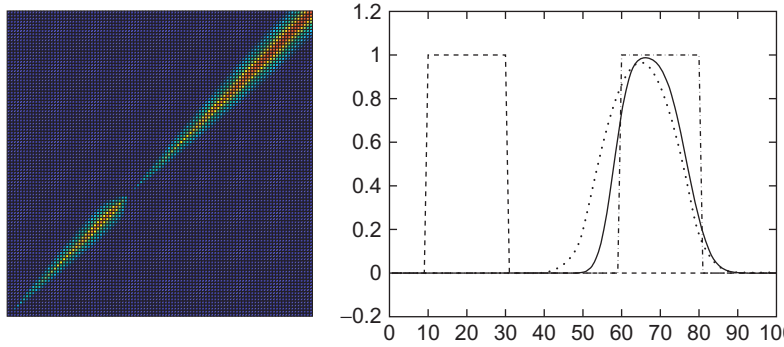

FIGURE 22.7 Assimilation experiment in an advection problem in which data are provided at grid point $i = 40$ at every time step. *Left panel*: The error-covariance matrix (left panel with lower values in black and higher values in lighter shades). The covariance matrix has the largest values along its diagonal, symptomatic of locality. Note, however, that these diagonal values are lower downstream of the data point because the accurate data reduce errors there. *Right panel*: The advected field, initial pattern (square signal on the left), what it should have been at a later time (square signal on the right), what it actually turned out to be with inaccurate velocity and numerical diffusion (dotted curve), and what the prediction became with the help data assimilation (solid curve). See `kalmanupwadv.m` for details of the implementation.

interior. In consequence, satellite data are immensely valuable in ocean forecasts, and it is essential to be able to utilize these surface data to inform density fields and currents in the interior of the ocean.

Because of our minimization of errors using linear combinations and the hypotheses of zero bias, the full Kalman filter is called a *Best Linear Unbiased Estimation* (BLUE) of the true state. Other linear methods presented previously, such as nudging, are suboptimal. It is interesting to note that the Kalman filter approach encompasses in some way the other assimilation methods. For example, if we were to prescribe a priori the forecast error covariance in the Kalman filter instead of letting it be calculated by the model, we would downgrade the Kalman filter to an Optimal Interpolation. If, furthermore, the error and observation covariance matrices were not just prescribed but also taken diagonal, then we recover the nudging scheme. Finally, should the nudging time scale be decreased towards zero, the procedure would be reduced to direct insertion (mere replacement of state variables by corresponding observations). In all cases, the term filter method is appropriate because only past data are utilized to estimate local fields.

22.5 INVERSE METHODS

The Kalman filter operation relies only on past information, the only type of information that is available in an operational forecast mode. Separately, it acknowledges that the dynamic model is not perfect. The procedure of data assimilation optimized with the Kalman filter, however, presents a major

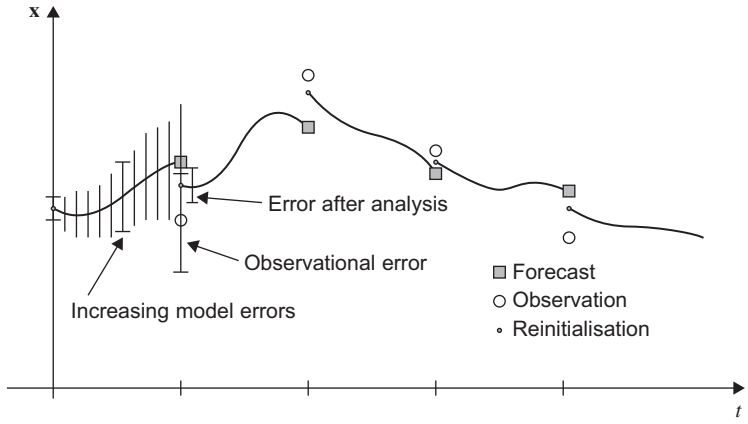

FIGURE 22.8 The Kalman Filter approach leads to a model trajectory that is interrupted each time data assimilation is performed. Vertical bars depicting the error level show that errors creep up during forecast intervals but are suddenly reduced every time assimilation takes place.

drawback in terms of model simulation analyzed over long time periods: The simulated model states (trajectories of state variables) are no longer continuous but exhibit a jump each time data assimilation is performed (Fig. 22.8), and these sudden variations in the values of dynamical variables may cause unphysical shocks. For some applications, it is desirable to avoid any jump to obtain a simulation that preserves the continuity of the physical system (Fig. 22.9).

When we are concerned with obtaining a continuous model trajectory, it is reasonable to assume the dynamical model is perfect except that errors may be introduced by inaccurate initial conditions, incorrect forcing, and possibly also incorrect model parameters such as an eddy-viscosity value. Those errors are then considered responsible for model predictions not conforming with reality. The idea we will now follow is that of optimizing those model components so that the model evolution remains as close as possible to the observations over an extended period of time. The state of the system is at any moment influenced not only by prior data but also by future data. Needless to say, this is not possible in a forecasting mode, but is nonetheless a useful method for re-analysis of past observations with an eye on model improvement. Such a method using data over an interval of time is termed a *smoother*.

With an *inverse model*, the goal is the minimization of model-data misfit over a finite time interval with N data sets $\mathbf{y}_n$ available along the way by searching for an optimal set of initial conditions $\mathbf{x}_0$. The method proceeds by minimizing a so-called *cost function* defined as

$$J = \sum_{n=0}^{N-1} \frac{1}{2} \left(\mathbf{H}_n \mathbf{x}_n - \mathbf{y}_n\right)^{\mathsf{T}} \mathbf{R}_n^{-1} \left(\mathbf{H}_n \mathbf{x}_n - \mathbf{y}_n\right) + J_b. \tag{22.47}$$

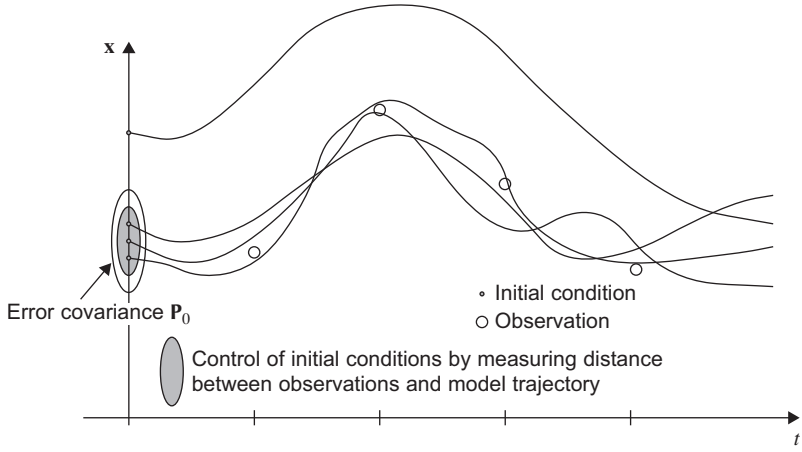

FIGURE 22.9 The adjoint method selects the model trajectory that best fits the observations over time. The initial conditions used to create different trajectories are drawn with a probability density function centered on a background state $\mathbf{x}_0^b$.

The quadratic first term is understood as a weighted sum of squares of misfits between observation ($\mathbf{y}_n$) and corresponding model realizations ($\mathbf{H}_n\mathbf{x}_n$). To this has been added a term J_b, which may become useful should we want to avoid a set of initial conditions $\mathbf{x}_0$ departing too much from a reference (background) field $\mathbf{x}_0^b$. In such case, this extra term takes the form

$$J_b = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right)^{\mathsf{T}} \mathbf{P}_0^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^b\right), \tag{22.48}$$

the minimization of which "pulls" the initial conditions toward their background values. Such a background state $\mathbf{x}_0^b$ of initial conditions can be supplied for example by previous forecasts or climatology.

In addition to the minimization of the cost function J, we need to enforce the constraint that the solution $\mathbf{x}$ is a model result satisfying

$$\mathbf{x}_{n+1} = \boldsymbol{\mathcal{M}}(\mathbf{x}_n). \tag{22.49}$$

For simplicity, however, we again resort to a linearized version,

$$\mathbf{x}_{n+1} = \mathbf{M}_n\mathbf{x}_n. \tag{22.50}$$

An elegant way to satisfy the combination of constraints is to use the method of Lagrange multipliers. By means of these multipliers $\boldsymbol{\lambda}_n$, the set constraints

(22.50) are merged into the cost function to form the following expression:

$$J = \sum_{n=0}^{N-1} \frac{1}{2} (\mathbf{H}_n \mathbf{x}_n - \mathbf{y}_n)^{\mathsf{T}} \mathbf{R}_n^{-1} (\mathbf{H}_n \mathbf{x}_n - \mathbf{y}_n) + \sum_{n=0}^{N-1} \boldsymbol{\lambda}_n^{\mathsf{T}} (\mathbf{x}_{n+1} - \mathbf{M}_n \mathbf{x}_n) + \frac{1}{2} \left(\mathbf{x}_0 - \mathbf{x}_0^b \right)^{\mathsf{T}} \mathbf{P}_0^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^b \right). \tag{22.51}$$

This augmented cost function is then optimized with respect to all variables, including the Lagrangian multipliers. This forms a new variational problem called *4D-Var*. Note that we chose to take the initial observations $\mathbf{x}_0^b$ into account in the cost function as well as all observations $\mathbf{y}_n$ along the way (up to $n = N-1$) but not the simulation outcome $\mathbf{x}_N$, which can thus be regarded as the forecast based on an evolution that is constrained to pass as close as possible to the N previous observations.

The optimization of (22.51) with respect to the Lagrange multipliers (i.e., derivatives with respect to $\boldsymbol{\lambda}_n$ set to zero) returns the model constraints of Eq. (22.50) for $n = 0$ to $N-1$. Variation with respect to the initial state $\mathbf{x}_0$ leads to

$$\nabla_{x_0} J = \mathbf{P}_0^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^b \right) + \mathbf{H}_0^{\mathsf{T}} \mathbf{R}_0^{-1} \left(\mathbf{H}_0 \mathbf{x}_0 - \mathbf{y}_0 \right) - \mathbf{M}_0^{\mathsf{T}} \boldsymbol{\lambda}_0, \tag{22.52}$$

which must vanish for the optimal solution. Variation of Eq. (22.51) with respect to each intermediate state $\mathbf{x}_m$ must also be zero. Realizing that $\mathbf{x}_m$ appears in each sum for both $n = m$ and $n = m-1$, we obtain the following condition in which we switched from m back to n:

$$\mathbf{H}_n^{\mathsf{T}} \mathbf{R}_n^{-1} (\mathbf{H}_n \mathbf{x}_n - \mathbf{y}_n) - \mathbf{M}_n^{\mathsf{T}} \boldsymbol{\lambda}_n + \boldsymbol{\lambda}_{n-1} = 0 \qquad n = 1, \ldots, N-1. \tag{22.53}$$

Finally, variation with respect to the final state $\mathbf{x}_N$ simply provides $\boldsymbol{\lambda}_{N-1} = 0$.

The different conditions can be recast into the following algorithm. We start with an estimate $\mathbf{x}_0$ of the initial condition and then perform the following operations

$$\mathbf{x}_{n+1} = \mathbf{M}_n \, \mathbf{x}_n, \qquad n = 0, \ldots, N-1 \tag{22.54a}$$

$$\boldsymbol{\lambda}_{N-1} = 0 \tag{22.54b}$$

$$\boldsymbol{\lambda}_{n-1} = \mathbf{M}_n^{\mathsf{T}} \boldsymbol{\lambda}_n - \mathbf{H}_n^{\mathsf{T}} \mathbf{R}_n^{-1} (\mathbf{H}_n \mathbf{x}_n - \mathbf{y}_n), \qquad n = N-1, \ldots, 1. \tag{22.54c}$$

In this process, we note that model-data misfits $\mathbf{H}_n \mathbf{x}_n - \mathbf{y}_n$ are driving the values of the Lagrange multipliers. All stationary conditions imposed on Eq. (22.51) are now satisfied except that $\nabla_{x_0} J$ given by Eq. (22.52) is not yet zero. The recurrence (22.54c) on the Lagrange multiplier can formally be extended to $n = 0$ so that $\boldsymbol{\lambda}_{-1}$ takes on the value of $-\nabla_{x_0} J$ if no background field is used ($\mathbf{P}_0^{-1} = 0$).

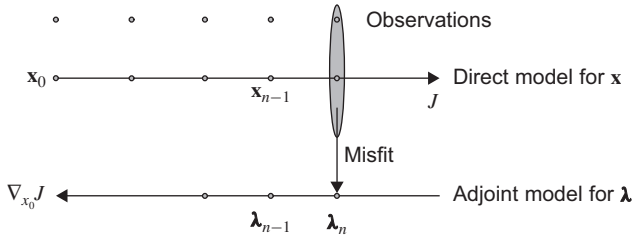

FIGURE 22.10 The forward integration starts from an initial guess of the control (adjustable) parameters and then provides the state variable over the simulation window. Misfits between observations and model results are stored, and error norms of the cost function J defined by Eq. (22.35) are accumulated. Then, the adjoint model is integrated backward in time, with the misfits used as forcing. Upon returning to the start $n = 0$, the gradient (22.52) can be calculated. If it is not zero, the optimum is not reached and an improved guess of the control parameters can be calculated by minimization tools using the cost function value J and its gradient with respect to the control parameters.

But because we now have access to a value of J and its gradient with respect to the variables on which we optimize our solution, the initial condition, we may use any mathematical minimization tool that searches for such minimum using gradients. The steepest descent or the most efficient *conjugate gradient* methods (see Section 7.8) are iterative methods that generate a succession of states (here for $\mathbf{x}_0$) that decrease the value of the cost function J depending on the gradient of the function. We therefore now have a relatively simple way of calculating those gradients by performing a forward integration of the model, called the direct model, and a backward integration to evaluate the Lagrange multipliers, called the inverse model, and finally the gradients (Fig. 22.10).

This apparently simple procedure masks several practical problems, however. The equations for the Lagrange multipliers are very similar to those of the direct model with the seemingly innocent difference that, instead of matrix $\mathbf{M}$, its transpose appears, and instead of applying $\mathbf{M}$ we apply its adjoint to $\boldsymbol{\lambda}_n$ to create the time series. Therefore, we speak of the *adjoint model* when referring to the backward integration for the Lagrange multipliers. In practice, since a numerical model never explicitly creates the matrix $\mathbf{M}$, this means that actual programming of an adjoint, nonlinear model is necessary, the action of which on $\boldsymbol{\lambda}$ is equivalent to applying the transpose model matrix. Another practical problem for time-varying models is the need to store intermediate model results, or to regenerate them on a need basis, over the entire simulation interval because of the backward integration. For more details on implementation aspects of the adjoint method, the reader is referred to Courtier, Thepaut and Hollingsworth, 1994 and for early theoretical developments to Talagrand and Courtier (1987).

The method is readily extended to optimizations of parameter values such as eddy viscosity and boundary conditions. Parameters to be optimized, called *control parameters*, can, for example, be introduced as additional state variables into the state vector, with an evolution equation of persistence. In all cases, it

must be kept in mind, however, that any inversion is only valid as long as the direct model itself is able to simulate correct results. In other words, should an inverse method be used on a grossly inadequate model in an attempt to improve it by estimating suitable parameter values, these parameter values run the risk of no longer having any physical meaning and to be just ad-hoc values. It might therefore be necessary instead to relax the hypothesis of a dynamically correct model and allow for modeling errors, as done when a Kalman filter is used.

Allowing the model solution to deviate somewhat from a pure model solution is achieved by replacing the strong constraint (22.50) with a so-called *weak constraint*, penalizing only strong departures from Eq. (22.50), with the stiffest penalties reserved for models that are most trustworthy. Doing so, we can adapt the cost function of the inverse model to form a *generalized inverse model* by minimizing

$$\begin{aligned} J = &\sum_{n=0}^{N-1} \frac{1}{2} (\mathbf{H}_n \mathbf{x}_n - \mathbf{y}_n)^{\mathsf{T}} \mathbf{R}_n^{-1} (\mathbf{H}_n \mathbf{x}_n - \mathbf{y}_n) \\ &+ \sum_{n=0}^{N-1} \frac{1}{2} (\mathbf{x}_{n+1} - \mathbf{M}_n \, \mathbf{x}_n)^{\mathsf{T}} \mathbf{Q}_n^{-1} \, (\mathbf{x}_{n+1} - \mathbf{M}_n \, \mathbf{x}_n) \\ &+ \frac{1}{2} \left(\mathbf{x}_0 - \mathbf{x}_0^b \right)^{\mathsf{T}} \mathbf{P}_0^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^b \right), \end{aligned} \tag{22.55}$$

where $\mathbf{Q}$ is the error covariance of the model.

Differentiation with respect to the initial condition provides

$$\nabla_{x_0} J = \mathbf{P}_0^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^b \right) + \mathbf{H}_0^{\mathsf{T}} \mathbf{R}_0^{-1} \left(\mathbf{H}_0 \mathbf{x}_0 - \mathbf{y}_0 \right) - \mathbf{M}_0^{\mathsf{T}} \mathbf{Q}_0^{-1} \, (\mathbf{x}_1 - \mathbf{M}_0 \, \mathbf{x}_0), \tag{22.56}$$

while differentiation with respect to intermediate states ($n = 1, \ldots, N-1$) leads to the additional conditions

$$\begin{aligned} &\mathbf{H}_n^{\mathsf{T}} \mathbf{R}_n^{-1} (\mathbf{H}_n \mathbf{x}_n - \mathbf{y}_n) - \\ &\mathbf{M}_n^{\mathsf{T}} \mathbf{Q}_n^{-1} \, (\mathbf{x}_{n+1} - \mathbf{M}_n \, \mathbf{x}_n) + \mathbf{Q}_{n-1}^{-1} \, (\mathbf{x}_n - \mathbf{M}_{n-1} \, \mathbf{x}_{n-1}) = 0, \end{aligned} \tag{22.57}$$

and differentiation with respect to the final state provides $\mathbf{x}_N = \mathbf{M}_{N-1} \, \mathbf{x}_{N-1}$.

This system of equations can be written in a more familiar way as

$$\mathbf{x}_n = \mathbf{M}_{n-1} \, \mathbf{x}_{n-1} + \mathbf{Q}_{n-1} \boldsymbol{\lambda}_{n-1}, \qquad n = 1, \ldots, N \tag{22.58a}$$

$$\boldsymbol{\lambda}_{N-1} = 0 \tag{22.58b}$$

$$\boldsymbol{\lambda}_{n-1} = \mathbf{M}_n^{\mathsf{T}} \boldsymbol{\lambda}_n - \mathbf{H}_n^{\mathsf{T}} \mathbf{R}_n^{-1} (\mathbf{H}_n \mathbf{x}_n - \mathbf{y}_n) \qquad n = N-1, \ldots, 1 \tag{22.58c}$$

with the need for gradient (22.56) to be zero. These equations are very similar to those of the adjoint method (22.54) with the additional term involving $\mathbf{Q}$ in the direct model allowing the propagation of model errors. Though similar, the practical solution is now more complicated than for the adjoint method and can be

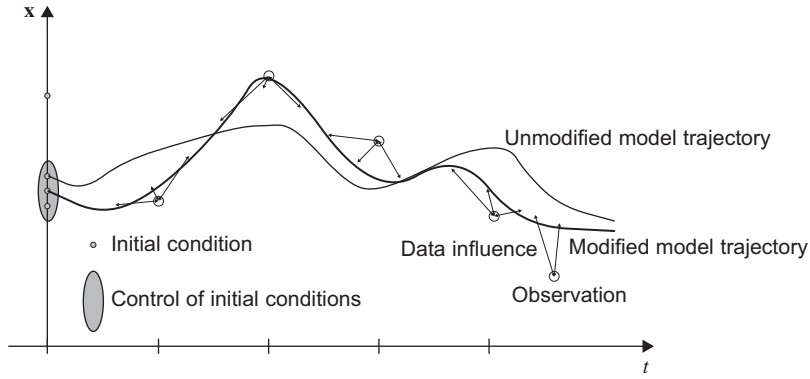

FIGURE 22.11 Generalized inverse methods allow the simultaneous optimization of initial conditions and the minimization of the departure of the model from observations.

obtained with the so-called *representer method* (Bennett, 1992). The generalized inverse method thus exhibits a solution that is not a true solution of the direct model but one that effectively reaches a compromise between observations and a true model solution (Fig. 22.11).

Variational approaches are attractive because they permit efficient model calibration, but error covariances of observations and model need to be prescribed a priori, which is a handicap. More importantly, except for the value of the cost function, which is an overall measure, the analysis is not accompanied by an error estimate on the final analysis contrary to the Kalman filter (22.33). In principle, it is possible to obtain such error estimates with variational methods by having recourse to the second derivatives of the cost function with respect to the control parameters, the Hessian matrix. Intuitively, if the cost function is decreasing sharply (i.e., if it can be said that it exhibits a "deep well"), the optimum is much better constrained than for a relatively flat cost function (Fig. 22.12). The Hessian (curvature) matrix allows therefore the calculation of the error covariance, a property demonstrated in Rabier and Courtier (1992).

Another way to obtain smooth trajectories with error estimates is to generalize the Kalman filter to include not only past values but also future ones. The so-called *Kalman smoother* does this and provides, for linear systems, identical results to those of the generalized inverse approach over the whole simulation interval.

Similarly it can be shown that the standard 4D-Var method and Kalman filter are equivalent to each other for a perfect model with identical data over a given time interval and initial error covariance, in the sense that they lead to the same final analysis for linear models and linear observation operators.

Although equivalence has been proven in some cases, the methods tend to lose their overlaps when the underlying hypotheses, such as linearity of the

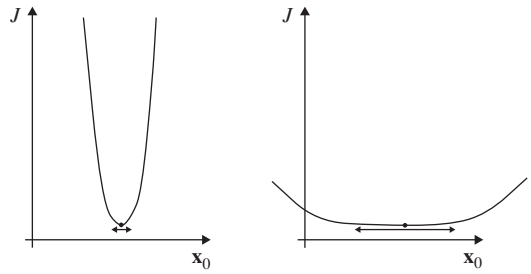

FIGURE 22.12 When the Hessian matrix of J—the set of its second derivatives with respect to adjustable parameters—is large, the minimum tends to be well constrained ("deep well"—left panel), and errors in the final state are small. On the contrary, errors in the optimal solution will be larger if the cost function exhibits a wide range over which the adjustable parameters may vary without penalizing significantly the value of J ("shallow well"—right panel).

model and unbiased information, are violated. The different methods then lead to different outcomes. Also, the practical implementation can differ significantly from method to method and can widely affect their algorithmic performance.

22.6 OPERATIONAL MODELS

Practical implementation of the aforementioned data assimilation methods is a concern when considering the typical number of operations that are to be performed routinely with an operational model. The nudging method barely adds a linear term to existing equations, and the overhead cost associated with this procedure is negligible. For Optimal Interpolation in its original form, there arises at least the need to invert a matrix of size proportional to the number of observations.[6] Since the matrix $\mathbf{H}$ is generally mostly composed of zeros, we need not pay attention to costs associated with multiplications by $\mathbf{H}$. Nevertheless, for a full matrix $\mathbf{HPH}^{\mathrm{T}}+\mathbf{R}$, the cost of its inversion is roughly P^3 operations. In addition, we need to store $\mathbf{HP}$ of size $M \times P$. This amounts to a major burden on computer resources if P is large. If we want to calculate the error covariance after the analysis, the required matrix multiplications demand PM^2 operations. If only the diagonal of the covariance matrix, i.e., if only the local error variances are sought, the number of operations drops to PM, still a large number.

The computational demand, of course, rises if we update the error covariance over time as needed in Kalman filtering. Granted, we never construct the matrix $\mathbf{M}$ to march the model forward, but multiplication by this matrix with a vector actually represents a model integration. Hence, if the matrix $\mathbf{M}$ multiplies another matrix of size $M \times M$, the cost estimate is that of M model integrations,

[6]This is assuming the covariance matrix $\mathbf{P}$ or, rather, $\mathbf{HP}$ can be easily calculated.

which is therefore the cost of updating the forecast error covariance, which must be compared with the cost of a single model integration as when no assimilation is performed.

For the 3D-Var method, the situation is not any better because it necessitates M^3 operations to invert each covariance matrix, unless their inverses are provided. In that case, we need M^2 operations to calculate each cost function, M operations to calculate its gradient, and K iterations to find the minimum. In order to reach the minimum, $K = M$ iterations are needed in theory, but good approximations can be found for K much less than M.

For 4D-Var, similar estimates apply with the need to perform a direct model integration and its adjoint, reverse integration for each evaluation of the gradient needed during the minimization. For K iterations, KM model integrations are required, while each evaluation of the cost function along the way requires M^2 operations, unless the error covariances have special forms (e.g., a diagonal matrix).

In view of the numbers M (on the order of 10^7 or 10^8) of model variables and P (on the order of 10^6) of observations in use for operational models, we clearly see a need for simplification. Otherwise, for the sizes of state vectors and observation arrays currently in use, Kalman filters and 3D-Var are out of the question unless covariance matrices are of a very special form. Reducing the complexity of an assimilation method to a computationally affordable version while retaining its advantages is where the "art" of modeling comes into play, by sifting the unnecessary elements from the essential features.

The type and quantity of observation greatly affects the possibility of simplifications. In the ocean, for operational purposes, observations are mostly surface data from satellites (sea-level anomaly with respect to an average dynamical position, sea surface temperature, color, sea ice and, hopefully in the future, salinity as well) and coastal data (sea level at tide gages), with Argo floats (e.g., Taillandier, Griffa, Poulain & Béranger, 2006) complementing the information with a few deep profiles. In the atmosphere, observational networks are much denser than at sea, and, in particular, daily radiosoundings from all over the globe contribute to data assimilation. Simplified assimilation methods generally take into account the type of data incorporated. For the ocean, one approach is to perform assimilation in two steps, one incorporating only surface information and the other profile data, with moreover specific simplifications brought to the covariances.

Reduction of complexity can be justified in most cases because the system's evolution includes a series of physically damped modes, which do not require correction since they fade away. Damped modes occur through the existence of attractors such as geostrophic equilibrium. However, unstable modes, so characteristic of our weather patterns, must absolutely be followed by the assimilation process.

Another avenue for simplifications is the size of the state vector. The very large number of numerical state variables results mostly from our numerical

need to have the grid size short compared with the scales of interest, i.e., $\Delta x \ll L$ and $\Delta z \ll H$. But this immediately indicates that we are using many more calculation points than the actual number of degrees of freedom that are needed. We can therefore try to downsize the burden of the data assimilation procedure from the number of computational points to the significantly lower number of significant degrees of freedom in the dynamics or number of observations.

One of the most popular approaches to simplify the assimilation procedure is the use of a *reduced rank* covariance matrix, by writing

$$\mathbf{P} \sim \mathbf{S}\mathbf{S}^{\mathsf{T}} \tag{22.59}$$

in which the reduced matrix $\mathbf{S}$ is of size $M \times K$, which is significantly less than $M \times M$ because K is much smaller than M. It is easily demonstrated that the rank[7] of $\mathbf{S}\mathbf{S}^{\mathsf{T}}$ is at most K, hence the name reduced rank. When adopting this simplification, we no longer need to store $\mathbf{P}$ because we store the much smaller matrix $\mathbf{S}$ and any matrix multiplication involving $\mathbf{P}$ reduces to successive matrix multiplications by $\mathbf{S}$ and its transpose. A multiplication of $\mathbf{P}$ with a square matrix of identical size $M \times M$ no longer requires M^3 operations but only $2KM^2$ operations.

We can also achieve reduction by saving computations at the level of matrix inversion. The effect of a reduced rank can be exemplified most easily with a diagonal matrix $\mathbf{R}$ of uncorrelated observational errors with the same error variance μ^2 at all locations. Defining the matrix $\mathbf{U} = \mathbf{H}\mathbf{S}$ of dimension $P \times K$ with $K \ll P$, we have

$$\mathbf{P}\mathbf{H}^{\mathsf{T}}\left(\mathbf{H}\mathbf{S}\mathbf{S}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}} + \mathbf{R}\right)^{-1} = \mathbf{S}\mathbf{U}^{\mathsf{T}}\left(\mathbf{U}\mathbf{U}^{\mathsf{T}} + \mu^2\mathbf{I}\right)^{-1} = \mathbf{S}\left(\mathbf{U}^{\mathsf{T}}\mathbf{U} + \mu^2\mathbf{I}\right)^{-1}\mathbf{U}^{\mathsf{T}}. \tag{22.60}$$

The last equality can be demonstrated directly by matrix operations or by using a special case of arguably the most useful matrix identity in data assimilation, the Sherman-Morisson formula (Analytical Problem 22.8). The last operation in Eq. (22.60) transforms the matrix to be inverted from a $P \times P$ matrix into a much smaller $K \times K$ one. This is where a major gain in computing is obtained. The same gain is attained when $\mathbf{R}$ is not diagonal but has an inverse that can be calculated easily.

In an ensemble forecast approach (e.g., Houtekamer & Mitchell, 1998), the direct model is used to create a series of simulations, each one being a slightly perturbed version of the others, with perturbations introduced through initial conditions, forcings, parameter values, or even topographic modifications. The ensemble of model results thus obtained permits statistical parameters to be estimated from the ensemble members. In practice, the convergence of variance estimations from K samples converges only as $1/\sqrt{K}$, and we therefore

[7] The rank of a matrix is the number of linearly independent columns.

anticipate needing a large ensemble or somehow creating an ensemble with optimal distributions of its members (e.g., Evensen, 2004).

Combining ensemble approaches and reduced rank approximations leads to a series of data assimilation variants with different implementations (e.g., Barth, Alvera-Azcárate, Beckers, Rixen & Vandenbulcke, 2007; Brasseur, 2006; Lermusiaux & Robinson, 1999; Pham, Verron & Roubaud, 1998; Robinson et al., 1998). The ensemble approach can even be extended to include not only simulation results of a single model with perturbed setups but also different models, aimed at modeling the same properties (super-ensemble approach) or even using different physical parameterizations and governing equations (hyper-ensemble approach). Such combinations can dramatically reduce errors, in particular biases (e.g., Rixen & Coelho, 2007).

Other simplifications are based on the reduction of the size of the state vector on which assimilation works. One possibility is to propagate error covariances with coarser grids or simplified models (e.g., Fukumori & Malanotte-Rizzoli, 1995). Particular dynamical balances can also be taken into account through covariances. If only one component of such a balance is observed, for example, sea surface height in a geostrophically balanced system, velocity does not need to be included in the assimilation step. Once a correction of sea surface height and density is made, corrections to velocity can be calculated by using geostrophy.

Operational models have been in use for some applications for quite some time, with weather forecasts based on numerical models having been initiated in the postwar period. These are now widespread with two major centers providing global forecasts (ECMWF and NCEP). Operational tidal models, hurricane predictions, and tsunami warning systems in the Pacific are also well established, incorporating data from observational networks at dedicated institutes. Ocean circulation forecasts began to appear (Mersea, Hycom, Hops etc.) well before public demand for tsunami predictions in December 2004.

A common aspect of operational models is that data assimilation was initially developed to reduce errors. It is fair, however, to request that operational models not only include forecasts but also provide the associated uncertainty by means of confidence intervals, which assimilation now permits to do. From a scientific point of view, we might well argue that the forecast corrections do not provide new insight into physical processes. In reality, the analysis of assimilation cycles can help in understanding error sources and verify that model results are statistically consistent with observations. Also verifying that the innovations (model-data departures) and error estimates are compatible with the statistical models in use allows us to identify inconsistencies. For example, innovations should on average be zero, lest a bias arises in the system that needs correction. Forecast verifications (Jolliffe & Stephenson, 2003) can teach us not only about dynamics but also about model and observation errors and can help to identify the model or observing system components that are most in need of improvement. Finally, strategies for adaptive sampling can emerge from such studies (e.g., Lermusiaux, 2007).

ANALYTICAL PROBLEMS

22.1. Analyze the exact solution of the following equations

$$\frac{du}{dt} = +\tilde{f}v - \frac{u - u_o(t)}{\tau} \tag{22.61}$$

$$\frac{dv}{dt} = -\tilde{f}u - \frac{v - v_o(t)}{\tau} \tag{22.62}$$

with $u_o(t) = \cos(ft)$, $v_0 = -\sin(ft)$ and a relaxation time $\tau = 1/(\alpha\tilde{f})$. These equations represent nudging to follow an inertial oscillation of frequency f while the "model" has a tendency to generate inertial oscillations with incorrect frequency $\tilde{f}$. Identify how nudging corrects an error on the initial condition, distinguishing the effect of an error on amplitude and phase. Then investigate how a difference between $\tilde{f}$ and f affects the solution.

22.2. Establish and describe an Optimal Interpolation method to interpolate in space knowing the spatial covariance of the true field and data errors at given locations.

22.3. An analysis method provides both optimal fields and their corresponding error estimates, as shown in Fig. 22.13 under the assumption that the background (or forecast) field has a uniform error variance. In view of the data distribution, the error field, and spatial structure of the analysis, how would you rank the following measures

- observational error
- forecast error

and include an estimate of their values? In your opinion, what is a rough estimate of a correlation length used for the analysis of the field? (*Hint*: Think about the way observations are propagated into the domain and the significance of covariance functions. In particular, investigate how the analysis behaves far away from any observation and then near a single observation far away from any other data location. For better reading of the fields displayed in Fig. 22.13, refer to `divashow.m`.)

22.4. Prove that the minimization of Eq. (22.35) leads to the same analysis as the Optimal Interpolation. What is the value of J at the optimum?

22.5. A weather forecast is typically limited to one or two weeks, after which the forecast skill is nil. Can you justify why climate models, which use the same governing equations of geophysical fluid dynamics, may be used to predict decades ahead? (*Hint*: Think about the significance of state variables and parameterizations.)

22.6. Bottom friction is generally parameterized by a quadratic law $\tau_b = \rho_0 C_d u^2$. Show that, even for an unbiased estimate of u, the bottom-stress estimate itself is biased. What is the error variance?

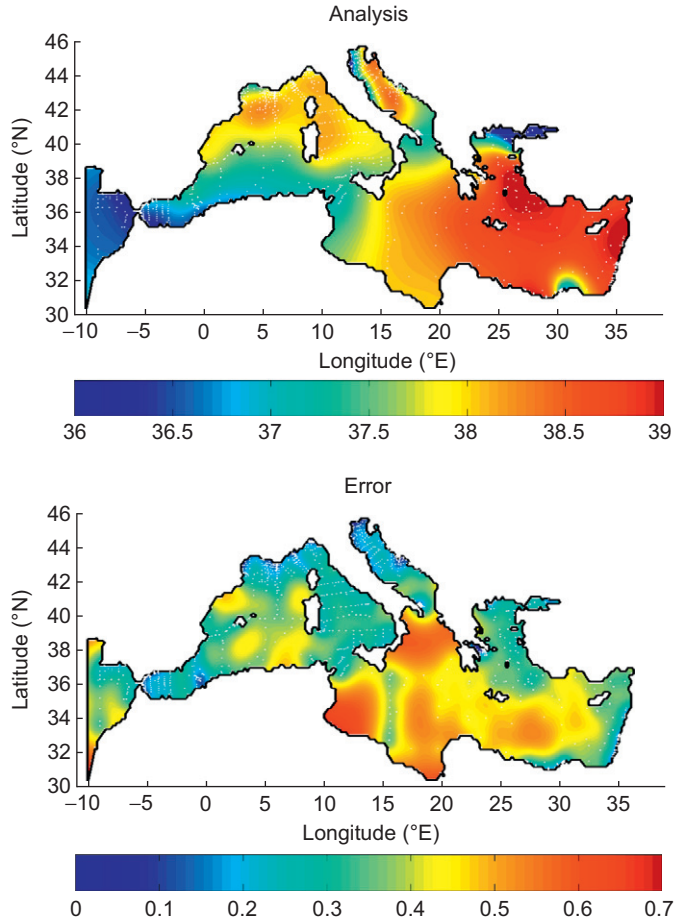

FIGURE 22.13 *Upper panel*: An Optimal Interpolation of surface-salinity data in the Mediterranean Sea with data locations indicated by white dots. *Lower panel*: The corresponding analysis error (in the same units as salinity), assuming uniform observational errors and a uniform forecast error. Note the impact of data distribution on the error estimates. (*Data Interpolating Variational Analysis (Diva), Brasseur, Blayo & Verron, 1996*).

22.7. Prove the following identity:

$$\mathbf{L}^{\mathrm{T}}\left(\mathbf{L}\mathbf{L}^{\mathrm{T}}+\mu^2\mathbf{I}\right)^{-1}=\left(\mathbf{L}^{\mathrm{T}}\mathbf{L}+\mu^2\mathbf{I}\right)^{-1}\mathbf{L}^{\mathrm{T}}. \tag{22.63}$$

Are there conditions to be satisfied for this to hold true?

22.8. Prove the Sherman-Morrison formula:

$$\left(\mathbf{A}+\mathbf{U}\mathbf{V}^{\mathrm{T}}\right)^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{U}\left(\mathbf{I}+\mathbf{V}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}^{\mathrm{T}}\mathbf{A}^{-1}. \tag{22.64}$$

22.9. Should the full state vector **x** be known with zero observational error, how would the Kalman gain matrix behave?

22.10. Derive the equation that steady-state forecast error covariance must satisfy, assuming the error-covariance matrix **P** reaches a stationary value. Show that it is a Ricatti equation. When do you expect the use of this error covariance to become interesting?

22.11. Can you see a reason for including an error of a perfect model ϵ^p different from zero in the definition (22.1) of the skill score? (*Hint*: Think about what a perfect model is able to predict.)

22.12. When do you think autocorrelation provides useful information on the predictability limit discussed in Fig. 22.4?

22.13. The Lyapunov equation (22.40) can also be used to recover classical results in error estimations. Suppose you use a state equation $\rho(T, S)$ and your measurements on T and S are plagued by errors with known error variance σ_T^2 and σ_S^2. Calculate the expected error variance on ρ and show that it can be recast into the form (22.40). Can you interpret **Q** in this case?

22.14. Altimetric data allow the detection of changes in sea surface elevation $\Delta\eta$ between successive passes of the satellite. If potential vorticity is conserved in the ocean at large scales, we can expect that the vertical density structure has not changed before the return of the satellite but has only been displaced vertically over a distance Δz. Calculate this displacement assuming that it is small and that pressure at a level of no-motion at $z = -z_0$ has not changed. Neglect density variations near the surface. What is the sign of the displacement if sea surface height is increased between the two passes? Interpret your finding in terms of baroclinic modes and explain how the result can be used for a simple data assimilation scheme (see Cooper & Haines, 1996).

NUMERICAL EXERCISES

22.1. Implement a nudging method and solve numerically Analytical Problem 22.1. When you decrease τ, what specific action do you need to take during time stepping? Investigate what happens when you add noise to the "observations" u_0 and v_0 and decrease the sampling rate of the "pseudo data."

22.2. Work with an Optimal Interpolation scheme on a one-dimensional gravity-wave problem already discretized in `oigrav.m` with 61 interior calculation points in a 61 km long and 100 m deep domain. Perform a twin experiment with a reference simulation starting from rest with a sinusoidal elevation of 40 cm amplitude and wavelength equal to the

domain size. From this simulation, you can extract pseudo data in order to compare at least two assimilation strategies:

- A sampling point for surface elevation η located near the left boundary, at one third of the total domain length, sampled, and assimilated every 10 seconds.
- An observation of the surface elevation at all grid points sampled and assimilated every 10 minutes.

In both cases, add random noise to your "data" with a standard deviation representative of altimetric precision (2 cm). Assume the noise is uncorrelated in time and space. For the assimilation experiment, start from initial conditions at rest and zero sea-level height.

Prescribe the forecast covariance as a function of distance between points x_i and x_j, proportional to the following correlation function

$$c(x_i, x_j) = \exp\left[-(x_i - x_j)^2/R^2\right] \tag{22.65}$$

where R is a correlation length. The covariance is obtained by multiplying this correlation function by the estimated variance of the forecast.

The observational error covariance can be specified according to the perturbation added. Look at the evolution of the simulations and quantify the error individually on η and velocity.

Which variance of the background field would you advocate in view of the initial conditions? Change the value and assess the effect on the assimilation behavior. Then, change the value of correlation length and finally try other combinations of space-time coverage. (*Hint*: For ease of programming, the two simulations (reference run and assimilation run) may be advanced simultaneously, and you can diagnose the error as you go, before assimilating, however.)

22.3. Experiment with `kalmanupwadv.m` by changing the error specifications, sampling rate, and sampling locations.

22.4. Develop an adjoint method for Numerical Exercise 22.3 and optimize the initial condition using the same "observations" as for the Kalman filter of Numerical Exercise 22.3. Use the exact advection velocity with $C = U\Delta t/\Delta x = 0.2$ and then $C = 1$ in order to witness the effect of model errors, remembering that $C = 1$ corresponds to perfect dynamics because the upwind scheme is exact in this case. What is the effect of reducing the number of observations and disregarding the background initial conditions? (*Hint*: Think about underdetermined and overdetermined problems.)

22.5. Estimate the computer memory needed to store **P** for today's weather forecast systems.

22.6. Implement an ensemble Kalman filter for the Lorenz equations (22.2) and the assimilation of observations of $x(t)$ only. Investigate the effect of the observation frequency by performing assimilation steps at 0.001, 0.01, 0.1, and 1 time intervals.

22.7. Work to improve the Optimal Interpolation of Numerical Exercise 22.2 by calculating and using covariance functions estimated from an ensemble run in which you perturb the initial condition.

22.8. Improve the assimilation scheme of Numerical Exercise 22.2 by implementing the full Kalman filter with an updated covariance matrix. For ease of programming, you might consider constructing explicitly the transition matrix $\mathbf{M}$.

Michael Ghil
1944–

Born in Budapest, Hungary, Michael Ghil spent his high-school years in Bucharest, Romania and then acquired an engineering education in Israel, where he served as a naval officer. After moving to the United States, he obtained his doctorate at the Courant Institute of Mathematical Sciences of New York University under Professor Peter Lax (see biography at end of Chapter 6).

His scientific work includes seminal contributions to climate system modeling, chaos theory, numerical and statistical methods, data assimilation, and mathematical economics. He provided self-consistent theories of Quaternary glaciation cycles, of the low-frequency variability of extratropical atmospheric flows, and of the midlatitude oceanic interannual variability. He is a prolific writer, with his name attached to a dozen books and well over two hundred research and review articles.

Professor Ghil takes turns teaching at the École Normale Supérieure in Paris and at the University of California in Los Angeles. He has enjoyed learning from and working with a large number of students and younger colleagues, with whom he often stays in contact. Many of these have attained considerable professional achievements in their own right, on three continents. Ghil speaks six languages fluently. (*Photo by Philippe Bruère, Compagnie des Guides, Chamonix—Mont Blanc*)

Eugenia Kalnay
1942–

Eugenia Kalnay was awarded a Ph.D in Meteorology from the Massachusetts Institute of Technology under the direction of Jule Charney (see biography end of Chapter 16). Following a position as Associate Professor in the same department, she became Chief of the Global Modeling and Simulation Branch at the NASA Goddard Space Flight Center, where she developed the accurate and efficient "NASA Fourth Order Global Model." Later as Director of the Environmental Modeling Center of the US National Weather Service (NWS), she spearheaded major improvements in the forecast skills of the NWS models. Many successful projects such as ensemble forecasting, 3D and 4D variational data assimilation, advanced quality control, seasonal and interannual dynamical predictions, were started under her leadership. She also directed the NCEP/NCAR 50-year Reanalysis Project, and the resulting re-analysis paper of 1996 is one of the most cited paper in all of geosciences. Moving to academia, Professor Kalnay cofounded at the University of Maryland the Weather-Chaos Group, a leader in Ensemble Kalman Filter methods.

Over the years, Kalnay has received numerous awards, including the prestigious IMO prize of the World Meteorological Organization, for her contributions to numerical weather prediction, data assimilation, and predictability. Kalnay, a key figure in this field, has pioneered many of the essential techniques, which she describes in her book *Atmospheric Modeling, Data Assimilation, and Predictability* (2003). (*Photo by Zhao-Xia Pu, used with permission*)