1

Methods for Assimilation of Observations: Application to Numerical Weather Prediction

Olivier TALAGRAND
LMD/IPSL, CNRS, ENS, Université PSL, Institut polytechnique de Paris, Sorbonne Université, Paris, France

1.1. Introduction

This chapter presents a general description of *assimilation of observations*. The guideline of the presentation is the development of assimilation in the context of numerical weather prediction. The case of numerical weather prediction is rather emblematic in that it has led to the gradual development of more and more powerful numerical algorithms for assimilation of observations, which have in turn significantly contributed to the improvement of weather prediction, while they were at the same time expanding to many diverse applications.

Weather prediction, like many other fields of present science and technology, is now largely dominated by numerical modeling. Meteorological forecasts are basically produced by numerical models which, started from the observed meteorological conditions, compute the expected evolution of the relevant parameters (pressure, temperature, wind and humidity) over periods which, depending on the particular purpose at hand, may vary from a few hours or less, to weeks or months. The numerical models are built on the physical laws that govern the evolution of the

Inversion and Data Assimilation in Remote Sensing,
coordinated by Yajing YAN. © ISTE Ltd 2024.

flow (namely, the laws of conservation of mass, energy and momentum, or at least as accurate a formulation of those laws as can be obtained). These laws are discretized on an appropriate three-dimensional spatial grid and integrated in time over the period of the forecast. As an example, the present operational model of the European Centre for Medium-Range Weather Forecasts (ECMWF), located in Reading in England, is built on a grid which covers the whole volume of the atmosphere, with horizontal and vertical resolutions of about 9 km and 2 km, respectively. This leads for the description of the instantaneous state of the whole atmosphere to a vector with dimension over 10^9. Two daily forecasts are produced at present by evolving the state vector over a 10-day period (with a time step of 450 s).

With regard to the initial conditions from which the forecasts must be started, the observing system consists of a large variety of different instruments. Radiosondes are launched from ground stations at conventional synchronous "synoptic" hours. The ground stations are, however, concentrated over continents, mostly in developed countries. Satellites measure the radiation emitted by the planet, providing information on the thermal structure of the atmosphere. They cover the surface of the Earth more homogeneously than ground stations, but perform observations continuously in time. Other types of observations (as performed, for instance, by commercial planes or drifting buoys at the surface of the ocean) are also available. At present, the number of scalar observed quantities that are used in operational weather prediction over a 24-h period lies in the range 10^7–10^8.

A large fraction of these observations is distributed over time, and there is much more information in the observations distributed over 12 or 24 h, over which the flow can significantly evolve, than in synoptic observations performed at a given time. *Assimilation* is the process through which all that information is combined in order to define the initial conditions of the forecast. Its purpose can be precisely defined as estimating the state of the observed system as accurately as possible using all available relevant information. That information essentially consists of two different parts. The observations stricto sensu, on the one hand, may vary in nature, accuracy, as well as in spatial and temporal resolution and distribution, and the physical laws that govern the flow, on the other hand, are available in practice in the form of a numerical and necessarily approximate model.

Assimilation is one of many inverse problems that are encountered in many fields of science and technology. Inverse problems arise in situations when we want to know the state of a system (often a physical system, but not necessarily so), while the available data are not in the format appropriate for the purpose at hand. They are encountered, for instance, in solid Earth geophysics (when we want to know the internal structure of the planet), non-destructive probing of structures of various types, navigation of aircraft, spacecraft or any mobile object, as well as in many other applications as described in the other chapters of the book. In most of these problems, data are affected with some uncertainty, and we may wish to know the resulting uncertainty on the final estimate. From a mathematical point of view, uncertainty is conveniently described by probability distributions. This leads to the fact that many inverse problems are stated as problems in Bayesian estimation, namely determining the probability distribution for the state of the system of interest, conditioned by the available data. Indeed, although the basic nature of the problems considered may be very different, the mathematical equations that are used for solving these problems are often the same (as can be seen in several chapters of this book).

It is in this general Bayesian perspective that assimilation of meteorological observations is described and discussed below. Compared to other inverse problems, there are two specific difficulties in assimilation of meteorological observations. Firstly, the observations used are distributed in time, so that the complex nonlinear dynamics of the atmospheric flow, and in particular the instabilities that constantly develop in the flow, must be taken into account. Secondly, as shown by the numbers given above, the numerical size of the problems to be solved is extremely large. This difficulty is aggravated by the need for the forecast that will follow to be delivered in time. Those two specific difficulties have had a very strong impact on the development of assimilation of meteorological observations.

Oceanography is another domain where assimilation of observations has been used early. The dynamics of the ocean is very similar to that of the atmosphere, with different spatial and temporal scales. Due to its opacity to any form of electromagnetic radiation, the ocean is much more difficult to observe than the atmosphere. However, similar problems arise as in meteorology, and much of what is said below also applies to the ocean. Actually, assimilation of oceanographic observations is at the origin of several of the developments that are described in the following.

Section 1.2 presents a classical and relatively simple problem in Bayesian estimation in the linear and Gaussian case. The relatively simple algorithm called *optimal interpolation* (OI) is described in section 1.3 as an application of that elementary linear and Gaussian approach. The main algorithms that are used in operational assimilation, namely, *Variational Assimilation* and *Ensemble Kalman Filter* (EnKF), are presented and discussed as extensions of the same approach in section 1.4. A new class of algorithms, namely *Particle Filters*, is then described and discussed (section 1.5). The new methods of artificial intelligence (AI) and machine learning (ML) are briefly presented in section 1.6, and various extensions and applications of assimilation are described in section 1.7.

Early studies on assimilation of meteorological and oceanographical observations are as follows: Daley (1991); Ghil and Malanotte-Rizzoli (1991); Talagrand (1997);

Kalnay (2002). More recent studies are as follows: Asch et al. (2016) and Carrassi et al. (2018).

In the following, $E()$ denotes the expectation of a (scalar or vector) random variable. $\mathcal{N}(a, \mathbf{C})$ denotes the multidimensional Gaussian probability distribution with expectation a and covariance matrix $\mathbf{C}$ (with a similar transparent notation for a uni-dimensional Gaussian probability distribution).

1.2. The linear and Gaussian case

A simple estimation problem is as follows: we want to determine the unknown *state vector* x of the system of interest, belonging to *state space* $\mathcal{S}$, with dimension n. The known available data are supposed to make up a vector z, belonging to *data space* $\mathcal{D}$, with dimension m. The data vector is related to the unknown x through the relationship:

$$z = \mathbf{\Gamma} x + \zeta \qquad [1.1]$$

where the *data operator* $\mathbf{\Gamma}$ is a known linear operator from state space into data space, and ζ is an unknown "error", meant to represent the effect of all possible uncertainties in the link between x and z (instrumental and/or representativeness errors, inaccuracies in the specification of $\mathbf{\Gamma}$, etc.). Assume that the error ζ is a random Gaussian variable, with expectation 0 and covariance matrix $\mathbf{S}$ ($\zeta \sim \mathcal{N}(0, \mathbf{S})$) (if the expectation of ζ was not 0, but known, it would have to be first subtracted from z). The probability distribution of x, conditioned to z, is then the Gaussian distribution $\mathcal{N}(x^a, \mathbf{P}^a)$, with:

$$x^a = (\mathbf{\Gamma}^T \mathbf{S}^{-1} \mathbf{\Gamma})^{-1} \mathbf{\Gamma}^T \mathbf{S}^{-1} z \qquad [1.2]$$

$$\mathbf{P}^a = E[(x^a - x)(x^a - x)^T] = (\mathbf{\Gamma}^T \mathbf{S}^{-1} \mathbf{\Gamma})^{-1} \qquad [1.3]$$

These quantities are properly defined under the only condition that rank$\mathbf{\Gamma} = m$. This is a *determinacy* condition, which means that z contains information, either directly or indirectly, on every component of x. It requires $m \geq n$. We will set $p = m - n$. The upper index a in equations [1.2] and [1.3] refers to the fact that, in the standard vocabulary of meteorology, the estimate x^a is called the *analysis* of the state of the system. The expectation $E[(x^a - x)]$ of the error in x^a is 0. The estimate x^a minimizes the statistical quadratic estimation error on any component of the state vector x.

Equations [1.2] and [1.3] are invariant in any invertible linear change of coordinates in either state or data space. This means, for instance, that an observed temperature profile in the data can be transformed into a geopotential profile, through the linear hydrostatic equation, without affecting the final analysis. It also means that the analysis of the horizontal wind field (if spatially discretized) can be performed either in terms of vector coordinates or in terms of divergence and vorticity.

The invariance also means that the analysis is independent of the possible choice of a norm or scalar product in either state or data space (it is not necessary, for instance, to decide for performing the analysis if a 1 $m.s^{-1}$ – error on the wind is larger or smaller than a 1 K – error on the temperature).

In the Bayesian perspective considered in this chapter, we can note that a sample of independent realizations of the conditional probability distribution $\mathcal{N}(x^a, \mathbf{P}^a)$ can easily be obtained. Perturb the data vector z according to:

$$z \rightarrow z' \equiv z + \zeta' \qquad [1.4]$$

where ζ' is distributed according to the data error distribution $\mathcal{N}(0, \mathbf{S})$. It is easily seen that x'^a, defined by performing the analysis on the perturbed data z', viz:

$$x'^a = (\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{S}^{-1}z' \qquad [1.5]$$

is distributed according to $\mathcal{N}(x^a, \mathbf{P}^a)$.

If the determinacy condition rank$\mathbf{\Gamma} = m$ is verified, it is always possible to decompose the data vector z as follows:

$$x^b = x + \zeta^b \qquad [1.6]$$

$$y = \mathbf{H}x + \epsilon \qquad [1.7]$$

where x^b belongs to state space and has therefore dimension n, while y has dimension $p = m - n$. In addition, the decomposition can be performed in such a way that the "errors" ζ^b and ϵ are statistically uncorrelated, viz.

$$E(\zeta^b\epsilon^T) = 0 \qquad [1.8]$$

The vector x^b will be called a *background* estimate of the state vector x, while the vector y will be called an *observation* vector, belonging to *observation space*, and associated with a linear observation operator $\mathbf{H}$ acting from state into observation space. This vocabulary corresponds to a common situation in geophysics, where the available data consist of a prior estimate x^b (for instance, the output of a recent forecast) and of a set of recent observations y. However, the representation [1.6]–[1.8] is of course absolutely general, and is not restricted to the particular situation of meteorological prediction.

Defining $\mathbf{P}^b \equiv E(\zeta^b \zeta^{bT})$ and $\mathbf{R} \equiv E(\epsilon\epsilon^T)$, equations [1.2] and [1.3] transform under conditions ([1.6]–[1.8]) into either one of the two following equivalent sets of equations:

$$x^a = x^b + \mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}(y - \mathbf{H}x^b) \qquad [1.9]$$

$$\mathbf{P}^a = \mathbf{P}^b - \mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}^b \qquad [1.10]$$

$$x^a = x^b + \mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1}(y - \mathbf{H}x^b) \qquad [1.11]$$

$$[\mathbf{P}^a]^{-1} = [\mathbf{P}^b]^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \qquad [1.12]$$

In equations [1.9] and [1.11], the analysis x^a is expressed as the sum of the background x^b and a correction proportional to the vector $d \equiv y - \mathbf{H}x^b$. The latter is the difference between the observation y and the analogue $\mathbf{H}x^b$ of the observations in the background. It represents the new information contained in the observation with respect to the background. It is called the *innovation vector*. The matrix applied to the innovation, viz.,

$$\mathbf{K} \equiv \mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1} = \mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1} \qquad [1.13]$$

is called the *gain matrix*.

It is seen in either [1.10] or [1.12] that $\mathbf{P}^b < \mathbf{P}^a$. The difference represents the improvement in the knowledge of the state vector x brought, in addition to the background x^b, by the observation vector y.

1.2.1. *Variational form*

It can be seen that the analysis x^a (equation [1.2]) minimizes the following scalar *objective function*, defined on the state space $\mathcal{S}$:

$$\xi \in \mathcal{S} \to \mathscr{J}(\xi) \equiv \frac{1}{2}(\mathbf{\Gamma}\xi - z)^T\mathbf{S}^{-1}(\mathbf{\Gamma}\xi - z) \qquad [1.14]$$

$\mathscr{J}(\xi)$ measures the difference in the data space between the data vector z and what the data operator $\mathbf{\Gamma}$ would produce if it was applied on ξ. It is seen in addition that the

Hessian $\mathscr{J}'' \equiv \mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma}$ of $\mathscr{J}$ (matrix of the second derivative) is the inverse of the analysis error covariance matrix $\mathbf{P}^a$:

$$\mathscr{J}'' = [\mathbf{P}^a]^{-1} \quad [1.15]$$

In the representation [1.6]–[1.7], the objective function $\mathscr{J}(\xi)$ assumes the following form:

$$\mathscr{J}(\xi) = \frac{1}{2}(\xi - x^b)^T[\mathbf{P}^b]^{-1}(\xi - x^b) + \frac{1}{2}(\mathbf{H}\xi - y)^T\mathbf{R}^{-1}(\mathbf{H}\xi - y) \quad [1.16]$$

One way to compute the analysis x^a is therefore to minimize the objective function in [1.14] or [1.16].

Equations [1.2]–[1.3] or [1.14] (or equivalently equations [1.9]–[1.10], [1.11]–[1.12] or [1.16]) fully solve the problem of Bayesian estimation in the case of linear operators and additive Gaussian errors. They have actually a broader significance. Coming back to equation [1.1], let us assume that we know nothing about the error ζ beyond the fact that it is a random variable with expectation 0 and covariance matrix $\mathbf{S}$. We want to determine the linear combination $z \in \mathcal{D} \rightarrow \mathbf{A}z \in \mathcal{S}$ that minimizes the statistical quadratic error $E[(\mathbf{A}z - x)(\mathbf{A}z - x)^T]$. The answer is $\mathbf{A}z = (\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{S}^{-1}z$ as in equation [1.2], the associated minimum statistical quadratic error being $(\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma})^{-1}$ as in equation [1.3].

The estimate $x^a = \mathbf{A}z$ is called in this context the *Best Linear Unbiased Estimate (BLUE)* of x from z. It is defined, together with the associated estimation error, by the same algebraic equations ([1.2] and [1.3]) as the Bayesian estimate, but it now has no Bayesian significance. The *BLUE* will not necessarily be the conditional expectation of x for given z. Perhaps, more importantly, the matrix $\mathbf{P}^a = (\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma})^{-1}$ will not be the conditional covariance matrix of x for given z, but only the error covariance matrix over all possible realizations of z.

The fact that equations [1.2] and [1.3] solve the more general problem of the *BLUE* gives them additional usefulness for the practical solution of estimation problems. These equations are used, independently of a possible strict Bayesian significance, in many situations. Actually, most algorithms that have been used, or are now used, in assimilation of meteorological or oceanographical observations can be described as more or less heuristic and empirical extensions of equations [1.2] and [1.3] or [1.9] and [1.10] to moderately non-Gaussian and/or nonlinear situations. Many variants exist as to the choice of the operators $\mathbf{\Gamma}$ or $\mathbf{H}$ (which can be nonlinear), the associated covariance matrices $\mathbf{S}$, $\mathbf{P}^b$ or $\mathbf{R}$, and as to the numerical algorithms used for performing the required computations. Algorithms such that OI, *3D* and *4D Variational Assimilation*, *Kalman Filter* and *Kalman Smoother* and their various extensions, which will be further described, can all be described as fundamentally emanating from equations [1.2] and [1.3]. The only exception is *Particle Filters*, which are totally independent of any linear or Gaussian hypothesis, and will be described in section 1.5.

Concerning nonlinearity, it can easily be introduced, if it is not too strong, into the above equations. Assume, for instance, that the available data are in the following form, analogous to equations [1.6]–[1.7]:

$$x^b = x + \zeta^b \qquad [1.17]$$

$$y = H(x) + \epsilon \qquad [1.18]$$

H being now a nonlinear operator from state space into observation space. Let us assume that the difference $x - x^b$ is small enough so that the vector $d' \equiv y - H(x^b)$, which is analogous to the innovation vector d introduced above, can be approximated as:

$$d' = y - H(x^b) = H(x) - H(x^b) + \epsilon \approx \mathbf{H}'(x^b)(x - x^b) + \epsilon \qquad [1.19]$$

where $\mathbf{H}'(x^b)$ is the Jacobian matrix (matrix of partial derivatives) of the components of $H(x)$ with respect to the components of x, taken at point x^b. Equations [1.17] and [1.19] then define a linear estimation problem of form [1.6–1.7], with unknown $x - x^b$.

Similarly, if the data operator $\mathbf{\Gamma}$ in equation [1.1] is replaced by a nonlinear operator Γ, the latter can be introduced in place of $\mathbf{\Gamma}$ in the objective function (equation [1.14]). Minimization of the new objective function can be expected to lead to a useful estimate of the state vector x, even if the minimum has no precise Bayesian significance. This is largely confirmed by the long experience of variational assimilation, described in the following two sections.

1.3. Optimal interpolation – three-dimensional variational assimilation

In the context of numerical weather prediction, where new forecasts are regularly performed, while new observations are permanently acquired, a natural way to define the initial conditions of a new forecast is to combine the result of the latest forecast with the observations that have been acquired in the meantime. That is exactly what equations [1.9]–[1.10] (or equivalently [1.11]–[1.12]) can achieve, the background x^b being the latest forecast, and the vector y consisting of the new observations. One difficulty here is to define the covariance matrix $\mathbf{P}^b$ of the background error. That was done first by assuming that $\mathbf{P}^b$ is independent of the state of the flow and may depend only on the season of the year and/or the geographical area. This was implemented early in numerical weather prediction, under the name of OI. Before we describe how $\mathbf{P}^b$ was defined in that context in more detail, it is appropriate to make some comments on the significance of equation [1.9].

The vector:

$$\mu \equiv (\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}(y - \mathbf{H}x^b) \qquad [1.20]$$

in the right-hand side of equation [1.9], which has dimension $p = m - n$ and components (μ_j) $(j = 1, \ldots, p)$, depends only on the data vectors x^b and y, and is common to all components of the analyzed state vector x^a. It is multiplied in equation [1.9] by the $(n \times p)$-matrix $\mathbf{P}^b\mathbf{H}^T$, whose columns c_j are the covariance vectors of the p observations with the state vector x. Equation [1.9] can therefore equivalently be written as:

$$x^a = x^b + \sum_{j=1,\ldots,p} \mu_j c_j \qquad [1.21]$$

which shows that the modification made by the analysis on the background x^b is a linear combination of p vectors in state space, called the *representers*, which are individually associated with the observations. The representer c_j associated with the component y_j of the vector y defines how y_j influences, through the corresponding covariances, the variables in state space. It is the representers that propagate, in the state space, the information contained in the observations.

One way to numerically compute the analysis x^a is therefore to solve the p-dimensional linear system which defines the vector μ in equation [1.20], and then to perform the multiplications implied by equation [1.21].

In OI, the covariance matrix $\mathbf{P}^b$ has often been defined on the basis of statistical climatological information. One important aspect is the need for enforcing an approximate geostrophic balance between the mass and velocity increments in middle latitudes. Derber and Bouttier (1999) define an approach in which the increments on the geopotential, divergence and vorticity fields are sequentially balanced.

We may often wish to build covariance matrices from explicit analytical expressions. However, we must be careful that the obtained matrices are not only symmetric, but also positive definite. That is actually a very strong constraint. Considering, for instance, homogeneous functions of the form C[$|M_1M_2|$], where

$|M_1 M_2|$ is the distance (e.g. spherical distance) between points M_1 and M_2, the class of functions C which define proper mathematical covariances is restricted (in brief terms, they are the functions whose spectral transform is real and non-negative). This problem is discussed, in particular, by Gaspari and Cohn (1999), who give explicit analytical covariance functions appropriate for geophysical applications.

As concerns numerical implementation, one major step has been to determine the analysis state x^a by minimization of an objective function of form [1.16]. This is called *Three-Dimensional Variational Assimilation* (*3D-Var*) and is achieved through an iterative process, each step of which requires the explicit determination of the gradient (partial derivatives) of the objective function $\mathscr{J}(\xi)$ with respect to ξ. The latter is called in this context the *control variable*. In view of the large numerical dimension of the state vector, the gradient is determined through the *adjoint* approach to be discussed in more detail in section 1.4.2. The variational approach avoids the need for explicit determination of the gain matrix. It also allows us, as already mentioned, to use nonlinear operators $H(x)$, as for instance the operators associated with a number of satellite observations. This has turned out to be extremely useful and efficient.

3D-Var has been used, and is still used, in operational applications, as well as for research purposes. It was first developed at the ECMWF, where it was actually used as a prior step before developing the more general *Four-Dimensional Variational Assimilation* (*4D-Var*) to be described in the next section. It is to be noted that the expression 3D-Var has come to denote, in the established vocabulary of assimilation of meteorological observation, not any process that is variational and implemented in three dimensions at a given time, but more precisely minimization, at a given time, of an objective function of form [1.16] (with a possibly nonlinear observation operator $H(x)$), where the background error covariance matrix $\mathbf{P}^b$ is independent of the current state of the flow; in particular, of the instability processes that may be at play, or may have been recently at play, in the flow.

1.4. Taking the dynamics of the flow into account

Instability processes, which fundamentally result from temperature gradients (which result themselves from differential heating), are ubiquitous in the atmosphere. Everything else being the same, the uncertainty in the state of the flow at a given time is likely to be larger in those components of the flow that have been subject to recent instability, and it is certainly preferable to take the corresponding instability processes into account in the assimilation. Figure 1.1 shows, for a given meteorological situation, the temporal evolution of the error on the state of the flow over a simulated 24-h forecast. The figure has been obtained by repeated integration of a simple, but basically realistic, model of the atmospheric circulation. The quantity represented is the spatial correlation of the error between a given fixed point (45N, 35W) and surrounding points. It is seen that as the forecast range increases from 0 to 6 and 24 h (respective panels from (a) to (c)), the spatial extent of the correlation pattern increases, while its shape is significantly modified. In addition, the amplitude of the uncertainty increases as well as a result of the sensitivity of the forecast to the initial condition (not shown). The important fact here is that the evolution visible in Figure 1.1 will be different in another meteorological situation, and that it is therefore highly desirable to take into account the specific instability properties of the current situation in each new assimilation.

Kalman Filtering and Variational Assimilation, which we are going to describe now, are two broad classes of algorithms that are widely used in assimilation of meteorological and oceanographical observations. They allow taking into account, in a mathematically consistent way, the dynamical evolution of the flow together with the associated uncertainty. Although these algorithms significantly differ in their numerical implementation, they both derive from the basic estimate (equations [1.2] and [1.3]) and are equivalent in the linear case. It is only in nonlinear situations that they can lead to different results.

We explicitly introduce the temporal evolution of the system under observation. The state space $\mathcal{S}$, still with dimension n, is now the space in which the state vector, denoted x_k at time k, evolves in time. The evolution is assumed to be governed by the equation:

$$x_{k+1} = \mathbf{M}_k x_k + \eta_k \qquad [1.22]$$

where the linear operator $\mathbf{M}_k$ is a known *model operator*, which expresses our knowledge of the dynamics governing the evolution of the flow. The last term η_k is called the *model error*. It corresponds to the fact that the model operator $\mathbf{M}_k$ will not in general be exact, and that the real state x_{k+1} at time $k+1$ will not in general be obtained by applying that operator to the real state vector x_k at time k (assuming the latter to be known). The model error is the error in the model representation of the dynamics of the flow. It should be stressed that it is not in general the error in a forecast produced by the model. The error in a forecast results, not only from the model error, but also from the error in the initial conditions of the forecast.

The model error η_k will be assumed to be random with expectation 0, to be uncorrelated in time, and to have covariance matrix $\mathbf{Q}_k$ at time k. This is expressed by the equation:

$$E(\eta_k \eta_{k'}^T) = \delta_{kk'} \mathbf{Q}_k \qquad [1.23]$$

where $\delta_{kk'}$ is the Kronecker symbol ($\delta_{kk'} = 1$ if $k = k'$, and $\delta_{kk'} = 0$ if $k \neq k'$).

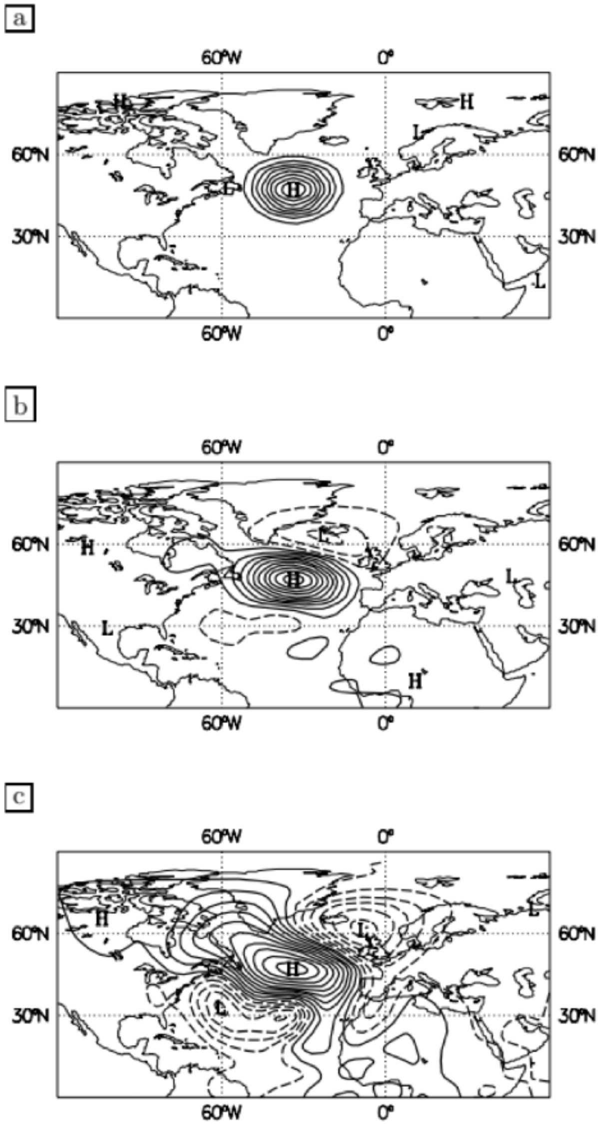

Figure 1.1. *Spatial correlation (between point 45N, 35W and surrounding points) of the error in a simulated numerical forecast at times 0, 6 and 24 h (from (a) to (c)). The field under consideration is the geopotential of the 500 hPa isobaric surface. The initial correlation (a) has been chosen as isotropic and everywhere positive. It is seen that, as the forecast proceeds, the correlation loses its isotropy while its spatial extension increases, and that negative values of correlation appear (dashed contours) (credit: F. Bouttier (Bouttier 1994))*

We assume that observations are available at successive times $k = 0, \ldots, K$, in the form:

$$y_k = \mathbf{H}_k x_k + \epsilon_k \quad [1.24]$$

where the observation vector y_k has dimension p, $\mathbf{H}_k$ is a known linear operator from the state space into the observational space, represented by a $p \times n$ matrix. ϵ_k is a random observational error with expectation 0, is uncorrelated in time, and has covariance matrix $\mathbf{R}_k$ at time k, viz.,

$$E(\epsilon_k \epsilon_{k'}^T) = \delta_{kk'} \mathbf{R}_k \quad [1.25]$$

(no explicit dependence of the dimension p of the observation vector with time is considered, but such a dependence could easily be introduced if necessary).

We assume that a background estimate x_0^b is available at the initial time $k = 0$, viz.,

$$x_0^b = x_0 + \zeta_0^b \quad [1.26]$$

with error covariance matrix:

$$E(\zeta_0^b \zeta_0^{bT}) = \mathbf{P}_0^b \quad [1.27]$$

Finally, we assume that the model errors η_k, the observation errors ϵ_k and the background error ζ_0^b are all mutually uncorrelated.

In these conditions, equations [1.22] to [1.27] define a linear estimation problem, which we will denote $\mathcal{P}$, of the general form (equation [1.1]), the unknown being now the consequence of system states $(x_0^T, \ldots, x_k^T, \ldots, x_K^T)^T$. The presence of the background (equations [1.26] and [1.27]), together with equation [1.22], ensures that the underlying determinacy condition is verified. The corresponding objective function (equation [1.16]) reads:

$$\begin{aligned}
&(\xi_0^T, \ldots, \xi_k^T, \ldots, \xi_K^T)^T \in \mathcal{S}^{K+1} \rightarrow \\
&\mathscr{J}_P(\xi_0, \ldots, \xi_k, \ldots, \xi_K) \equiv \frac{1}{2}(\xi_0 - x_0^b)^T [\mathbf{P}_0^b]^{-1} (\xi_0 - x_0^b) \\
&\qquad + \frac{1}{2} \sum_{k=0,\ldots,K} [y_k - \mathbf{H}_k \xi_k]^T \mathbf{R}_k^{-1} [y_k - \mathbf{H}_k \xi_k] \qquad [1.28] \\
&\qquad + \frac{1}{2} \sum_{k=0,\ldots,K-1} [\xi_{k+1} - \mathbf{M}_k \xi_k]^T \mathbf{Q}_k^{-1} [\xi_{k+1} - \mathbf{M}_k \xi_k]
\end{aligned}$$

where the term on the first line of the right-hand side measures the misfit of the unknown to the background x_0^b (equations [1.26] and [1.27]), the summation on the second line measures the misfit to the observations y_k (equations [1.24] and [1.25]) and the summation on the third line measures the misfit to the model equation (equations [1.22] and [1.23]).

1.4.1. *The Kalman Filter*

The Kalman Filter (Kalman 1960) is an algorithm for solving the estimation problem (equations [1.22] to [1.27]) recursively over time. It starts from the initial background x_0^b and the associated error covariance matrix $\mathbf{P}_0^b$. Assume a background estimate x_k^b has been obtained at a later time k from data (initial background, model equation and observations) at previous times, together with an error covariance matrix $\mathbf{P}_k^b$. The observation vector y_k is then used, through equations [1.9]–[1.10], to update these background estimates and to obtain a new, analyzed estimate x_k^a together with its error covariance matrix $\mathbf{P}_k^a$, viz.,

$$x_k^a = x_k^b + \mathbf{P}_k^b \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^b \mathbf{H}_k^T + \mathbf{R}_k)^{-1} (y_k - \mathbf{H}_k x_k^b) \quad [1.29]$$

$$\mathbf{P}_k^a = \mathbf{P}_k^b - \mathbf{P}_k^b \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^b \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \mathbf{H}_k \mathbf{P}_k^b \quad [1.30]$$

Following this *analysis step*, a *forecast step* is performed to time $k+1$, leading to:

$$x_{k+1}^b = \mathbf{M}_k x_k^a \quad [1.31]$$

$$\mathbf{P}_{k+1}^b = \mathbf{M}_k \mathbf{P}_k^a \mathbf{M}_k^T + \mathbf{Q}_k \quad [1.32]$$

The background x_{k+1}^b is obtained in equation [1.31] by applying the model equation (equation [1.22]) in its deterministic form to the analysis x_k^a. As for the corresponding estimation error covariance matrix $\mathbf{P}_{k+1}^b$, given by equation [1.32], it is obtained from equations [1.22] and [1.23], together with the hypothesis of temporal decorrelation of the errors. The term $\mathbf{M}_k \mathbf{P}_k^a \mathbf{M}_k^T$ on the right-hand-side corresponds to the error at time $k+1$ that results from the analysis error at time k, after it has been transported by the model $\mathbf{M}_k$. The term $\mathbf{Q}_k$ represents the effect of the model error η_k between times k and $k+1$ (equations [1.22] and [1.23]). Since the model error is assumed to be decorrelated from all previous errors, the corresponding covariance matrices simply add up.

The Kalman filter, starting from the initial conditions x_0^b and $\mathbf{P}_0^b$, thus consists of an alternative sequence of analysis and forecast steps of the respective forms [1.29]–[1.30] and [1.31[–[1.32]. At any time k, it solves the estimation problem $\mathcal{P}$ stated above. More precisely, if all errors in the data are globally Gaussian, the Gaussian probability distribution $\mathcal{N}(x_k^b, \mathbf{P}_k^b)$ [respectively $\mathcal{N}(x_k^a, \mathbf{P}_k^a)$] is the probability distribution for the real state x_k, conditioned by all data prior to (respectively up to) time k. In the case that the errors are not Gaussian, but the model and observation operators are still linear, a similar statement applies; the filter does not producing conditional probability distributions for the state x_k, but *BLUEs* for that state.

The idea of applying the Kalman Filter to assimilation of meteorological observations came at a very early stage, the first contribution being apparently by Jones (1965), with other early works by Petersen (1968); Talagrand and Miyakoda (1971) and Ghil et al. (1981).

Equation [1.32] achieves what has been identified above as a desirable aspect of assimilation, namely describing the temporal evolution the uncertainty on the state of the flow. Its cost is, however, prohibitive in any practical situation. Taken literally, it requires $2n$ multiplications by the matrix $\mathbf{M}_k$, that is, $2n$ integrations of the model between times k and $k+1$. That would be totally impossible in real time with a realistic model. Two approaches have been defined to circumvent that difficulty: *reduced order filters,* on the one hand, in which the uncertainty is assumed to be concentrated in a low-dimensional subspace of the state space, and *ensemble filters.* on the other hand, in which the uncertainty is represented by a finite number of points (model states) in state space.

1.4.1.1. *Reduced order filters*

There are several variants of reduced order filters. We succinctly describe here the *reduced rank square root Kalman filter* (RRSQRT), introduced by Verlaan and Heemink (1997, 2001). In that approach, the analysis is performed in the subspace spanned by $r \ll n$ eigenvectors of the background error covariance matrix $\mathbf{P}^b$ (we drop the time index k). More precisely, an $n \times n$ covariance matrix $\mathbf{P}$ can be written as:

$$\mathbf{P} = \mathbf{E}\mathbf{E}^T \qquad [1.33]$$

where the column vectors of the $(n \times n)$-matrix $\mathbf{E}$ are the (orthogonal) eigenvectors of $\mathbf{P}$ (the modulus of each vector being the square root of the associated positive eigenvalue). The RRSQRT consists of writing the matrix $\mathbf{P}^b$ in the form:

$$\mathbf{P}^b \approx \mathbf{E}^b\mathbf{E}^{bT} \qquad [1.34]$$

where $\mathbf{E}^b$ is an $(n\times r)$-matrix, the columns of which are the r dominant eigenvectors of $\mathbf{P}^b$ (or some appropriate approximation thereof). Setting $\mathbf{\Psi} \equiv (\mathbf{HE^b})^\mathbf{T}$, the analysis gain matrix (equation [1.13]) becomes:

$$\mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1} \approx \mathbf{E}^b\mathbf{\Psi}(\mathbf{\Psi}^T\mathbf{\Psi} + \mathbf{R})^{-1} \qquad [1.35]$$

while the analysis error covariance matrix $\mathbf{P}^a$ (equation [1.10]) becomes:

$$\mathbf{P}^a \approx \mathbf{E}^a \mathbf{E}^{aT} \qquad [1.36]$$

with:

$$\mathbf{E}^a = \mathbf{E}^b [\mathbf{I}_r - \mathbf{\Psi}(\mathbf{\Psi}^T \mathbf{\Psi} + \mathbf{R})^{-1} \mathbf{\Psi}^T]^{1/2} \qquad [1.37]$$

The matrix $\mathbf{\Psi}$ has dimension $p \times r$, so that, if the number p of observations is small, formulas [1.35] and [1.36]–[1.37] allow us to compute the (approximate) analysis x^a and the matrix $\mathbf{E}^a$ at a low cost.

In the forecast step, the r column vectors of $\mathbf{E}^a$ are evolved with the model (equation 1.22). A model error, if it is to be introduced, is not introduced in RRSQRT by an explicit perturbation term as in equation [1.22]. It can be introduced by decreasing the number of column vectors to be integrated, and replacing them by vectors meant to represent the dominant directions of the model error.

The main advantage of RRSQRT, provided that the number r of eigenvectors required to describe the dynamics (equation [1.22]) to a good degree of accuracy is small, is its low cost. Another advantage is the definition of the covariance matrices by their square root, as in equation [1.36]. This ensures the necessary positive definiteness of the covariance matrices, which can be lost through poor numerical conditioning in other algorithms. On the other hand, the dominance of the column vectors in $\mathbf{E}^b$ may be lost in the temporal integration of the model, thus degrading the accuracy of the filter. New, dominant eigenvectors may have to then be redefined.

Other forms of somewhat similar reduced order filters have been defined and studied, such as the *Singular Evolutive Extended Kalman Filter* (SEEK) (Pham et al. 1998) and its extension to nonlinear dynamics, the *Singular Evolutive Interpolated Kalman Filter* (SEIK) (Pham 2001).

It should be noted that these reduced order filters require the identification of directions of largest uncertainty in state space. That, contrary to what is the case for the Kalman Filter itself, which is norm-independent, requires the use of a norm in that space.

1.4.1.2. *The Ensemble Kalman Filter*

The EnKF (originally proposed for oceanographical applications by Evensen (1994), see also Evensen (2009)) is a Monte Carlo type method, in which the uncertainty on the state of the flow is at any time described by a finite ensemble of L

points in the state space (i.e. of state vectors). This ensemble is evolved through successive alternate analysis and forecast steps, similar to what is done in the standard Kalman Filter.

Let us assume that an ensemble of L background estimates $\{x_l^b\}$ $(l = 1, \ldots, L)$ is available at time k, at which an observation $y = \mathbf{H}x + \epsilon$ is available (we again drop the time index). The purpose is to update each background estimate x_l^b through a formula of form [1.9] in order to produce an analysis estimate x_l^a. To this end, an estimate of the background error covariance matrix $\mathbf{P}^b$ must be available. One possible choice is to take the sample covariance matrix of the ensemble $\{x_l^b\}$, viz.,

$$\mathbf{P}_{En}^b \equiv \frac{1}{L-1} \sum_{l=1,\ldots,L} (x_l^b - \bar{x}^b)(x_l^b - \bar{x}^b)^T \qquad [1.38]$$

where $\bar{x}^b$ is the average $\bar{x}^b \equiv \frac{1}{L} \sum_{l=1,\ldots,L} x_l^b$. In addition, the observation vector y is transformed through perturbations of the form:

$$y \rightarrow y_l = y - \epsilon_l \quad l = 1, \ldots, L \qquad [1.39]$$

into an ensemble of L approximate observations. The perturbations ϵ_l are independent realizations of the probability distribution of the observation error ϵ. The y_l's are meant to sample the uncertainty on the observation y. The couples $\{x_l^b, y_l\}$ are then combined together through the following formula, of form [1.9]:

$$x_l^a = x_l^b + \mathbf{P}_{En}^b \mathbf{H}^T (\mathbf{H}\mathbf{P}_{En}^b \mathbf{H}^T + \mathbf{R})^{-1} (y_l - \mathbf{H}x_l^b) \quad l = 1, \ldots, L \qquad [1.40]$$

This produces an ensemble of L analyzed states $\{x_l^a\}$, from which we can compute, for example, a mean analysis or an analysis error covariance matrix.

In the forecast step of the filter, the model (equation [1.22]) is integrated from the L analyzed states $\{x_l^a\}$. If a model error η_k is present, it can be added as a random perturbation in the course of each integration. And if, contrary to what is the case in equation [1.22], the model itself is nonlinear, the L integrations can also be performed with that model, thus eliminating the need for any linearity hypothesis in the dynamics of the system.

Equation [1.40] requires a linear observation operator $\mathbf{H}$ in the gain matrix. If the observation operator, denoted now H, is nonlinear, one solution is to replace it by its local Jacobian $\mathbf{H}'$ (see equation [1.19]). Another solution is to note that the matrices $\mathbf{P}^b\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}$ in equation [1.9] are actually the covariance matrices $E[(x - \bar{x}^b)d^T]$ and $E(dd^T)$. These can be approximated here as sample covariances from the ensembles $\{x_l^b - \bar{x}^b\}$ and $\{y_l - H(x_l^b)\}$ $(l = 1, \ldots, L)$ (if the latter solution is used, it avoids computation of the matrix $\mathbf{P}_{En}^b$ in equation [1.38]).

Experience shows that EnKF leads to very good results, even with ensembles of relatively small size, on the order of a few tens or at most 100–200, much smaller in any case than the dimension of the state space. Through the temporal integration of the ensemble, EnKF takes into account the temporal evolution of the uncertainty of the flow. This has a significant impact on its performance. As an example, Meng and Zhang (2008) have made a methodical comparison, over a long period, of an EnKF and a 3D-Var system, using the same model and the same observations. They have found a systematic advantage for the former. This advantage is due to the temporal integration of the ensemble.

Practical implementation of EnKF nevertheless raises a number of difficulties. The relatively small dimension of the ensembles, if it allows us to perform the required computations in the first place, has some disadvantages. It is observed that the covariances defined by the sample matrix $\mathbf{P}_{En}^b$ (equation [1.38]) show unrealistic large values at long spatial distances. This is due to undersampling by a relatively small ensemble. In addition, the matrix $\mathbf{P}_{En}^b$ has rank $L - 1$ at most, so that the increments defined by equation [1.40] are restricted from the start to a space of relatively small dimension, which does not allow us to properly sample the uncertainty on the analysis. Both these aspects are dealt with heuristically by an appropriate method of *localization*. The simplest localization method consists of modifying background covariances so that they smoothly decrease to zero at long distances. This can be achieved by replacing $\mathbf{P}_{En}^b$ with the matrix $\mathbf{P}_{Loc}^b$ defined by:

$$\mathbf{P}_{Loc}^b \equiv \mathbf{L} \circ \mathbf{P}_{En}^b \qquad [1.41]$$

In this expression, $\mathbf{L}$ is a symmetric isotropic definite positive matrix whose entries are close to unity at short distance and smoothly decrease with distance to become equal to zero beyond a distance d. The notation $\circ$ denotes Schur multiplication, or entry-wise multiplication of the two matrices. The matrix $\mathbf{P}_{Loc}^b$ thus defined is symmetric definite positive, which ensures that it is a proper covariance matrix, and achieves the intended purpose of small (and even zero) covariances at long distances. The choice of an appropriate matrix $\mathbf{L}$ can be based on the compactly supported covariance functions defined by Gaspari and Cohn (1999).

In addition, the localized matrix $\mathbf{P}_{Loc}^b$ is usually of full rank, which avoids the a priori concentration of the analyzed sample on a subspace that may be far from the observations.

Another difficulty with EnKF is that analyzed ensembles of small dimension tend to collapse and lose dispersion. Too small dispersion can lead to divergence of the filter, which will "trust" the background ensemble too much, and gradually diverge from the observations. The collapse is due, at least in part, to the nonlinear dependence of the gain matrix (equation [1.13]) with respect to the background error covariance matrix $\mathbf{P}^b$. This creates a systematic curvature bias, which decreases the dispersion of the ensembles. The usual procedure to avoid the collapse is to inflate the analyzed ensemble by an appropriate coefficient.

EnKF is widely used in many practical applications. Many variants have been defined. Most of these include both localization and inflation of ensembles, with coefficients largely determined through problem-dependent trial-and-error experimentation. One particular approach has been defined by Houtekamer and Mitchell (2001), who use two ensembles, the covariance matrix used for updating one ensemble being obtained from the other ensemble. A description of many of those variants can be found in Chapter 6 of Asch et al. (2016). Much work has been done also on the theoretical aspects of EnKF. For instance, Bocquet (2011) and Bocquet et al. (2015) have defined a filter which takes into account the fact that the background ensemble $\{x_l^b\}$, being a finite size sample, does not exactly represent the probability distribution of the background error. This approach must avoid the need for inflation.

In the Bayesian perspective taken here, it is interesting to know whether EnKF achieves Bayesianity. The standard Kalman Filter (equations [1.29]–[1.32]) achieves Bayesianity in the linear and Gaussian case under the hypothesis of decorrelation of errors in time. Under the same hypothesis, assuming the background ensemble $\{x_l^b\}$ to be a sample of independent realizations of the underlying conditional probability distribution $\mathcal{N}(x^b, \mathbf{P}^b)$, can a similar statement be made about the analysis $\{x_l^a\}$? If the background error covariance matrix used in the update (equation [1.40]) was equal to $\mathbf{P}^b$, the remark made about equations [1.4]–[1.5] would show that the answer is positive. Now, as just stressed above, the matrix $\mathbf{P}^b_{En}$ in equation [1.40] is not equal to $\mathbf{P}^b$ (nor would any other matrix defined from the finite ensemble $\{x_l^b\}$, such as $\mathbf{P}^b_{Loc}$ in equation [1.41]), so that the answer will in general be negative. However, we can note that the forecast step of EnKF, if started from a possible Bayesian analysis ensemble, would preserve Bayesianity of the ensuing background ensemble.

Le Gland et al. (2011) have studied the asymptotic properties of the EnKF when the dimension L of the ensemble grows to infinity. They have found that the distribution of the ensemble tends to an asymptotic limit, but that it is only in the linear case that the limit coincides with the Bayesian probability distribution.

The Kalman Filter, as described above, either in its standard form (equations [1.29]–[1.32]) or in its ensemble form, carries information forward in time only. It does not propagate the information backward in time (that is not necessary if assimilation is used only for defining initial conditions of forecasts, but can be useful for a posteriori assimilation of observations, as done in climate reanalyses). Kalman Filter is not capable to take into account temporal correlation between errors (indeed, temporal decorrelation is necessary for the optimality of the standard filter). Such temporal correlation certainly exists in reality. Concerning observation errors, observations performed at successive times by the same instrument (for instance, a satellite-borne radiometer) are certainly correlated. Also, as concerns model errors, they are certainly correlated, at least over short periods of time.

The algorithms that are capable of time-backward propagation of the information contained in the observations are called *smoothers*. One obvious approach for that is of course to consider a four-dimensional problem, using as data vector the set of all data (observations and model) that are available over the intended assimilation window, and taking as unknown the sequence of all model states over that period. That is what minimization of the objective function (equation [1.28]) would achieve in the linear case (in the absence of temporal correlation if we restrict ourself to the expression [1.28]). We will describe how variational assimilation, based on the minimization of an objective function integrated over time, is implemented in operational weather prediction in the next section. However, we mention that a large number of smoothers, some of which are based on a sequential approach similar to Kalman Filtering, have been developed and studied by many authors. All these algorithms propagate the information backward in time, while some are capable of taking temporal correlation into account. We refer the reader interested by a more complete description of smoothers to section 6.8 of Asch et al. (2016) as well as to Cosme (2015).

1.4.2. *Four-dimensional variational assimilation*

1.4.2.1. *Strong constraint variational assimilation*

As already said, equations [1.22]–[1.27] define a linear estimation problem of the general form (equation [1.1]). However, before considering it as such, we consider a different nonlinear problem, which was at the origin of four-dimensional variational assimilation. We consider the dynamical model to be nonlinear, but exact, so that equation [1.22] is changed into:

$$x_{k+1} = M_k(x_k) \qquad [1.42]$$

where $\mathbf{M}_k$ is now a nonlinear operator. A solution of equation [1.42] is entirely determined by the specification of x_0 at initial time 0. Similarly, we assume a nonlinear operator H_k in equations [1.24]–[1.25]. Except for that, the observation errors in equations [1.24]–[1.25] are still assumed to be uncorrelated in time, with error covariance matrix $\mathbf{R}_k$ at time k. A background estimate x_0^b is still assumed to be available at time $k = 0$, with error covariance matrix $\mathbf{P}_0^b$. We then define from equations [1.42], [1.24] and [1.25], the objective function:

$$\begin{aligned}\xi_0 \in \mathcal{S} \rightarrow \mathscr{J}_{\mathcal{S}}(\xi_0) \equiv \frac{1}{2}(\xi_0 - x_0^b)^T[\mathbf{P}_0^b]^{-1}(\xi_0 - x_0^b) \\ + \frac{1}{2}\sum_{k=0,\ldots,K}[y_k - H_k(\xi_k)]^T\mathbf{R}_k^{-1}[y_k - H_k(\xi_k)]\end{aligned} \quad [1.43]$$

subject to $\xi_{k+1} = M_k(\xi_k), \quad k = 0, \ldots, K-1$

For any $\xi_0 \in \mathcal{S}, \mathscr{J}(\xi_0)$ measures the misfit between the solution of equation [1.42] emanating from ξ_0, on the one hand, and the background x_0^b and the observations y_k, on the other hand.

Minimization of $\mathscr{J}(\xi_0)$ is called *strong constraint four-dimensional variational assimilation (*Strong Constraint 4D-Var). The words “Strong Constraint” mean that, in contrast with the “Weak Constraint” to be discussed below, the model equation [1.42] is supposed to be exact, with the consequence that the sequence of analyzed states exactly verify that equation.

The main problem with the numerical minimization of a scalar function such as $\mathscr{J}_{\mathcal{S}}(\xi_0)$ is the dimension of the control variable ξ_0. Minimization must be performed by an iterative gradient process, each step of which requires the explicit knowledge of the gradient of the function to be minimized with respect to the control variable. The cost of explicit computation of the gradient by finite perturbation of the components of the control variable, followed by finite difference, would be prohibitive. The most efficient way to compute the gradient is through the *adjoint* approach. The adjoint approach, which was first introduced in the context of assimilation of observations by Penenko and Obraztsov (1976), was later developed by Lewis and Derber (1985) and Le Dimet and Talagrand (1986). It is based on the following mathematical fact. Consider a mapping $u \rightarrow v = F(u)$ with input u and output v (with any dimensions), and a scalar function $\mathscr{J}(v)$ of the output v. The gradient of the compound function $\mathscr{J}_o F$ with respect to the input u, which we simply denote $\nabla_u \mathscr{J}$, is equal to:

$$\nabla_u \mathscr{J} = \mathbf{F}'^T \nabla_v \mathscr{J} \quad [1.44]$$

where $\nabla_v \mathscr{J}$ is the gradient of $\mathscr{J}$ with respect to the output v, and $\mathbf{F}'$ is the Jacobian of the mapping F, taken at point u. This allows us to proceed, through the transpose (or adjoint) $\mathbf{F}'^T$, from the gradient with respect to the output to the gradient with respect to the input. If the mapping F is a composition of the form $F_N \circ \ldots \circ F_1$, the transpose $\mathbf{F}'^T$ is the numerical product, taken in reverse order, of the transposes of the Jacobians of those components, viz.,

$$\mathbf{F}'^T = \mathbf{F}_1'^T \ldots \mathbf{F}_N'^T \qquad [1.45]$$

In the case of the objective function (equation [1.43]), the adjoint approach, once the input ξ_0 and the corresponding solution $\xi_k (k = 0, \ldots, K)$ of equation [1.42] are known, leads to the following sequence of operations for the computation of the gradient of $\mathscr{J}_S$ with respect to ξ_0:

– Compute at the final time K the vector:

$$\lambda_K = \mathbf{H}'^T_K \mathbf{R}_K^{-1} [H_K(\xi_K) - y_K] \qquad [1.46]$$

– Then, proceeding backwards in time, compute for $k = K - 1, \ldots, 1$:

$$\lambda_k = \mathbf{M}'^T_k \lambda_{k+1} + \mathbf{H}'^T_k \mathbf{R}_k^{-1} [H_k(\xi_k) - y_k] \qquad [1.47]$$

– Finally compute at time $k = 0$:

$$\lambda_0 = \mathbf{M}'^T_0 \lambda_1 + \mathbf{H}'^T_0 \mathbf{R}_0^{-1} [H_0(\xi_0) - y_0] + [\mathbf{P}_0^b]^{-1} (\xi_0 - x_0^b) \qquad [1.48]$$

In these equations, the matrices $\mathbf{M}'_k$ $(k = 0, \ldots, K - 1)$ and $\mathbf{H}'_k$ $(k = 0, \ldots, K)$ are the Jacobians of the respective operators M_k and H_k, taken at point ξ_k. The vectors λ_k $(k = 0, \ldots, K)$, which have dimension n, are called the *adjoint variables*. The vector λ_0 is the required gradient of the objective function (equation [1.43]) with respect to the initial condition $u \equiv \xi_0$, viz., $\lambda_0 = \nabla_u \mathscr{J}_S$.

In agreement with equation [1.45], the adjoint computations (equations [1.46]–[1.48]) are performed backwards in time, starting from the final time K. Their numerical cost is of the same order of magnitude as the cost of the direct integration of equation [1.42] (about twice as costly in the case of an atmospheric model, that ratio being basically independent of the dimension n of the state vector). It is also seen that they require the explicit knowledge of the current solution $\{\xi_k\}$ $(k = 0, \ldots, K)$ of equation [1.42]. The current solution is used as argument for the observation operators H_k, and also, in the case of nonlinear operators M_k and H_k, as arguments in the Jacobians $\mathbf{M}'_k$ and $\mathbf{H}'_k$. It must therefore be computed before the adjoint computations are started, and then kept in memory (or recomputed in the course of the adjoint computations). That high requirement of memory size is the price to be paid for the low computational cost of the adjoint approach.

As said, the objective function $\mathscr{J}_S$ is minimized through an iterative procedure. Each step of that procedure requires one direct integration of equation [1.42], started from the current value of the control variable ξ_0, followed by one backward integration of the adjoint equations (equations [1.46]–[1.48]). Although costly, that is much more economical than computation of the gradient by explicit perturbation of the control variable.

The impact of 4D-Var, as compared with 3D-Var, is very significant. Thépaut et al. (1993) compare the increments produced at the end of a 24-h assimilation by 3D-Var and 4D-Var. While the former shows the increment structure imposed by the a priori choice of the background error covariance matrix $\mathbf{P}^b$ (equation [1.16]), the latter shows a much more complicated structure, which results from the evolution of the flow over the 24 h of assimilation, and in particular of the instabilities that have occurred over that period.

It is the adjoint approach that has made variational assimilation possible in the first place. Strong constraint variational assimilation using the adjoint approach was first introduced in operational numerical weather prediction by ECMWF in 1997 with an assimilation window of 6 h. It led to significant improvements in the quality of the ensuing forecasts in comparison with the previous 3D-Var (see, for example, Klinker 2000; Rabier 2000).

Contrary to Kalman Filtering, variational assimilation does not compute an explicit estimate of the uncertainty on the state of the flow, either in the final analysis or at any other step of the process. However, the Jacobian $\mathbf{M}'_k$ describes to first order the evolution of the difference between two close solutions of the model equation [1.42]. The evolution of the uncertainty on the state of the flow is thus represented, not explicitly, but in a way that allows us to minimize the objective function $\mathscr{J}_S$ (that would produce in the linear case, the same estimate as Kalman Filter at the final time K). Another basic difference is that variational assimilation, in contrast to standard sequential Kalman filter, is smoother in that it adjusts the assimilating model to the whole set of observations available over the assimilation window, and propagates the information contained in the observations both forward and backward in time. This is not important when assimilation is used for defining the initial conditions of a forecast, but can be important for other uses, for instance, for reanalysis of past observations where it is preferable to use all available observations, performed either before or after analysis time.

As for the Kalman filter, approximations have been made for reducing the cost of the computations. These approximations essentially consist of simplifying the assimilating model (equation [1.42]) together with its adjoint (exact consistency between the assimilating model and its adjoint is necessary for the success of the minimization). This is usually achieved by decreasing the resolution of the model, which means that the minimization is performed only in the large scales of the flow, the smaller scales not being modified. In the *incremental approach* described by Courtier et al. (1994), the approximations that are made vary in the successive iterations of the minimization process. The whole procedure is defined in such a way that, owing to appropriate successive linear approximations, the global minimization consists of a sequence of minimizations of more easily minimized exactly quadratic objective functions.

On the other hand, Variational Assimilation, as described above, does not allow cycling for one assimilation window to the next. As described above, it has to be restarted from scratch for each new window. Purely sequential assimilation such as the Kalman Filter, on the contrary, allows continuous introduction of new observations. However, the method of *Ensemble of Data Assimilations*, described in section 1.4.3, allows for carrying partial information from one window of Variational Assimilation to the next.

Strong constraint Variational Assimilation and EnKF are the two main algorithms that are used for assimilation in operational large-dimension numerical weather prediction. No significant difference seems to have been observed between the results they produce. A clean comparison experiment (same model, same observations, same assumed errors) was made by Buehner et al. (2010a, 2010b) who found that, for the same total computational cost, both approaches led to results which, although different in some aspects, were of similar global quality.

1.4.2.2. *Weak Constraint Variational Assimilation*

Weak Constraint Variational Assimilation takes into account the presence of errors in the numerical model, as in equation [1.22]. It is seen from equations [1.28] and [1.43] that, in the case of nonlinear model or observation operators, the corresponding objective function to be minimized reads:

$$
\begin{aligned}
&(\xi_0^T, \ldots, \xi_k^T, \ldots, \xi_K^T)^T \in \mathcal{S}^{K+1} \rightarrow \\
&\mathscr{J}_w(\xi_0, \ldots, \xi_k, \ldots, \xi_K) \equiv \frac{1}{2}(\xi_0 - x_0^b)^T [\mathbf{P}_0^b]^{-1} (\xi_0 - x_0^b) \\
&\qquad + \frac{1}{2} \sum_{k=0,\ldots,K} [y_k - H_k(\xi_k)]^T \mathbf{R}_k^{-1} [y_k - H_k(\xi_k)] \\
&\qquad + \frac{1}{2} \sum_{k=0,\ldots,K-1} [\xi_{k+1} - M_k(\xi_k)]^T \mathbf{Q}_k^{-1} \\
&\qquad \times [\xi_{k+1} - M_k(\xi_k)]
\end{aligned}
\qquad [1.49]
$$

which differs from equation [1.43] in that the control variable now consists of the whole sequence of model states $\{\xi_k\}_{k=0,\ldots,K}$, and that additional terms, measuring the misfit of that sequence to the dynamical equation [1.42], are present. Equation [1.49] differs from equation [1.28] in that it allows for nonlinear model and observation operators.

Weak Constraint Variational Assimilation has been abundantly studied. One major difficulty is that it requires an explicit specification for the model error covariance matrix $\mathbf{Q}_k$. It is very difficult to quantitatively assess the model error, which can vary significantly with the variables under consideration and the current situation of the flow (not to mention the fact that it can also be correlated in time). On the other hand, Weak Constraint Assimilation has a distinct numerical advantage. The model error term regularizes the objective function $\mathscr{J}_w(\xi_0, \ldots, \xi_k, \ldots, \xi_K)$ (as any smoothing term actually does) and makes the minimization easier. In spite of that, it does not seem that Weak Constraint Variational Assimilation could easily be implemented in large-scale operational applications, but positive results have been obtained. The operational assimilation system of ECMWF now includes a weak constraint component in the assimilation of stratospheric fields.

1.4.3. *Ensemble methods*

EnKF and Variational Assimilation have been widely used in operational systems for assimilation, with constant improvements but no fundamental changes, for two decades. Research for more powerful methods nevertheless continues, as we are going to describe now.

As said before, it is natural, in the general Bayesian logic that we have described for assimilation, to develop methods which produce an ensemble of state points meant to sample the probability distribution of the state of the observed system, conditioned by the available data. EnKF is of course one such method. However, other ensemble methods have been defined and studied.

It has been mentioned (equations [1.4]–[1.5]) that, in the linear and Gaussian case, an ensemble of independent realizations of the conditional probability distribution $P(x|z) = \mathcal{N}(x^a, \mathbf{P}^a)$ can easily be obtained. Perturb the data z following the probability distribution of the data error, and perform the analysis (equation [1.2]) on the perturbed data, for instance, by minimizing an objective function of form [1.14].

This approach can of course be implemented with nonlinear model and observation operators, even if there is no guarantee as to the Bayesianity of the ensembles that will be obtained. It has been implemented under the name of

Ensemble of Data Assimilations (EDA) at ECMWF and Météo-France. It has not been implemented there for the full assimilation, but only using a simplified model, in order to obtain a sample of analyzed states that are used at the end of the assimilation window, together with a static estimate of the background error covariance matrix, for defining the matrix $\mathbf{P}_0^b$ to be used in the objective function $\mathscr{J}_S(\xi_0)$ (equation [1.43]). Concerning the Bayesianity of the ensembles produced by EDA in nonlinear situations, Jardak and Talagrand (2018) have performed experiments on nonlinear systems, in particular on the chaotic system defined by Lorenz (1996). They have found that, as long as can be judged from statistics performed on the obtained ensembles, EDA produces, for the same ensemble size, ensembles that are as good an estimate of the real state and of the associated uncertainty as in the linear and Gaussian case, in which EDA is exactly Bayesian.

A number of ensemble algorithms, which may borrow from both the variational and the Kalman filter approach, have been defined, many of which are described and discussed in Chapter 7 of Asch et al. (2016). Variational assimilation requires the use of the adjoint of the assimilating model. Development, validation and maintenance of an adjoint code, although it is a rather straightforward task, require time and resources. The possibility of avoiding that task, or at least of significantly reducing its cost, has led to the development of new algorithms. Liu et al. (2008, 2009), following ideas initially put forth by Robert et al. (2005), instead of performing the minimization of an objective function in the entire state space, performed it in a relatively small subspace. The idea is that, if the dimension of the minimizing space is small enough, the gradient of the objective function with respect to the control variable can be computed by finite perturbation of the latter, thus avoiding the need for the adjoint of the dynamical model (or of possibly nonlinear observation operators). The gradient can then be computed through the transpose of a small dimension explicitly computed Jacobian matrix. This approach, of which numerous variants exist, is commonly called *4DEnVar*. Liu et al. (2008, 2009) minimized a quadratic approximation of the objective function, which leads to small-dimension matrix computations. The question of the definition of the initial background error covariance matrix remains, and the new question of how to choose the small dimension minimization space arises. A number of possible solutions, including for instance running an EnKF in parallel with the variational assimilation, have been studied.

Along the same idea of restricting the minimization to an appropriately defined subspace, Bocquet and Sakov (2014) have introduced the *Iterative Ensemble Kalman Smoother* (IEnKS). The gradient of the objective function with respect to the control variable is again computed by finite difference in a small dimension control space, a finite difference approximation of the Hessian of the objective function being determined at the same time. Once the minimizing solution has been determined, the control subspace is regenerated, on the basis of a Gaussian hypothesis, from the approximate Hessian. The IEnKS has been shown, on nonlinear systems of relative small dimensions, to outperform, in terms of final accuracy, both Strong Constraint Variational Assimilation and the EnKF.

1.4.4. *Stability and instability*

Experience shows that various algorithms that have been used for ensemble assimilation produce useful results for relatively small ensemble sizes, of the order of a few tens or at most a few hundreds, while they are used with numerical models that have state dimensions on the order of 10^{3-5} or more. This is fortunate in that it is largely what makes them useful, but it is also somewhat surprising that these relatively small dimensions can lead to useful results. Also, it raises the question of what a good choice of ensemble size is.

Consider for simplicity the case of a perfect model. Over the assimilation window, the error along the unstable components of the flow will increase, while the error along the stable modes will decrease. As a result, at the end of a variational assimilation (assuming that the observations are uniformly distributed over time), the residual error in the minimizing solution will be concentrated in the stable modes at the beginning of the assimilation window, and in the unstable modes at the end of the window. This aspect was studied by Pires et al. (1996), who found in numerical experiments that this was indeed the case. More recently, Gurumoorthy et al. (2017) and Bocquet et al. (2017) have rigorously proven that, in the case of an exact linear model, the error in the Kalman Filter concentrates asymptotically in the unstable manifold of the model. A consequence is that, along an assimilation window, late observations bring information on the unstable modes, which have developed over the window, and are therefore easier to observe. Early observations, on the other hand, bring information on the stable modes of the flow. This can be particularly important for a posteriori assimilation. Observations performed at the beginning and at the end of the assimilation window do not contain the same type of information and cannot be considered as interchangeable. Having observations distributed in time is not the same as having the same quantity of observations, with the same accuracy, concentrated at one time. The former says more on the stable or unstable character of the flow and of the components of the flow along the corresponding modes.

It results that an important aspect of assimilation is to monitor in one way or another and to prevent as much as can be done, the development of instabilities in the course of assimilation. Also, in the case of variational assimilation as well as of any form of smoother, it is important to monitor and control the stable components of the flow. Trevisan et al. (2010) have shown that concentrating the control variable of Variational Assimilation in the unstable subspace of the flow (identified by an ensemble integration of the model itself, which concentrates the dispersion of the forecasts in that subspace) can be more efficient than keeping the control variable in the entire state space.

Of course, in a linear case, Kalman Filter in its standard form, if implemented correctly, will explicitly describe the evolution of the uncertainty through equation [1.32]. In a nonlinear case, if a realistic evolution of the uncertainty is included in the assimilation (for instance, in an Ensemble Filter), the growth of instability should be monitored and taken into account in the analysis step of the filter.

However, the important fact is that in geophysical flows, the number of unstable components is relatively small. In large-scale meteorology and oceanography, the dominant instabilities result from the latitudinal temperature gradient. At any time, there are only a small number (typically a few tens) of "centers of action" where instabilities develop. Moreover, the physical process of instability, the so-called *baroclinic instability*, is fundamentally the same in those various centers of action. This is basically the reason that makes it possible to implement ensemble assimilation, in either meteorological or oceanographical applications, with relatively small ensembles.

All simplified forms of Kalman Filter or Variational Assimilation that have been presented above are based on the restriction of the estimation to a small dimension subspace of the entire state space. No systematic study of how that subspace must be chosen (in particular, in terms of the current situation of the flow) seems to have been performed. However, it is clear from the preceding discussion that it must include at least the directions along which instabilities have developed in the recent past.

1.5. Particle filters

All the algorithms that have been described up to now are fundamentally based on the linear and Gaussian approach described in section 1.2. This approach has been heuristically (and successfully) extended to moderately nonlinear and non-Gaussian situations. However, the basic equations that are used in those algorithms are fundamentally of the form [1.9]–[1.10] or [1.14], and all algorithms that have been used up to now in large-scale meteorological or oceanographical applications are of the same general form.

Particle filters are totally independent of any linear or Gaussian hypothesis (see, for example, Doucet et al. (2001) for an introduction). As their name suggests, they are ensemble methods in which the state of the system is described at any time, and as in EnKF, by a set of points in state space. However, a probability is now attached to each of these points, and it is both the particles and their associated probabilities that evolve in the course of the assimilation.

We give here a simple description of particle filters, without going into full mathematical details. Let us assume that a background ensemble $\{x_l\}$ $(l = 1, \ldots, L)$ is available at an observation time (we drop again the time index), together with associated probabilities, which we denote $\{p_l^b \equiv P(x_l)\}$ $(l = 1, \ldots, L)$. Let y be the observation vector at the same time. In order to update the probabilities with y and to obtain new probabilities p_l^a, we use Bayes' law, viz.,

$$p_l^a \equiv P(x_l|y) \propto P(y|x_l)P(x_l) \qquad [1.50]$$

where the proportionality coefficient, if required, can be obtained from the condition that $\sum_l p_l^a = 1$. The conditional probability $P(y|x_l)$ (the so-called *likelihood*) can easily be obtained if, for instance, y is of form [1.18], the probability distribution of the error ϵ being independent of the probability distribution of the x_l's. Then $P(y|x_l) = P[\epsilon = y - H(x_l)]$. It is seen that, in this analysis step, the particles x_l are not modified, but their probabilities are.

In the forecast phase to the next observation time, the particles x_l are evolved according to the model equation [1.42] (or through any process that is consistent with the probability distribution defined by the weights p_l^a; see, for example, Carrassi et al. (2018)). If a random model error is present, an independent realization of that error has to be included in each of the L model integrations. This is easy to achieve if the model error is supposed to be independent of the distribution of the x_l's. In the forecast phase, the particles x_l are modified, but their probabilities are not.

It is clear that this procedure is independent of any linear or Gaussian hypothesis. As described, it depends, as do the standard Kalman Filter and most of the variants of the latter, on a hypothesis of independence in time of the errors (in precise mathematical terms, the appropriate condition is a condition of *Markovianity*). It is clearly Bayesian, at least in the sense that, if it is initialized from an ensemble of independent realizations of an initial probability distribution, it will produce at any later stage a sample of independent realizations of the probability distribution of the system state at that later stage, conditioned by the information (observations and model) that has been used in the meantime.

Particle filters are very efficient in small dimension systems, but when the dimension increases, it is observed that the ensembles tend to collapse, all the probability tending to concentrate on one particle. This is a manifestation of the *curse of dimensionality*, which is the fact that it is in practice very costly, if not simply impossible, to describe probability distributions in large dimensional spaces. Snyder et al. (2008) estimate that, in order to avoid ensemble collapse, the size of the ensemble must grow exponentially with the dimension of the state space. Active research is being pursued for Particle Filter schemes that might circumvent the curse of dimensionality. A general approach, called *regularization*, is to remove from the ensemble particles with low probability, and to replace them by appropriately chosen new particles in the region of phase space with high probability. For a review of particle filters as applied to high-dimensional systems, see, for example, van Leeuwen (2017).

1.6. Artificial intelligence

The continuing growth of the quantity of available data and of computing power has stimulated in many fields the development of AI and ML. These are methods that can produce knowledge through pure statistical processing of large amounts of data, without use of (or in addition to) a priori explicitly stated physical laws (or, for that matter, biological or economical laws, or any explicit statement relative to the behavior of the system of interest). ML is largely based on the use of *neural networks*, which establish explicit numerical links between given sets of "inputs" and "outputs". Atmospheric science requires in all its aspects large amounts of data and powerful computing, and is therefore a potential field of applications of these new developments. Data assimilation and ML, even if they are at present distinctly different in the way they are implemented and in the algorithms they use, are basically two approaches for achieving the same purpose, viz., extract as useful information as possible from large sets of data. In the context of atmospheric science, and in relation with data assimilation, ML is being considered for two classes of applications (which are not mutually exclusive): (1) identification of laws governing the evolution of the observed system (see, for example, Brajard et al. (2020)); and (2) identification of observation operators. Observation operators have not been discussed much above, but they relate the observations to the state variables of the assimilating model (equation [1.18]). They are often meant to represent complex, poorly known, physical processes, sometimes through their statistical impact at a spatial or temporal scale that is larger than the scale at which the observations are performed. An example is the interaction between clouds and radiation, which is poorly known even at the microscale at which it takes place, and has to be represented in assimilation at a much larger scale. We can hope that the new possibilities offered by AI and ML will lead to significant progress in, at least, the identification of observation operators. For more on the use of ML in the context of data assimilation, see, for example, Geer (2021).

1.7. Extensions and applications

When numerical weather prediction was first developed, in the early 1950s, what mattered was of course the forecast itself. The definition of the initial conditions, performed originally as a simple interpolation from observing stations to model gridpoints, was a very minor task. But then difficulties gradually arose, which led to including more and more components in the process of definition of the initial conditions: need for ensuring realistic spatial scales, together with approximate geostrophic balance, in the initial fields, introduction of observations that are distributed in time, with the need for using the forecast model even before the start of the forecast (it is on that occasion that the word *assimilation* was coined), introduction of observations that are linked to the model through complex, sometimes poorly known, relationships (for instance satellite observations), and need for monitoring and controlling the instabilities in the system. This resulted in major technical and numerical developments, and assimilation has now become a very substantial part of the whole process of numerical weather prediction. One day of assimilation now requires the same amount of computing resources as for five days of prediction or more (including ensemble prediction). There is little doubt that the gradual development of more and more powerful methods for assimilation has significantly contributed, together with the improvement of numerical models and of the observing system, to the improvement of the quality of the meteorological forecasts.

However, even in the world of meteorological prediction, assimilation has extended much beyond the simple definition of the initial conditions of a forecast. It is a basic diagnostic in the assessment of the quality of a model. The degree to which a model is capable to fit the observations through assimilation is one measure, among others, of the quality of the model. Removing a particular set of observations from an assimilation experiment gives a direct measure of the information content of that set of observations. Assimilation is systematically used for *Observing Systems Simulation Experiments* (OSSEs), intended at evaluating a priori the gain that possible future observing systems could bring on the knowledge of the state of the atmosphere and on the quality of the ensuing forecasts. Assimilation is also at the core of *reanalyses*, in which past observations are assimilated with present models and algorithms (Kalnay et al. 1996). Such reanalysis programs now process observations going back as far as the beginning of the 20th century (Laloyaux et al. 2018). Reanalyses are very useful, even fundamental, for climatic studies, especially now in the context of an evolving climate. However, they are also used for making reforecasts on past situations, which give additional and interesting insight on the performance and quality of present models.

Beyond atmospheric science, assimilation has naturally extended to many other aspects of geophysical fluid dynamics. Oceanography has already been mentioned. Not only do there exist programs for assimilation of observations of the ocean, but interactive assimilation of both meteorological and oceanographic observations in coupled models is of great potential interest, for instance, for seasonal prediction, and is the subject of active research. The difficulty here is the gap between typical time scales of the atmospheric and oceanic dynamics, which does not allow direct use of algorithms such as the ones that are described above. A discussion of the associated problems can be found in Penny and Hamill (2017).

Assimilation has now been used in almost all branches of climate science. We can mention atmospheric chemistry, hydrology, paleoclimatology, glaciology, the dynamics of the biosphere and the vegetation cover. It has become extraterrestrial with the study of the dynamics of planetary atmospheres. Outside strictly climatic science, it has been used to the study of seismology, plate tectonics and magnetism (not only terrestrial, but also solar), volcanology and to applications which actually go beyond physical and chemical sciences. Chapters 11 and 12 of Asch et al. (2016) give examples of applications of assimilation of observations in physical and chemical sciences, as well as in human and social sciences. Also, for a recent example, we can mention a work by Evensen et al. (2021) on the dynamics of the COVID-19 epidemic.

1.8. References

Asch, M., Bocquet, M., Nodet, M. (2016). *Data Assimilation: Methods, Algorithms, and Applications*. SIAM, Philadelphia, PA.

Bocquet, M. (2011). Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Proc. Geoph.*, 18, 735–750.

Bocquet, M. and Sakov, P. (2014). An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 140, 1521–1535.

Bocquet, M., Raanes, P.N., Hannart, A. (2015). Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation. *Nonlinear Proc. Geoph.*, 22, 645–662.

Bocquet, M., Gurumoorthy, K.S., Apte, A., Carrassi, A., Grudzien, C., Jones, C.K. (2017). Degenerate Kalman filter error covariances and their convergence onto the unstable subspace. *SIAM/ASA Journal on Uncertainty Quantification*, 5, 304–333.

Bouttier, F. (1994). Sur la prévision de la qualité des prévisions météorologiques. PhD Thesis, Université Paul-Sabatier, Toulouse.

Brajard, J., Carassi, A., Bocquet, M., Bertino, L. (2020). Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model. *Journal of Computational Science*, 44, 101171.

Buehner, M., Houtekamer, P.L., Charette, C., Mitchell, H.L., He, B. (2010a). Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Monthly Weather Review*, 138, 1567–1586.

Buehner, M., Houtekamer, P.L., Charette, C., Mitchell, H.L., He, B. (2010b). Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Monthly Weather Review*, 138, 1550–1566.

Carrassi, A., Bocquet, M., Bertino, L., Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Climate Change*, 9(5), e535.

Cosme, E. (2015). Méthodes d'assimilation rétrospective de données et application à l'océanographie. HDR Thesis, Université Grenoble Alpes, Grenoble.

Courtier, P., Thépaut, J.-N., Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.*, 120, 1367–1387.

Daley, R. (1991). *Atmospheric Data Analysis*. Cambridge University Press, Cambridge.

Derber, J. and Bouttier, F. (1999). A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, 51A, 195–222.

Doucet, A., de Freitas, N., Gordon, N. (eds) (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo Methods in Practice*. Springer, Berlin.

Evensen, G. (1994). Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99(C5), 10143–10162.

Evensen, G. (2009). *The Ensemble Kalman Filter*. Springer, Berlin.

Evensen, G., Amezcua, J., Bocquet, M., Carrassi, A., Farchi, A., Fowler, A., Houtekamer, P.L., Jones, C.K., de Moraes, R.J., Pulido, M. et al. (2021). An international initiative of predicting the SARS-CoV-2 pandemic using ensemble data assimilation. *Foundations of Data Science*, 3(3), 413–477.

Gaspari, G. and Cohn, S.E. (1999). Construction of correlation functions in two and three dimensions. *Q. J. Roy. Meteor. Soc.*, 125(554), 723–757.

Geer, A. (2021). Learning earth system models from observations: Machine learning or data assimilation? *Phil. Trans. R. Soc. A.*, 379(2194). doi: 10.1098/rsta.2020.0089.

Ghil, M. and Malanotte-Rizzoli, P. (1991). Data assimilation in meteorology and oceanography. *Adv. in Geophys.*, 33, 141–266.

Ghil, M., Cohn, S., Tavantzis, J., Bube, K., Isaacson, E. (1981). Applications of estimation theory to numerical weather prediction. In *Dynamic Meteorology-Data Assimilation Methods*, Bengtsson, L., Ghil, M., Källén, E. (eds). Springer, Berlin.

Gurumoorthy, K.S., Grudzien, C., Apte, A., Carrassi, A., Jones, C.K. (2017). Rank deficiency of Kalman error covariance matrices in linear time-varying system with deterministic evolution. *SIAM J. Control Optim.*, 55, 741–759.

Houtekamer, P. and Mitchell, H.L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.*, 129, 123–137.

Jardak, M. and Talagrand, O. (2018). Ensemble variational assimilation as a probabilistic estimator – Part 2: The fully non-linear case. *Nonlin. Processes Geophys.*, 25, 589–604. doi: 10.5194/npg-25-589-2018.

Jones, R.H. (1965). Optimal estimation of initial conditions for numerical prediction. *J. Atmos. Sci.*, 22, 658–663.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *J. Fluid. Eng.*, 82, 35–45.

Kalnay, E. (2002). *Atmospheric Modeling and Data Assimilation and Predictability*. Cambridge University Press, Cambridge.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J. et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, 77, 437–471.

Klinker, E., Rabier, F., Kelly, G., Mahfouf, J.-F. (2000). The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration. *Q. J. R. Meteorol. Soc.*, 126, 1191–1215.

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J.-R., Broennimann, S., Buizza, R., Dalhgren, P., Dee, D., Haimberger, L., Hersbach, H. et al. (2018). Cera-20c: A coupled reanalysis of the twentieth century. *Journal of Advances in Modeling Earth Systems*, 10(1172), 97–110. doi: 10.1029/2018MS001273.

Le Dimet, F.X. and Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, 38A, 97–110.

Le Gland, F., Monbet, V., Tran, V.-D. (2011). Large sample asymptotics for the ensemble Kalman filter. In *The Oxford Handbook of Nonlinear Filtering*, Crisan, D. and Rozovskii, B. (eds). Oxford University Press, Oxford.

van Leeuwen, P. (2017). Particle filters for nonlinear data assimilation in high-dimensional systems. *Annales de la faculté des sciences de Toulouse : Mathématiques*, 6(26), 1051–1085. doi: 10.1007/BF02427924.

Lewis, J. and Derber, J.C. (1985). The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, 37A, 309–322.

Liu, C., Xiao, Q., Wang, B. (2008). An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Weather Rev.* 136, 3363–3373.

Liu, C., Xiao, Q., Wang, B. (2009). An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing system simulation experiments with advanced research WRF (ARW). *Mon. Weather Rev.*, 137, 1687–1704.

Lorenz, E.N. (1996). Predictability: A problem partly solved. In *Seminar on Predictability*. ECMWF Reading, Berkshire.

Meng, Z. and Zhang, F. (2008). Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVAR in a month-long experiment. *Mon. Weather Rev.*, 136, 3671–3682.

Penenko, V. and Obraztsov, N.N. (1976). A variational initialization method for the fields of the meteorological elements. *Soviet Meteorol. Hydrol.*, 11, 1–11.

Penny, S. and Hamill, T.M. (2017). Coupled data assimilation for integrated earth system analysis and prediction. *Bull. Amer. Meteor. Soc.*, 98, ES169–E172.

Petersen, D. (1968). On the concept and implementation of sequential analysis for linear random fields. *Tellus* 20, 673–686.

Pham, D. (2001). Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Weather Rev.*, 129, 1194–1207 [Online]. Available at: https://journals.ametsoc.org/view/journals/mwre/129/5/1520-0493_2001_129_1194_smfsda_2.0.co_2.xml.

Pham, D., Verron, J., Roubaud, M.C. (1998). Singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Marine Systems*, 16, 323–340. doi: 10.1016/S0924-7963(97)00109-7.

Pires, C., Vautard, R., Talagrand, O. (1996). On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus*, 48A, 96–121.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., Simmons, A. (2000). The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, 126, 1143–1170.

Robert, C., Durbiano, S., Blayo, E., Verron, J., Blum, J., Le Dimet, F.X. (2005). A reduced-order strategy for 4D-VAR data assimilation. *J. Marine Systems*, 57, 70–82.

Snyder, C., Bengtsson, T., Bickel, P., Anderson, J.L. (2008). Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.*, 136, 4629–4640.

Talagrand, O. (1997). Assimilation of observations, an introduction. *J. Meteor. Soc. Japan*, 75(1B), 191–209.

Talagrand, O. and Miyakoda, K. (1971). The assimilation of past data in dynamical analysis. Part II. *Tellus*, 23, 318–327.

Thépaut, J., Hoffman, R.N., Courtier, P. (1993). Interactions of dynamics and observations in a 4-dimensional variational assimilation. *Mon. Weather Rev.*, 121, 3393–3414.

Trevisan, A., D'Isidoro, M., Talagrand, O. (2010). Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension. *Q. J. R. Meteorol. Soc.*, 136, 487–496.

Verlaan, M. and Heemink, A. (1997). Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrol. Hydraul.*, 11, 349–368. doi: 10.1007/BF02427924.

Verlaan, M. and Heemink, A. (2001). Nonlinearity in data assimilation applications: A practical method for analysis. *Monthly Weather Review*, 129, 1578–1589.