Variational Assimilation

Olivier Talagrand

Laboratoire de Météorologie Dynamique / CNRS
École Normale Supérieure, Paris, France, Talagrand@lmd.ens.fr

1 Introduction

The expression *variational assimilation* designates a class of assimilation algorithms in which the fields to be estimated are explicitly determined as minimizers of a scalar function, called the *objective function*, that measures the misfit to the available data. In particular, *four-dimensional variational assimilation*, usually abbreviated as *4D-Var*, minimizes the misfit between a temporal sequence of model states and the observations that are available over a given assimilation window. As such, and contrary to the standard Kalman filter and, more generally, to sequential algorithms for assimilation, it propagates the information contained in the data both forward and backward in time.

From a numerical point of view, variational algorithms require the minimization of a scalar function defined over a large dimensional space. That is possible in practice through the systematic use of the *adjoint* of the assimilating model.

We first describe variational assimilation in the context of statistical linear estimation, which also underlies the theory of the Kalman filter (Section 2). This leads to the definition of a general form for the objective function to be minimized. Minimization methods and the adjoint approach for computing gradients, are then succinctly described (Section 3), as well as practical implementation of variational assimilation (Section 4). A number of problems, associated in particular with the strong non-linearity of the governing equations, are discussed (Section 5). The adjoint approach is further discussed, concerning in particular uses other than variational assimilation (Section 6). Conclusions follow in Section 7.

A large part of what follows is derived in the framework of Bayesian and statistical estimation. $\mathcal{E}[\ ]$ will denote statistical expectation, and $\mathcal{N}(a,\ C)$ the Gaussian probability distribution (either scalar or vector) with expectation a and covariance C. The superscript ^{T} will denote transposition.

2 Variational assimilation in the context of statistical linear estimation

For an elementary introduction, consider the following situation. One wants to determine an unknown scalar quantity x^t (i.e. true state) from two observations of the form

$$z_1 = x^t + \varepsilon_1 \tag{1a}$$

$$z_2 = x^t + \varepsilon_2 \,. \tag{1b}$$

In these expressions, ε_1 and ε_2 are observational errors, whose exact values are unknown, but whose statistical properties are known. More precisely, it is assumed that these errors are centred ($\mathcal{E}[\varepsilon_1]=\mathcal{E}[\varepsilon_2]=0$), mutually uncorrelated ($\mathcal{E}[\varepsilon_1\varepsilon_2]=0$), and have respective variances $\mathcal{E}[\varepsilon_1^2]=s_1$ and $\mathcal{E}[\varepsilon_2^2]=s_2$. We look for an estimate of x, of the form $x^a \equiv \alpha_1 z_1 + \alpha_2 z_2$, ($\alpha_1+\alpha_2=1$), with α_1 and α_2 chosen so as to minimize the statistical quadratic estimation error $s \equiv \mathcal{E}[(x^a-x)^2]$. The answer is

$$x^a = \frac{s_2 z_1 + s_1 z_2}{s_1 + s_2} \tag{2}$$

i.e., each of the two measurements is weighted in inverse proportion to the variance of the error on that measurement. The corresponding quadratic estimation error, which minimizes s, and which we denote s_a, is given by

$$\frac{1}{s_a} = \frac{1}{s_1} + \frac{1}{s_2}. \tag{3}$$

The same estimate x^a would be obtained by considering z_1 as a "background" estimate for x, and z_2 as an "observation" (or the reverse), and then applying the standard formulas for the Kalman filter.

The same estimate can also be obtained as the minimizer of the function

$$x \rightarrow J(x) \equiv \frac{1}{2}\left[\frac{(x-z_1)^2}{s_1} + \frac{(x-z_2)^2}{s_2}\right]. \tag{4}$$

The meaning of this expression is clear. The squared deviation of x from either one of the two observations is weighted in inverse proportion of the variance of the error on that observation. Minimization of $J(x)$ therefore imposes that x must fit either observation to within its own accuracy. This leads to the estimate given by Eqs. 2-3.

Variational assimilation, as it is implemented at present in meteorological and oceanographical applications (see chapters *Numerical Weather Prediction*, Swinbank; *Ocean Data Assimilation*, Haines), minimizes a function which generalizes Eq. 4. In particular, in the linear case, and as in the elementary example above, it minimizes the statistical quadratic estimation error (on any component of the estimated fields individually), and is actually another algorithm for solving the same problem as the Kalman filter.

Consider the following more general estimation problem. Estimate an unknown vector $\mathbf{x}^t$ (with components x_i^t, i =1,…, n), belonging to *state space* $\mathcal{S}$, with dimension n, from a known *data vector* $\mathbf{z}$ (with components z_j, j =1,…, m), belonging to *data space* $\mathcal{D}$, with dimension m, of the form

$$\mathbf{z} = \boldsymbol{\Gamma}\mathbf{x}^t + \boldsymbol{\varepsilon}. \tag{5}$$

In Eq. 5, $\mathbf{\Gamma}$ is a known linear operator from $\mathcal{S}$ into $\mathcal{D}$, called the *data operator*, and represented by an $m\times n$-matrix. $\boldsymbol{\varepsilon}$ is a random vector in $\mathcal{D}$, called the *error vector*. The problem is, therefore, to invert the operator $\mathbf{\Gamma}$, taking into account, as far as possible, the statistical properties of the error $\boldsymbol{\varepsilon}$. The estimate of $\mathbf{x}^t$ is sought in the form of a linear (and *a priori* non-homogeneous) function of $\mathbf{z}$, *viz*.

$$\mathbf{x}^a = \mathbf{a} + \mathbf{A}\mathbf{z} \,, \tag{6}$$

where $\mathbf{a}$ is a vector of $\mathcal{S}$, and $\mathbf{A}$ is a linear operator from $\mathcal{D}$ into $\mathcal{S}$. $\mathbf{a}$ and $\mathbf{A}$ are to be determined under the following two conditions:

(i) The estimate $\mathbf{x}^a$ is invariant in a change of origin in state space (for instance, if the unknown $\mathbf{x}^t$ contains temperatures, the result must be independent of whether those temperatures are expressed in degrees Celsius or in Kelvins);

(ii) For any component x_i^t of $\mathbf{x}^t$, the statistical expectation of the square of the corresponding estimation error $x_i^a - x_i^t$ is minimized.

The solution to this problem is given by

$$\mathbf{x}^a = (\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{S}^{-1}(\mathbf{z}-\boldsymbol{\mu}) \tag{7}$$

[*i. e.*, $\mathbf{A} = (\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{S}^{-1}$ and $\mathbf{a} = -\mathbf{A}\boldsymbol{\mu}$], where $\boldsymbol{\mu} \equiv \mathscr{E}[\boldsymbol{\varepsilon}]$ and $\mathbf{S} \equiv \mathscr{E}[(\boldsymbol{\varepsilon}-\boldsymbol{\mu})(\boldsymbol{\varepsilon}-\boldsymbol{\mu})^{\mathrm{T}}]$ are, respectively, the expectation and covariance matrix of the error $\boldsymbol{\varepsilon}$. It is seen that $\mathbf{A}$ is a left-inverse of $\mathbf{\Gamma}$ (*i.e.*, $\mathbf{A}\mathbf{\Gamma} = \mathbf{I}_n$, where $\mathbf{I}_n$ is the unit matrix of order n), with the consequence that the estimate $\mathbf{x}^a$ is unbiased, ($\mathscr{E}[\mathbf{x}^a-\mathbf{x}^t] = 0$), and that the corresponding estimation error has covariance

$$\mathbf{P}^a \equiv \mathscr{E}[(\mathbf{x}^a - \mathbf{x}^t)(\mathbf{x}^a - \mathbf{x}^t)^T] = (\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma})^{-1} \,. \tag{8}$$

Condition (ii) above means that the trace of $\mathbf{P}^a$ is the minimum trace that can be obtained among all possible linear estimates of $\mathbf{x}$.

Equations 7-8 generalize Eqs. 2-3. The estimate $\mathbf{x}^a$ is called the *Best Linear Unbiased Estimate* (*BLUE*) of $\mathbf{x}$ from $\mathbf{z}$ (the term *Best Linear Unbiased Estimator* is also used). Its explicit determination requires the knowledge of (at most) the expectation $\boldsymbol{\mu}$ and the covariance matrix $\mathbf{S}$ of the data error $\boldsymbol{\varepsilon}$.

Taking Eq. 7 at face value, the unambiguous definition of the *BLUE* requires the matrix $\mathbf{S}$, and then the matrix $\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma}$, to be invertible. The need for invertibility of $\mathbf{S}$ is only apparent (without going into full details, $\mathbf{S}$ is singular when some components of $\mathbf{x}$ are exactly observed; it then suffices to restrict the estimation to those components that are not exactly observed). The condition for invertibility of $\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma}$, once $\mathbf{S}$ is invertible, is on the other hand real. It is equivalent to the condition that the null space of the data operator $\mathbf{\Gamma}$ is restricted to the 0-vector

$$\mathbf{\Gamma}\mathbf{x} = \mathbf{0} \;\Leftrightarrow\; \mathbf{x} = \mathbf{0} \tag{9}$$

or, equivalently, that $\mathbf{\Gamma}$ has rank equal to the dimension n of $\mathbf{x}$. This means that the data vector $\mathbf{z}$ contains information, either directly or indirectly, on every component of $\mathbf{x}$. The problem of determining $\mathbf{x}$ from $\mathbf{z}$ is *overdetermined*. This requires that $m \geq n$. There must be at least as many scalar data in $\mathbf{z}$ as there are scalar parameters to be determined. We will set $m = n+p$. The condition given by Eq. 9 will be called the *determinacy* condition.

The *BLUE* possesses a number of important properties.

- As already mentioned, the operator $\mathbf{A}$ is a left-inverse of $\mathbf{\Gamma}$. This means that, if the data are exact ($\boldsymbol{\varepsilon} = 0$ in Eq. 9), then so is the estimate $\mathbf{x}^a$;
- The *BLUE* is invariant in a change of origin in either data or state space. It is also invariant in any invertible linear change of coordinates in either space. This means, for instance, that a profile of observed temperatures can be transformed, through the hydrostatic equation, into a profile of geopotential values without altering the estimated fields. It also means that the horizontal wind can be estimated in terms of geometrical coordinates, or in terms of its divergence and vorticity. The result will be the same. This condition of invariance also means that the *BLUE* is independent of the choice of a scalar product, either in state or data space. For instance, for any symmetric definite positive matrix $\mathbf{C}$, the quantity $(\mathbf{x}^a-\mathbf{x})^T\mathbf{C}(\mathbf{x}^a-\mathbf{x})$, which is one (among infinitely many) measure of the magnitude of the estimation error $(\mathbf{x}^a-\mathbf{x})$, is minimized by the *BLUE*. The invariance of the *BLUE* in any invertible change of linear coordinates can also be expressed by saying that Eqs. 7-8 are more than vector-matrix equations. They are *tensor equations*, valid in any system of linear coordinates;
- When the data error $\boldsymbol{\varepsilon}$ is Gaussian, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{S})$, the *BLUE* achieves Bayesian estimation, in the sense that the conditional probability distribution for the state vector $\mathbf{x}$, given the data vector $\mathbf{z}$, is the Gaussian distribution with expectation $\mathbf{x}^a$ and covariance matrix $\mathbf{P}^a$, as given by Eqs. 7-8. In condensed notation, $\mathbf{P}(\mathbf{x} \mid \mathbf{z}) = \mathcal{N}(\mathbf{x}^a, \mathbf{P}^a)$.

It is easily verified that the *BLUE* $\mathbf{x}^a$ can be obtained as the minimizer of the following scalar function, defined over state space

$$\mathbf{x} \rightarrow J(\mathbf{x}) \equiv \frac{1}{2}[\Gamma\mathbf{x}-(\mathbf{z}-\boldsymbol{\mu})]^T\mathbf{S}^{-1}[\Gamma\mathbf{x}-(\mathbf{z}-\boldsymbol{\mu})]\,. \qquad (10)$$

This expression generalizes Eq. 4. Its significance is clear. For any vector $\mathbf{x}$ in state space, $\mathbf{\Gamma x}$ is what the data operator $\mathbf{\Gamma}$ would produce if it was applied on $\mathbf{x}$. $J(\mathbf{x})$ is then a measure of the magnitude of the discrepancy between $\mathbf{\Gamma x}$ and the unbiased data vector $\mathbf{z}$-$\boldsymbol{\mu}$. Through the inverse covariance matrix $\mathbf{S}^{-1}$, that measure possesses two notable properties. First, it weights the data according to their accuracy. Second, it is physically non-dimensional, making it possible to combine in a consistent way data of a different physical nature.

Variational assimilation, as it exists at present in meteorology and oceanography, minimizes objective functions of the form of Eq. 10, with the only difference, to be discussed later, that moderately non-linear operators $\mathbf{\Gamma}$ are used. What follows is a

more detailed description of how variational assimilation is implemented in practice, and of the main results it produces.

The first step in the minimization of a function such as that given by Eq. 10 is to remove the bias in the data by subtracting the error expectation $\boldsymbol{\mu}$ from the data vector. Unless specified otherwise, it will be assumed below that this has been done, and the expectation $\boldsymbol{\mu}$ will not appear any more explicitly in the equations. But it must be kept in mind that implementation of variational assimilation requires the prior knowledge, and subtraction from the data, of the error expectation, or *bias*. Failure to properly remove the bias in the data will, in general, result in the presence of residual biases in the estimated fields (chapter *Bias Estimation*. Ménard, discusses bias in data assimilation).

When the determinacy condition (Eq. 9) is verified, the data vector $\mathbf{z}$ can always be transformed, through linear invertible operations, into two components of the following form. First, an explicit estimate of the true state vector $\mathbf{x}^t$, of form

$$\mathbf{x}^b = \mathbf{x}^t + \boldsymbol{\varepsilon}^b, \tag{11}$$

where $\boldsymbol{\varepsilon}^b$ is an error; second, an additional set of data, of the form

$$\mathbf{y} = \mathbf{H}\mathbf{x}^t + \boldsymbol{\varepsilon}^o, \tag{12}$$

with dimension $p = m\text{-}n$. In this equation, $\mathbf{H}$ is a linear operator, represented by a $p\times n$ -matrix, and $\boldsymbol{\varepsilon}^o$ is an error. In addition, the transformations that lead to Eqs. 11-12 can always be defined in such a way that the errors $\boldsymbol{\varepsilon}^b$ and $\boldsymbol{\varepsilon}^o$ are uncorrelated

$$\mathscr{E}[\boldsymbol{\varepsilon}^b(\boldsymbol{\varepsilon}^o)^T] = 0. \tag{13}$$

It is in the form of Eqs. 11-12 that data are most often available in meteorological and oceanographical applications. The component $\mathbf{x}^b$ is a prior, or *background* estimate of the unknown state vector $\mathbf{x}$ at a given time k (usually a recent forecast, or a climatological estimate). As for the additional vector $\mathbf{y}$, it consists of observations depending on the state vector through the *observation operator* $\mathbf{H}$. The uncorrelation hypothesis (Eq. 13), although certainly disputable, is often (if not always) made. Equations 11-12, together with Eq. 13, are also assumed in the standard Kalman filter. We stress here that Eqs. 11-13 are no more restrictive than, but exactly equivalent to, Eq. 5 together with the determinacy condition, Eq. 9.

Introducing the covariance matrices of the errors $\boldsymbol{\varepsilon}^b$ and $\boldsymbol{\varepsilon}^o$

$$\mathbf{P}^b \equiv \mathscr{E}[\boldsymbol{\varepsilon}^b(\boldsymbol{\varepsilon}^b)^T], \quad \mathbf{R} \equiv \mathscr{E}[\boldsymbol{\varepsilon}^o(\boldsymbol{\varepsilon}^o)^T], \tag{14}$$

Eqs. 7-8 take the following form, used in particular in the Kalman filter

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{P}^b\mathbf{H}^T[\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}]^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^b) \tag{15a}$$

$$\mathbf{P}^a = \mathbf{P}^b - \mathbf{P}^b\mathbf{H}^T[\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}]^{-1}\mathbf{H}\mathbf{P}^b. \tag{15b}$$

We recall that the vector

$$\mathbf{d} \equiv \mathbf{y} - \mathbf{H}\mathbf{x}^b , \tag{16}$$

is called the *innovation vector*, and that the matrix $\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R}$, the inverse of which appears in Eqs. 15a-b, is the covariance matrix of $\mathbf{d}$

$$\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R} = \mathscr{E}[\mathbf{d}\mathbf{d}^T] . \tag{17}$$

As for the objective function (Eq. 10), it takes under decomposition of Eqs. 11-12 the following form

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T[\mathbf{P}^b]^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}) . \tag{18}$$

The meaning of this expression is clear. The first term on the right hand side of Eq. 18 is a measure of the deviation of $\mathbf{x}$ from the background, while the second term is a measure of the deviation from the observation.

Several situations are encountered in the practice of meteorology and oceanography, which we are going to describe in some detail, giving more explicit expressions for the general form (Eq. 18) of the objective function.

The simplest situation is when a background $\mathbf{x}^b$, of form given by Eq. 11, is available at some time k, together with observations, of form given by Eq. 12, that have been performed at the same time (or over a period of time short enough so that the flow can be considered stationary). Minimization of the objective function (Eq. 18) will produce an estimate of the state of the flow at time t. One then speaks in that case of *three-dimensional variational analysis*, often abbreviated as *3D-Var*.

A different, more complex, situation is encountered when one wants to assimilate observations that are distributed over a period of time over which the evolution of the flow cannot be neglected. Let us assume observations are available at successive times $k = 0, 1, \ldots, K$, of the form

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k^t + \boldsymbol{\varepsilon}_k^o , \tag{19}$$

where $\mathbf{x}_k^t$ is the exact true state of the flow at time k, $\mathbf{H}_k$ is a linear observation operator, and $\boldsymbol{\varepsilon}_k^o$ an observational error with covariance matrix $\mathbf{R}_k$. The observational errors are assumed to be uncorrelated in time. It is assumed in addition that the temporal evolution of the flow is described by the equation

$$\mathbf{x}_{k+1}^t = \mathbf{M}_k\mathbf{x}_k^t + \boldsymbol{\eta}_k , \tag{20}$$

with known model linear operator $\mathbf{M}_k$, and random model error $\boldsymbol{\eta}_k$.

Assume in addition that a background $\mathbf{x}_0{}^b$, with error covariance matrix $\mathbf{P}_0{}^b$, and error uncorrelated with the observational errors in Eq. 19, is available at time $k = 0$.

If the model error is ignored, any initial condition $\mathbf{x}_0$ at time $k = 0$ defines a model solution

$$\mathbf{x}_{k+1} = \mathbf{M}_k \mathbf{x}_k \qquad k = 0, \ldots, K-1. \tag{21}$$

The objective function

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T [\mathbf{P}_0^b]^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{k=0}^{K}(\mathbf{H}_k \mathbf{x}_k - \mathbf{y}_k)^T [\mathbf{R}_k]^{-1}(\mathbf{H}_k \mathbf{x}_k - \mathbf{y}_k) \tag{22}$$

which is of the general form given by Eq. 10, measures the distance between the model solution (Eq. 21) and the data. Minimization of $J(\mathbf{x}_0)$ will define the initial condition of the model solution that fits the data most closely. Following a terminology first introduced by Sasaki (1970a, b, c), this is called *strong constraint four-dimensional variational assimilation*, often abbreviated as *strong constraint 4D-Var*. The words "strong constraint" stress the fact that the model identified by Eq. 21 must be exactly verified by the sequence of estimated state vectors.

If the model error is taken into account, Eq. 20 defines an additional set of "noisy" data. We assume the model error $\boldsymbol{\eta}_k$ in Eq. 20 to have covariance matrix $\mathbf{Q}_k$, to be uncorrelated in time and to be uncorrelated with observation and background errors. Equation 10 then gives the following expression for the objective function defining the *BLUE* of the sequence of states $\{\mathbf{x}_k , k = 0,\ldots,K\}$

$$\begin{aligned} J(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_K) = {} & \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T [\mathbf{P}_0^b]^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) \\ & + \frac{1}{2}\sum_{k=0}^{K}(\mathbf{H}_k \mathbf{x}_k - \mathbf{y}_k)^T [\mathbf{R}_k]^{-1}(\mathbf{H}_k \mathbf{x}_k - \mathbf{y}_k) \\ & + \frac{1}{2}\sum_{k=0}^{K-1}(\mathbf{x}_{k+1} - \mathbf{M}_k \mathbf{x}_k)^T [\mathbf{Q}_k]^{-1}(\mathbf{x}_{k+1} - \mathbf{M}_k \mathbf{x}_k) . \end{aligned} \tag{23}$$

The objective function is now a function of the whole sequence of states $\{\mathbf{x}_k , k=0,\ldots,K\}$. Minimization of an objective function of the form given by Eq. 23, where the model equations are present as noisy data to be fitted by the analysed fields like any other data, is called, again according to the terminology introduced by Sasaki (1970a, b, c), *weak constraint four-dimensional variational assimilation*, abbreviated as *weak constraint 4D-Var*.

Equations 22-23, with appropriate redefinition of the state and observation spaces, are particular cases of Eq. 10. Another type of variational algorithm can be defined from Eq. 15a, which can be written as

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{P}^b \mathbf{H}^T \mathbf{w} , \tag{24}$$

where the vector $\mathbf{w} \equiv [\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}]^{-1}\,\mathbf{d}$ minimizes the objective function

$$K(\boldsymbol{\nu}) \equiv \frac{1}{2}\boldsymbol{\nu}^T[\mathbf{HPH}^T + \mathbf{R}]\boldsymbol{\nu} - \mathbf{d}^T\boldsymbol{\nu} \ . \tag{25}$$

This function is defined on the dual of the observation space, which has dimension p. Minimization of Eq. 25 corresponds to the *dual* approach to variational assimilation, by opposition to the *primal* approach, given by Eq. 18. The dual approach is also known as defining the *Physical Space Assimilation System* (*PSAS*, pronounced "pizzazz"; the word *Physical* is historical). Just as Eqs. 18, 22-23 are particular forms of Eq. 10, the dual approach can be used in any of the situations corresponding to those three equations. Depending on the conditions of the problem, and especially on the relative dimension of the state and observation space, it may be more advantageous to use the primal or the dual approach. A significant difference is that the dual approach uses the error covariance matrices $\mathbf{P}^b$ and $\mathbf{R}$ in their direct forms, while the primal approach requires their inverses. Another difference is that the dual approach requires an explicit background $\mathbf{x}^b$, while the primal approach can be implemented, in the general form given by Eq. 10, without an explicit background (it only requires the determinacy condition, Eq. 9).

All forms of variational assimilation given by Eqs. 18, 22-23 and 25 have been used, or at least extensively studied, for assimilation of meteorological and oceanographical observations. The theory of the *BLUE* requires the data operators ($\boldsymbol{\Gamma}$, $\mathbf{H}$ and $\mathbf{M}_k$ in the above notations) to be linear. In practice, this condition is rarely verified. In particular, variational assimilation of form given by Eq. 22 or Eq. 23 is almost always implemented with a non-linear model. From a heuristic point of view, it is clear that, if the non-linearity is in a sense sufficiently small, variational assimilation, even if it does not solve a clearly identified estimation problem, is likely to produce useful results (this point will be further discussed in Section 5 below). The dual approach, on the other hand, explicitly uses the transpose observation operator $\mathbf{H}^T$, and requires exact linearity.

3 Minimization methods. The adjoint approach

3.1. Gradient methods for minimization

Variational assimilation aims at minimizing an objective function of one of the forms defined in the previous section. The objective functions we will consider can be exactly quadratic or not. We will make a slight change of notation, and will systematically denote by $\mathbf{x}$, and will call *control variable*, the argument of the function to be minimized; in Eq. 23, the control variable is the whole sequence $\mathbf{x}_0$, …, $\mathbf{x}_K$, while it is $\boldsymbol{\nu}$ in Eq. 25. The control variable belongs to the *control space*, whose dimension will be denoted by N. We will denote by $\partial J/\partial \mathbf{x}$ the gradient of J with respect to $\mathbf{x}$, *i.e.*, the N-vector whose components are the partial derivatives of J with respect to the components x_i of $\mathbf{x}$, *viz.*,

$$\frac{\partial J}{\partial \mathbf{x}} = \left(\frac{\partial J}{\partial x_i} \right)_{i=1,\ldots,N} . \tag{26}$$

The gradient is equal to 0 at the minimum of the objective function. One way to determine the minimum could conceivably be (as is actually often done in simple small dimension problems) to determine analytical expressions for the components of the gradient, and then to solve a system of N scalar equations for the minimizing components of $\mathbf{x}$. In meteorological and oceanographical applications, the complexity of the computations defining the objective function (in 4D-Var, these calculations include the temporal integration of a numerical dynamical model of the flow over the assimilation window) makes it totally inconceivable even to obtain analytical expressions for the gradient. Another way to proceed is to implement an iterative minimization algorithm, which determines a sequence of successive approximations $\mathbf{x}^{(l)}$ of the minimizing value of $\mathbf{x}$, *viz.*,

$$\mathbf{x}^{(l+1)} = \mathbf{x}^{(l)} - \mathbf{D}^{(l)} , \tag{27}$$

where $\mathbf{D}^{(l)}$ is at every iteration an appropriately chosen vector in control space. One possibility is to choose $\mathbf{D}^{(l)}$ along the direction of the local gradient $\partial J/\partial \mathbf{x}$. Algorithms which are based on that choice, called *steepest descent* algorithms, turn out, however, not to be numerically very efficient. In other algorithms, the vector $\mathbf{D}^{(l)}$ is determined as a combination of the local gradient and of a number of gradients computed at previous steps of the iteration, Eq. 27 (see, *e.g.*, Bonnans *et al*. 2003). All minimization methods that are efficient for large dimensions are of the form given by Eq. 27, and require the explicit determination, at each iteration step, of the local gradient $\partial J/\partial \mathbf{x}$. They are called *gradient methods*. Since one cannot hope to obtain an analytical expression for the gradient, it must be determined numerically. One possibility could be to determine it by finite differences, by imposing in turn a perturbation Δx_i on all components x_i of the control vector, and approximating the partial derivative $\partial J/\partial x_i$ by the difference quotient

$$\frac{\partial J}{\partial x_i} \approx \frac{J(\mathbf{x} + \Delta x_i) - J(\mathbf{x})}{\Delta x_i} . \tag{28}$$

This, however, would require N explicit computations of the objective function, *i.e.*, in the case of four-dimensional assimilation, N integrations of the assimilating model. Although that has actually been done for variational assimilation of meteorological observations, in an experimental setting, and with a relatively small dimension model (Hoffman 1986), it would clearly be impossible in any practical application.

3.2 The adjoint method

The *adjoint method* allows numerical computation of the gradient of a scalar function at a cost that is at most a few times the cost of the direct computation of that function. Adjoint equations are an extremely powerful mathematical and numerical tool. They are central to the theory of *optimal control*, *i.e.*, the theory of how the behaviour of a physical system can be controlled by acting on some of its components (see for instance the book by Lions 1971). Adjoint equations can also be used for solving mathematical problems in their own right. The use of adjoint equations in meteorological and oceanographical applications was advocated by the Russian school of mathematics at an early stage of development of numerical modelling of the atmosphere and ocean (see, *e.g.*, Marchuk 1974). We are going to demonstrate the method of adjoint equations in the special case of strong constraint 4D-Var (Eq. 22), in the most general case where the model and observation operators can be non-linear.

In order to stress the possible non-linearity of the model and observation operators, we now introduce the non-linear model operator, $\mathcal{M}_k(\ )$, and the non-linear observation operator, $\mathcal{H}_k(\ )$. The notation for operators used hitherto in this chapter, $\mathbf{M}_k$ and $\mathbf{H}_k$ (denoting linear model and observation operators, respectively), being reserved hereafter for the *Jacobians* (matrices of partial derivatives) of $\mathcal{M}_k(\ )$ and $\mathcal{H}_k(\ )$, respectively. We rewrite Eqs. 21-22 with non-linear operators as

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T [\mathbf{P}_0^b]^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{k=0}^{K}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k)^T [\mathbf{R}_k]^{-1}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k) \tag{29a}$$

with

$$\mathbf{x}_{k+1} = \mathcal{M}_k(\mathbf{x}_k), \quad k = 0,\ldots,K-1. \tag{29b}$$

Our purpose is to determine the gradient $\partial J/\partial \mathbf{x}_0$ of J with respect to $\mathbf{x}_0$. That gradient is characterized by the property that, for any perturbation $\delta\mathbf{x}_0$ of $\mathbf{x}_0$, the corresponding variation of J is, to first order with respect to $\delta\mathbf{x}_0$, equal to

$$\delta J = \left(\frac{\partial J}{\partial \mathbf{x}_0}\right)^T \delta\mathbf{x}_0 \,. \tag{30}$$

The perturbation $\delta\mathbf{x}_0$ results at later times in perturbations which, through differentiation of Eq. 29b, are given to first order by

$$\delta\mathbf{x}_{k+1} = \mathbf{M}_k \delta\mathbf{x}_k \,, \quad k = 0,\ldots,K-1 \,, \tag{31}$$

where, as said, $\mathbf{M}_k$ is the Jacobian of $\mathbf{x}_{k+1}$ with respect to $\mathbf{x}_k$. Equation 31 is called the *tangent linear equation* of Eq. 29b. Although the dependence is not explicit in Eq.

31, it must be kept in mind that the Jacobian $\mathbf{M}_k$ will, in general, depend in the non-linear case on the local value of $\mathbf{x}_k$.

As for the first order variation of the objective function J, it is given by differentiation of Eq. 29a, *viz.*,

$$\delta J = (\mathbf{x}_0 - \mathbf{x}_0^b)^T [\mathbf{P}_0^b]^{-1} \delta\mathbf{x}_0 + \sum_{k=0}^{K} (\mathscr{H}_k(\mathbf{x}_k) - \mathbf{y}_k)^T \mathbf{R}_k^{-1} \mathbf{H}_k \delta\mathbf{x}_k \,, \tag{32}$$

where $\mathbf{H}_k$ is the local Jacobian of $\mathscr{H}_k$, and where the $\delta\mathbf{x}_k$'s are given by Eq. 31.

δJ is a compound function of $\delta\mathbf{x}_0$ through the $\delta\mathbf{x}_k$'s. Our purpose is to "skip" the intermediate $\delta\mathbf{x}_k$'s, and to obtain a direct dependence of δJ with respect to $\delta\mathbf{x}_0$ of form given by Eq. 30. To that end, we introduce at each time $k = 1,\ldots, K$ a vector $\boldsymbol{\lambda}_k$, belonging to the dual of state space (and therefore with dimension n), and to be defined more precisely later. We form the products $\boldsymbol{\lambda}_k^T (\delta\mathbf{x}_k - \mathbf{M}_{k-1}\delta\mathbf{x}_{k-1})$, which, according to Eq. 31, are equal to 0. Subtracting those products from the right-hand side of Eq. 32 yields

$$\begin{aligned} \delta J = (\mathbf{x}_0 - \mathbf{x}_0^b)^T [\mathbf{P}_0^b]^{-1} \delta\mathbf{x}_0 &+ \sum_{k=0}^{K} (\mathscr{H}_k(\mathbf{x}_k) - \mathbf{y}_k)^T \mathbf{R}_k^{-1} \mathbf{H}_k \delta\mathbf{x}_k \\ &- \sum_{k=0}^{K} \lambda_k^T (\delta\mathbf{x}_k - \mathbf{M}_{k-1}\delta\mathbf{x}_{k-1}) \end{aligned} \tag{33}$$

(subtracting rather than adding the products is of course arbitrary, but convenient). We now transform Eq. 33 by first using the fact that the transpose of a matrix product is the product of the corresponding transposes, taken in reversed order. For instance, the product $(\mathscr{H}_k(\mathbf{x}_k) - \mathbf{y}_k)^T \mathbf{R}_k^{-1} \mathbf{H}_k$ is equal to $([\mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathscr{H}_k(\mathbf{x}_k) - \mathbf{y}_k)])^T$ (where use has been made of the fact that the covariance matrix $\mathbf{R}_k$ is symmetric), thus transforming the (scalar) quantity $(\mathscr{H}_k(\mathbf{x}_k) - \mathbf{y}_k)^T \mathbf{R}_k^{-1} \mathbf{H}_k \delta\mathbf{x}_k$ into the scalar product of the two n–vectors $\mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathscr{H}_k(\mathbf{x}_k) - \mathbf{y}_k)$ and $\delta\mathbf{x}_k$. Performing that operation on all terms in Eq. 33 and gathering all terms with common factor $\delta\mathbf{x}_k$ yields

$$\begin{aligned} \delta J = &\{ [\mathbf{P}_0^b]^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) + \mathbf{H}_0^T \mathbf{R}_0^{-1} (\mathscr{H}_0(\mathbf{x}_0) - \mathbf{y}_0) + \mathbf{M}_0^T \boldsymbol{\lambda}_1 \}^T \delta\mathbf{x}_0 \\ &+ \sum_{k=1}^{K-1} \{ \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathscr{H}_k(\mathbf{x}_k) - \mathbf{y}_k) - \boldsymbol{\lambda}_k + \mathbf{M}_k^T \boldsymbol{\lambda}_{k+1} \}^T \delta\mathbf{x}_k \\ &+ \{ \mathbf{H}_K^T \mathbf{R}_K^{-1} (\mathscr{H}_K(\mathbf{x}_K) - \mathbf{y}_K) - \boldsymbol{\lambda}_K \}^T \delta\mathbf{x}_K \,. \end{aligned} \tag{34}$$

This expression is valid for any choice of the $\boldsymbol{\lambda}_k$'s. It is seen that choosing

$$\boldsymbol{\lambda}_K = \mathbf{H}_K^T \mathbf{R}_K^{-1} (\mathscr{H}_K(\mathbf{x}_K) - \mathbf{y}_K) \,, \tag{35a}$$

and then recursively

$$\boldsymbol{\lambda}_k = \mathbf{M}_k^T \boldsymbol{\lambda}_{k+1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathscr{H}_k(\mathbf{x}_k) - \mathbf{y}_k) \quad \text{for} \quad k = K-1,\ldots,1 \ , \tag{35b}$$

eliminates all $\delta\mathbf{x}_k$ terms in Eq. 34, except the $\delta\mathbf{x}_0$ term. Defining further

$$\boldsymbol{\lambda}_0 = \mathbf{M}_0^T \boldsymbol{\lambda}_1 + [\mathbf{P}_0^b]^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) + \mathbf{H}_0^T \mathbf{R}_0^{-1} (\mathscr{H}_0(\mathbf{x}_0) - \mathbf{y}_0) \tag{35c}$$

there remains

$$\delta J = \boldsymbol{\lambda}_0^T \delta\mathbf{x}_0 \ , \tag{36}$$

which shows that $\boldsymbol{\lambda}_0$ is the required gradient of the objective function with respect to the initial condition $\mathbf{x}_0$ (see Eq. 30).

Equations 35a-c make up the *adjoint* of the tangent linear equation, Eq. 31. The word "adjoint" comes from the fact that Eqs. 35a-c are built on the transpose matrices $\mathbf{H}_k^T$ and $\mathbf{M}_k^T$, which are particular cases of the more general notion of adjoint operators. The adjoint equation is defined for the particular solution $\mathbf{x}_k$ of the basic equation (29b) for which the gradient is to be determined. It depends on that solution through the terms $\mathscr{H}_k(\mathbf{x}_k) - \mathbf{y}_k$ and, in the case of either a non-linear model operator $\mathscr{M}_k$ or a non-linear observation operator $\mathscr{H}_k$, through the transpose Jacobians $\mathbf{M}_k^T$ and/or $\mathbf{H}_k^T$. It is often said for convenience that Eqs. 35a-c define the adjoint of the basic model given by Eq. 29b, but it must be kept in mind that the adjoint equation is defined for a particular solution of that model.

The computations to be performed for determining the gradient $\partial J/\partial\mathbf{x}_0$ for given initial condition $\mathbf{x}_0$ are now clearly defined:

(1) Starting from $\mathbf{x}_0$, integrate the basic equation (29b). Store the corresponding solution $\mathbf{x}_k$ in memory;

(2) Starting from the "final" condition (Eq. 35a) at time K, integrate the adjoint equations 35b-c backward in time. The required gradient is $\boldsymbol{\lambda}_0$. The direct solution $\mathbf{x}_k$ is necessary for computing the terms $\mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathscr{H}_k(\mathbf{x}_k) - \mathbf{y}_k)$ and, in case the basic model (Eq. 29b) is non-linear, for determining the transpose Jacobian $\mathbf{M}_k^T$.

The determination of the gradient therefore requires one forward integration of the basic model (Eq. 29b), followed by one backward integration of the adjoint model (Eqs. 35a-c). The latter is a modified form of the direct model, and the corresponding cost must be of similar magnitude to the cost of integrating the direct model. It can be rigorously shown that, in terms of the number of arithmetic operations to be performed, the cost of one adjoint computation of the gradient $\partial J/\partial\mathbf{x}_0$ is at most four times the cost of one computation of the objective function, J. In meteorological and oceanographical applications, the cost of one adjoint integration (in terms of elapsed computer time) is typically twice the cost of one direct integration. This ratio is basically independent of the dimension N of the control variable, and makes the adjoint computation of a gradient much more economical than the N direct model integrations that would be required if the

gradient was to be computed by explicit perturbations. It is this fact that made variational assimilation possible at all in the first place.

Not surprisingly, there is a price to be paid for this major reduction in computing time. The price, as seen above, is the necessity to store in memory the direct solution $\mathbf{x}_k$. More precisely, what has to be kept in memory (or else to be recomputed in the course of the adjoint integration) are all quantities that are arguments of non-linear operations in the direct integration. Relaxing the storage constraint, for instance by using a more economical approximate adjoint, is difficult. Experience shows that minimization algorithms, especially efficient ones, are very sensitive to even slight misspecification of the gradient. The question of how the cost of variational assimilation can be reduced will be discussed in the next section**.**

The description that has just been given of the adjoint method is fundamentally sufficient for 4D-Var. It obviously covers the case of 3D-Var (minimization of an objective function of form given by Eq. 18), which does not involve a dynamical model of the flow. In that case, of course, only the transpose Jacobian $\mathbf{H}^T$ of the observation operator is needed.

The first attempt at using the adjoint approach for variational assimilation of meteorological observations was made by Penenko and Obraztsov (1976), on a simple one-level linear atmospheric model, and with synthetic data. Later attempts were made by Lewis and Derber (1985), Le Dimet and Talagrand (1986) and Talagrand and Courtier (1987). Courtier and Talagrand (1987) first used real data, while Thacker and Long (1988) made the first attempt at using adjoint equations for variational assimilation of oceanographical observations. Thépaut and Courtier (1991) first used a full primitive equation meteorological model. These early works showed that variational assimilation of meteorological or oceanographical observations was numerically feasible at an acceptable cost, and produced physically realistic results. Variational assimilation was progressively applied to more and more complex numerical models. It was introduced in 1997 in operational prediction, in the strong-constraint formulation, at the European Centre for Medium-Range Weather Forecasts, ECMWF (Klinker *et al.* 2000), and in 2000 at the French Meteorological Service (Météo-France). In both places, operational implementation of variational assimilation has resulted in significant improvements of the ensuing forecasts (see chapter *Assimilation of Operational Data*, Andersson and Thépaut). Some of these improvements were due to side effects not directly linked to the variational character of the assimilation, but others, especially in a number of specific meteorological situations, were due to better consistency between the assimilated states and the dynamics of the atmospheric flow. Since then, other meteorological services, such as the Japan Meteorological Agency, the Meteorological Office (United Kingdom), the Meteorological Service of Canada and the China Meteorological Administration, have introduced variational assimilation in their operational prediction system. All these schemes are of the strong-constraint form, and use a 6-hour assimilation window (12-hour in the case of ECMWF). In addition, ECMWF, after having produced several sets of reanalysed past observations, all based on sequential assimilation algorithms, is now running a new reanalysis project (the *ERA-Interim* project, http://www.ecmwf.int/research/era/do/get/era-interim) based on variational

assimilation. A specific advantage of variational assimilation in the case of reanalysis of past data is that it propagates information both forward and backward in time, thus allowing the use of observations that have been performed after estimation time.

Similar developments have taken place in oceanography, and variational assimilation using the adjoint of oceanographic circulation models is now commonly used for many diverse applications (although not so far for operational oceanographic prediction). Those applications include determination of the initial conditions of the flow, as described above (see, *e.g.*, Weaver and Anderson 1997; Vialard *et al.* 2003; Ricci *et al.* 2005), but also identification of "parameters", such as wind stress at the surface of the ocean (Vossepoel *et al.* 2004). Egbert *et al.* (1994) and Louvel (2001) used the dual approach through minimization in dual observation space of an objective function of form given by Eq. 25. In that approach, each iteration of the minimization process requires first a backward integration of the adjoint model, followed by a forward integration of the tangent linear model. Variational assimilation has also extended to other fields of geophysics and environmental sciences, such as atmospheric chemistry (Fisher and Lary 1995; Errera and Fonteyn 2001; Elbern *et al.* 2007; Lahoz *et al.* 2007 – see also chapters in Part D, *Chemistry*), or surface hydrology (Reichle 2000 – see chapter *Land Surface Data Assimilation*, Houser *et al.*). Other extensions of the variational methodology, that have largely benefited from the experience in meteorology, have been to terrestrial magnetism (Fournier *et al.* 2007; Sun *et al.* 2007) and seismology (Tromp *et al.* 2005).

4 Practical implementation

If the principle of variational assimilation and of the adjoint method is conceptually perfectly clear and rigorous, practical implementation of variational assimilation raises a number of serious problems. We will discuss below the specific problems associated with the development and validation of a code for performing the adjoint computations defined by Eq. 35, and are going to consider first a number of purely numerical problems.

4.1 The incremental approach

The developments of the previous section seem to require that it is the adjoint of the complete model (Eq. 29b) that has to be used for the computation of the gradient of the objective function. A Numerical Weather Prediction (NWP) model is an extremely complex and lengthy code, and the ensuing "all-or-nothing" choice (take the complete adjoint of the model, or else do nothing) seems particularly impractical. Simplifying the adjoint equation as such, without modification of the direct model nor of the objective function, is not an appropriate solution. That would lead to an approximate gradient of the objective function, and, as has already been said, experience shows that minimization algorithms, especially efficient ones, are very sensitive to even slight misspecification of the gradient. A convenient and versatile solution, known as the *incremental approach* to variational assimilation, has been

introduced by Courtier *et al.* (1994). Several variants of that approach exist. We are going to describe the one that is conceptually the simplest.

The basic idea is to simplify the dynamical model (Eq. 29b) to a form that is both more economical and more manageable, in particular as concerns the adjoint. But that is not done on the model (Eq. 29b) itself, but rather on the tangent linear model (Eq. 31). A *reference solution* $\mathbf{x}_k^{(0)}$ of the basic equation (29b) having been determined (emanating for instance from the background $\mathbf{x}_0^b = \mathbf{x}_0^{(0)}$), the corresponding tangent linear model (Eq. 31) is modified to

$$\delta\mathbf{x}_{k+1} = \mathbf{L}_k \delta\mathbf{x}_k , \quad k = 0,\ldots,K-1 , \tag{37}$$

where $\mathbf{L}_k$ is, at any time k, an appropriately chosen "simpler" operator than the Jacobian $\mathbf{M}_k$. Consistency then requires to modify the basic model (Eq. 29b) in such a way that the tangent linear equation corresponding to solution $\mathbf{x}_k^{(0)}$ is Eq. 37. This is achieved by making the initial condition $\mathbf{x}_0 \equiv \mathbf{x}_0^{(0)} + \delta\mathbf{x}_0$ evolve into $\mathbf{x}_k \equiv \mathbf{x}_k^{(0)} + \delta\mathbf{x}_k$, where $\delta\mathbf{x}_k$ itself evolves according to Eq. 37. That makes the basic dynamics linear.

As for the objective function (Eq. 29a), several possibilities exist, at least when the observation operators are non-linear. One possibility is to linearize those operators just as the model operator $\mathscr{M}_k$ has been linearized. This leads to replacing the quantity $\mathscr{H}_k(\mathbf{x}_k)$ by $\mathscr{H}_k(\mathbf{x}_k^{(0)}) + \mathbf{N}_k \delta\mathbf{x}_k$, where $\mathbf{N}_k$ is an appropriate simplified linear operator (possibly, but not necessarily, the Jacobian of $\mathscr{H}_k$ at point $\mathbf{x}_k$). The objective function (Eq. 29a) is then replaced by

$$\begin{aligned} J_1(\delta\mathbf{x}_0) = \; & \frac{1}{2}(\delta\mathbf{x}_0 + \mathbf{x}_0^{(0)} - \mathbf{x}_0^b)^T [\mathbf{P}_0^b]^{-1} (\delta\mathbf{x}_0 + \mathbf{x}_0^{(0)} - \mathbf{x}_0^b) \\ & + \frac{1}{2}\sum_{k=0}^{K} (\mathbf{N}_k \delta\mathbf{x}_k - \mathbf{d}_k)^T \mathbf{R}_k^{-1} (\mathbf{N}_k \delta\mathbf{x}_k - \mathbf{d}_k) , \end{aligned} \tag{38}$$

where the $\delta\mathbf{x}_k$'s are subject to Eq. 37, and where $\mathbf{d}_k \equiv \mathbf{y}_k - \mathscr{H}_k(\mathbf{x}_k^{(0)})$ is the innovation at time k.

The function given by Eq. 38 is an exactly quadratic function of the initial perturbation $\delta\mathbf{x}_0$. The minimizing perturbation $\delta\mathbf{x}_{0,m}$ defines a new initial state $\mathbf{x}_0^{(1)} \equiv \mathbf{x}_0^{(0)} + \delta\mathbf{x}_{0,m}$, from which a new solution $\mathbf{x}_k^{(1)}$ of the basic equation (Eq. 29b) is computed. The process is then repeated for solution $\mathbf{x}_k^{(1)}$.

This defines a system of two-level nested loops for minimization of the original objective function (Eq. 29a). The fundamental advantage of the incremental approach is that it allows one to define at will the simplified linearized operators $\mathbf{L}_k$ and $\mathbf{N}_k$. Many degrees of freedom are available for ensuring an appropriate trade-off between practical implementability and meteorological accuracy and usefulness. The simplified dynamics in Eq. 37 can itself be modified in the course of the minimization, by progressively introducing more and more complex dynamics or "physics" in the successive outer loops.

It is the incremental method which, after the adjoint method, makes variational assimilation feasible. It is implemented, either in the form that has just been described or in slightly different variants, in most (if not all) operational NWP systems that use variational assimilation. At ECMWF, it is implemented with two outer loops, the approximations introduced in the linearized dynamics (Eq. 37) consisting first, of a reduced spatial resolution (from triangular spectral truncation T799 to T255 for the second outer loop) and, second, of a simplified "physical" package.

An obvious question is whether the nested-loop process of the incremental process converges and, if it does, to what it converges. In the case where the linearized operators $\mathbf{L}_k$ and $\mathbf{N}_k$ vary from one outer loop to the next, the possible convergence of the process can depend on the way those operators vary. In particular, convergence to the minimum of the original objective function (Eq. 29a) is possible only if the linear operators $\mathbf{L}_k$ and $\mathbf{N}_k$ converge to the corresponding Jacobians $\mathbf{M}_k$ and $\mathbf{H}_k$ at that minimum. The question of the convergence of the incremental process has been studied in some detail by Trémolet (2007) on the ECMWF 4D-Var system. Numerical tests show that the process does not converge asymptotically, at least in the conditions in which it is implemented at ECMWF. The way the incremental approach is implemented, at ECMWF and elsewhere, is largely based on empirical tuning.

4.2 First-Guess-At-the-right-Time 3D-Var

An extreme case of the incremental approach is what is called *First-Guess-At-the-right-Time 3D-Var*, or *FGAT 3D-Var*. It can be described as a process of form of Eqs. 37-38 in which the simplified linear operator $\mathbf{L}_k$ is taken as the identity operator. This process is four-dimensional in that the observations distributed over the assimilation window are compared with their analogues in a time-evolving reference integration of the assimilating model. But it is three-dimensional in that the minimization of the objective function (Eq. 38) does not use any explicit dynamics other than the trivial dynamics expressed by the unit operator, and that the numerical implementation is in effect three-dimensional. The FGAT 3D-Var approach, which is implemented through a unique minimization (no nested loops), has been shown to improve the quality of the assimilated fields, simply through the fact that is effectively uses a more exact innovation vector than does standard 3D-Var, in which all observations over the assimilation window are compared to the same first-guess field.

5 Further considerations on variational assimilation

Independently of its numerical and algorithmic properties, the major advantage of variational assimilation is that it takes into account, through the adjoint equation, the temporal evolution of the uncertainty in the state of the flow, at least over the assimilation window. Although (contrary to the Kalman filter) it does not explicitly compute the evolution of the uncertainty as such (and, in particular, does not produce an explicit estimate of the uncertainty in the estimated fields), it determines an approximation of the minimizing solution of the objective function (Eq. 29), which

depends on the dynamics of the flow, and of the temporal evolution of the uncertainty. This was shown in full detail by Thépaut *et al*. (1993), who compared the impact of individual observations in a 3D-Var process, which ignores the temporal evolution of the uncertainty, and a 4D-Var process. The impact was significantly different, and strongly dependent on the dynamical state of the flow, in the latter case.

Significant impact does not of course mean positive impact. All operational implementations of 4D-Var have been preceded by the development and implementation of a 3D-Var system. This is very convenient in that it allows progressive introduction of the various components of the full 4D-Var system. But it also provides the opportunity for systematic comparison of 3D-Var and 4D-Var. The comparison has always shown the superiority of 4D-Var, in particular in terms of the quality of the ensuing forecasts. Similar comparisons have also been performed, with the same conclusions, on other, non-operational assimilation systems. See also Lorenc and Rawlins (2005) for a detailed discussion of 3D-Var and 4D-Var.

All operational implementations of 4D-Var have so far been of the strong constraint form. In spite of the constant improvement of NWP models, the hypothesis of a perfect model is of course highly disputable. Weak-constraint assimilation, which corresponds to minimization of an objective function of form given by Eq. 23, would certainly be desirable. It however requires a quantitative estimate, in the form of the covariance matrix $\mathbf{Q}_k$, of the model error. A reliable estimate may be difficult to obtain. Derber (1989) has suggested identifying a possible systematic bias in the model by introducing that bias in the control variable. Other authors (Zupanski 1997; Trémolet 2006) have studied algorithms of the general form given by Eq. 23. There is some indication (*M. Fisher, pers. comm.*) that weak constraint variational assimilation could be useful over longer assimilation windows (24 hours or more) than used in strong constraint assimilation. That is easily understandable in view of the fact that the perfect model hypothesis becomes less and less valid as the length of the assimilation window increases.

The primal weak-constraint objective function (Eq. 23) becomes singular in the limit of a perfect model ($\mathbf{Q}_k$=0). As already said, the dual approach uses the data error covariance matrices in their direct form, so that the dual objective function (Eq. 25), as defined for weak constraint variational assimilation, is regular for $\mathbf{Q}_k$=0. This means that the same dual algorithm can be used for both strong- and weak-constraint variational assimilation. This is an attactive feature of the dual approach.

Courtier (1997) has shown that, subject to an appropriate preconditioning of the dual variable $\boldsymbol{\nu}$ in Eq. 25, the numerical conditioning (and therefore the numerical cost) of the dual algorithm is the same as that of the primal approach. In variational assimilation, it is actually the repeated numerical integrations of the direct and adjoint models that takes the major part of the computations, and the numerical cost of strong- and weak-constraint variational assimilation is fundamentally the same. This point is discussed in more detail in Louvel (2001).

The dual approach requires strict linearity of the operator $\mathbf{H}$ in Eq. 25 which, in the case of variational assimilation, means strict linearity of the model and observation operators. Auroux and Blum (2002, 2004) have introduced a double-loop algorithm (which has some similarity with the incremental approach described

above) in which successive linear problems of form given by Eq. 25 are solved, each one being based on a linearization about the result of the previous one.

More generally, and independently of the particular numerical algorithm that is used, the validity of the linear approach defined by Eqs. 7 and 10 is questionable in meteorological and oceanographical applications. It has already been said that, from a purely heuristic point of view, the linear approach must be valid if the non-linearities are in a sense small enough. A more accurate description of the real situation that is encountered in meteorology and oceanography is given, rather than by Eqs. 11-12, by

$$\mathbf{x}^b = \mathbf{x}^t + \boldsymbol{\varepsilon}^b \,, \tag{39}$$

$$\mathbf{y} = \mathcal{H}^*(\mathbf{x}^t) + \boldsymbol{\varepsilon} \,, \tag{40}$$

where $\mathcal{H}^*$ ($\mathcal{H}$-star) denotes a non-linear observation operator. In the case of 3D-Var, $\mathcal{H}^*$ is the observation operator at estimation time. In the case of 4D-Var, the vector $\mathbf{y}$ denotes the complete temporal sequence of observations, and the operator $\mathcal{H}^*$ includes the (non-linear) dynamical model. The knowledge of the data (Eqs. 39-40) is equivalent to the knowledge of Eq. 39 together with what can be called the non-linear innovation vector

$$\mathbf{d} \equiv \mathbf{y} - \mathcal{H}^*(\mathbf{x}^b) = \mathcal{H}^*(\mathbf{x}^t) - \mathcal{H}^*(\mathbf{x}^b) + \boldsymbol{\varepsilon} \,. \tag{41}$$

If the background $\mathbf{x}^b$ is close enough to the real unknown state $\mathbf{x}^t$, $\mathbf{d}$ can be approximated by

$$\mathbf{d} \approx \mathbf{H}(\mathbf{x}^t - \mathbf{x}^b) + \boldsymbol{\varepsilon} \,, \tag{42}$$

where $\mathbf{H}$ is here the Jacobian of the full operator $\mathcal{H}^*$ at point $\mathbf{x}^b$. If the so-called *tangent linear approximation* defined by Eq. 42 is valid, Eqs. 39-42 define an estimation problem that is linear with respect to the deviation $\mathbf{x}^t$-$\mathbf{x}^b$ of the real state with respect to the background $\mathbf{x}^b$. Equations 15 and 18 are then valid, $\mathbf{H}$ being the Jacobian of $\mathcal{H}^*$. In the case of 4D-Var, this leads to minimization of an objective function of the incremental form given by Eqs. 37-38, where the operators $\mathbf{L}_k$ and $\mathbf{N}_k$ replace the exact Jacobians $\mathbf{M}_k$ and $\mathbf{H}_k$ along the (full non-linear) reference model solution.

Both direct (see, *e*.g., Lacarra and Talagrand 1988) and indirect evidence shows that the tangent linear approximation is valid for large scale geostrophic atmospheric flow (scales larger than 200 km) up to about 24-48 hours. This limit, however, rapidly decreases with decreasing spatial scales, to be of the order of a few hours for convective scales. For oceanic geostrophic flow (scales larger than a few tens of kilometres), the limit is a few weeks.

The developments of this chapter are therefore fully valid within those limits. It is to be stressed, however, that in circumstances where the tangent linear approximation is known or hypothesized to be valid, the linearization in Eq. 42 is rarely performed explicitly. Either fully non-linear operators are kept in the objective function to be minimized, or (as is actually the case in the incremental approach described above) approximations that go further than Eq. 42 are implemented. The only case where the linearization given by Eq. 42 seems to have explicitly been

implemented is in the above-mentioned works of Auroux and Blum (2002, 2004) relative to the dual approach, which requires exactly linear operators.

But the question arises of what is to be done in circumstances when the tangent linear approximation is not valid. In the context of 4D-Var, there are actually two different questions, depending on the strength of the non-linearities. If the non-linearities are weak, the minimization of an objective function of the general form given by Eq. 29 remains numerically feasible, but may not be justified on the basis of estimation theory. If the non-linearities are strong, even the numerical minimization of the objective function, owing for instance to the presence of distinct minima, can raise difficulties.

These questions have not been discussed so far in much depth. One can mention the work of Pires *et al*. (1996), who studied variational assimilation for a strongly chaotic non-linear system (specifically, the celebrated three-parameter system of Lorenz 1963). These authors have shown that the objective function given by Eq. 29 possesses an increasing number of local minima with increasing length of the assimilation window. This can be easily understood in view of the repeated folding in state space that is associated with chaos. They have defined a procedure, called *Quasi-Static Variational Assimilation* (*QSVA*), in which the length of the assimilation window, starting from a value for which the objective function (Eq. 29) possesses a unique minimum, is progressively increased. Each new minimization is started from the result of the previous one. This allows one to keep track of the absolute minimum of the objective function, at least if the temporal density of observations is in a sense high enough. QSVA has been implemented on a quasi-geostrophic atmospheric model by Swanson *et al*. (1998) who have been able to usefully extend variational assimilation (in the hypothesis of a perfect model) to assimilation windows as long as five days. This is largely beyond the limit of validity of the tangent linear approximation. QSVA, or a similar algorithm, could possibly be implemented in operational practice, for instance by using successive overlapping assimilation windows.

Other developments have taken place recently at the research level. Carrassi *et al*. (2008) have defined a 3D-Var system in which the control variable, instead of consisting of the whole state vector, is restricted to the deviations from the background along the (relatively few) unstable modes of the system. This approach is now being extended to 4D-Var (*Trevisan, pers. comm.*). A somewhat similar work has been performed by Liu *et al*. (2008), who have developed a low-order incremental 4D-Var system. The background error covariance matrix $\mathbf{P}_0^b$ (Eq. 38) is defined, not on the basis of an *a priori* statistical model, but on the basis of the dispersion of an ensemble of background forecasts. As in Carrassi *et al*. (2008), the control space is not the entire state space, but the state spanned by the background forecasts. Taking advantage of the relatively small dimension of the control space, and of the linearity associated with the incremental character of the procedure, it is not necessary to use an adjoint code for computing the gradient of the objective function. That can be achieved through simple transposition of an appropriate matrix. The results obtained are competitive with a fully-fledged 4D-Var. The "ensemble" feature of those works give them similarity with the Ensemble Kalman Filter (see chapter *Ensemble Kalman Filter: Current Status and Potential*, Kalnay).

Both those works suggest that it could be possible to achieve substantial numerical gain, without significant degradation of the final results (and even maybe without the use of an adjoint), by restricting the control variable to an appropriate subspace of the whole state space.

All the algorithms that have been described above are based on the minimization of an objective function of the general form given by Eqs. 10, 18 or 29, which is quadratic in terms of the data-minus-unknown differences, with weights equal to the inverse of the covariance matrices of the corresponding errors. Equations 10 and 18 correspond to least-variance statistical linear estimation, while Eq. 29 corresponds to an extension to weakly non-linear situations. Other forms for the objective function have also been considered. In particular, Fletcher and Zupanski (2006) and Fletcher (2007), following a general Bayesian approach, propose to maximize the conditional probability density function for the state of the flow, given the data. In the case of linear data operators and Gaussian errors, this leads to minimization of an objective function of form given by Eq. 10. Those authors consider the case of lognormal distributions, which are more appropriate for bounded variables such as humidity. This leads to a significantly different form for the objective function.

6 More on the adjoint method

The adjoint method has been demonstrated above in the particular case of the objective function given by Eq. 29. It is actually very general, and defines a systematic approach for computing the (exact) gradient of a differentiable scalar function with respect to its arguments. Although this may not be obvious from the above developments, the adjoint method consists in a systematic use of the chain rule for differentiation of a compound function. Proceeding backward through the original sequence of computations, it recursively computes the partial derivatives of the scalar function under consideration with respect to the variables in those computations (see, *e.g.*, Talagrand 2003). As such, the adjoint method can be used not only for optimization purposes, as in variational assimilation, but (actually more simply) for determination of gradients as such, and for sensitivity studies.

The advantages and disadvantages of variational assimilation will be further discussed in the Conclusions below (Section 7). But its major disadvantage (at least for variational assimilation as it exists at present) is probably the need for developing the adjoint code which performs computations in Eq. 35. Not only must the adjoint code be developed, but it must be carefully validated, since experience shows that even minor errors in the computed gradient can significantly degrade the efficiency of the minimization (if not totally inhibit it). In addition, NWP models are constantly modified, and the corresponding modifications must be made on the adjoint code. Writing the adjoint of a code at the same time as the direct code involves only a rather small amount of additional work (10% or 20%). But developing the adjoint of an already existing code can require a substantial amount of work, and can be a very tedious and time-consuming task. On the other hand, the fact that adjoint computation is in essence a systematic use of the chain rule for differentiation leads to perfectly defined “adjoint” coding rules, which make the development of an

adjoint code, if lengthy and tedious, at least totally straightforward. These rules are described in, *e.g.*, Talagrand (1991), Giering and Kaminski (1998) or Kalnay (2002).

Those same rules are at the basis of "adjoint compilers", *i.e.*, software pieces that are designed to automatically develop the adjoint of a given code (see, *e.g.*, http://www.fastopt.de/; Hascoët and Pascual 2004). The adjoint of a particular piece of code is independent of the rest of the code, and automating the derivation of the adjoint instructions for a sequence of coding instructions, which is a purely local operation, is relatively easy. Other aspects, such as the choice and management of non-linear variables to be kept in memory from the direct integration, or to be recomputed in the course of the adjoint integration, require a global view of the code, and are more difficult to automate. For that reason, the use of these software pieces still requires experience of adjoint coding as well as some preparatory work, but they are nevertheless extremely useful, and very substantially reduce the amount of time and work necessary for developing the adjoint of an atmospheric or oceanic circulation model.

The adjoint approach is used in assimilation of meteorological and oceanographical observations for numerically solving, through an iterative minimization process, an optimization problem. Now, as said above, what the adjoint equations really do is simply compute the gradient of one scalar output of a numerical process with respect to (potentially all) the input parameters of that process. As such, the adjoint approach can be used for sensitivity studies of outputs with respect to inputs, independently of any optimization or minimization. It will be useful to use the adjoint approach when the number of output parameters whose sensitivity is sought is smaller than the number of input parameters with respect to which the sensitivity is sought (in the inverse case, direct perturbation of the input parameters will be more economical).

Actually, the first proponents of the use of the adjoint approach in meteorology and oceanography had primarily sensitivity studies in mind (Marchuk 1974; Hall *et al.* 1982). Adjoint models have been used to perform sensitivity studies of many different kinds: sensitivity of the atmospheric flow with respect to initial or lateral boundary conditions (Errico and Vukicevic 1992; Rabier *et al.* 1992; Gustafsson *et al.* 1998); sensitivity of the global oceanic circulation to parameters (Marotzke *et al.* 1999); sensitivity of biogeochemical processes (Waelbroeck and Louis 1995); and sensitivity of atmospheric chemical processes (Zhang *et al.* 1998). See also the special issue of *Meteorologische Zeitschrift* (Ehrendorfer and Errico 2007) devoted to Adjoint Applications in Dynamic Meteorology. Two specific types of applications are worthy of particular mention. The first one has to do with the identification, for a particular situation, of the unstable components of the flow. In its simplest form, this amounts to determining the so-called *singular vectors* of the flow, *i.e.*, the perturbations that amplify most rapidly, over a period of time, in the tangent linear approximation (Lacarra and Talagrand 1988; Farrell 1989; Urban 1993). This has been extended by Mu and colleagues (Mu 2000; Mu *et al.* 2003) to *Non-Linear Singular Vectors* (*NLSV*s), *i.e.*, perturbations that amplify most rapidly in the full non-linear evolution. A condition must then be imposed on the initial amplitude of the perturbation, which leads to a (technically more difficult to solve) constrained optimization problem. Both linear and non-linear singular vectors allow accurate diagnostic and analysis of instability (Moore and Farrell 1993; Mu and Zhang 2006;

Rivière *et al*. 2008). A related, but more specific, application is the identification of the components of the flow to which a particular feature of the future evolution of the flow (such as, for instance, the deepening of a depression) is most sensitive. This allows one to "target" observations in order to optimize the prediction of the feature under consideration. This has been implemented successfully on the occasion of specific campaigns (see, *e.g.*, Langland *et al*. 1999; Bergot and Doerenbecher 2002). Observation targeting through adjoint methods is further discussed in Buizza *et al*. (2007). Another, potentially very promising, application of the adjoint method is the determination of the sensitivity of analysed and predicted fields to observations. It is then the adjoint of the whole assimilation and prediction process, and not only of the assimilating model, that has to be used (Langland and Baker 2004). This has led to very useful diagnostics of the value and usefulness of various types of observations (*Langland, Cardinali, pers. comm.*).

7 Conclusion

Variational assimilation has now become a basic tool of numerical meteorology and oceanography, and a major component of operational NWP in several major meteorological services. Together with the Ensemble Kalman Filter (see chapter *Ensemble Kalman Filter: Current Status and Potential*, Kalnay), it is one of the two most advanced and powerful assimilation methods. The specific advantages of variational assimilation are rather obvious. It is very versatile and flexible, and allows for easy introduction of a new type of observation in an assimilation system. It suffices to specify the corresponding observation operator and the first- and second-order statistical moments of the associated error. It automatically propagates information both forward and backward in time, and makes it easy to take into account temporal correlation between errors (either observation or model errors). To the author's knowledge, this last possibility has been used so far on only one occasion, for taking into account temporally correlated errors in high frequency observations of surface pressure (Järvinen *et al*. 1999). But it can be extremely useful, especially for the treatment of model error and of the associated temporal correlation (time will presumably come when this will be necessary).

Variational assimilation is costly in that it requires the development, validation and maintenance of the adjoint of the assimilating model, as well as of the various observation operators. This is a time-consuming task. However, owing to the gain in experience and expertise, and to the continuous improvement of adjoint compilers, that task progressively becomes easier and easier. And, as discussed in the previous section, adjoints, once they are available, can be used for many other applications than assimilation, and in particular to powerful diagnostic studies.

Assimilation of meteorological and oceanographical observations may be at a turning point. It seems that the limits of what can be obtained from statistical linear estimation (*i.e.*, from Eq. 7 and its various generalizations to weakly non-linear situations) are being reached. The only exception is likely Quasi-Static Variational Assimilation, discussed in Section 5, which is based on minimization of objective functions of form given by Eq. 29, but whose limits have not been identified. Statistical linear estimation is at the basis of variational assimilation and of the

"Kalman" component of the Ensemble Kalman filter. It can legitimately be said that the ultimate purpose of assimilation is to achieve Bayesian estimation, *i.e.*, to determine the conditional probability distribution for the state of the atmosphere (or the ocean), given all the relevant available information. In view of the large dimension of the state of the atmosphere, the only possible way to describe the conditional probability distribution seems to be through an ensemble of points in state space, as indeed the Ensemble Kalman filter already does. A basic question is then to determine whether it is possible to develop methods for ensemble variational assimilation, which would produce a Bayesian ensemble, while retaining the specific advantages of variational assimilation, namely easy propagation of information both forward and backward in time, and possibility to easily take error temporal correlations into account. Some results suggest that this should be possible.

References

Auroux, D. and J. Blum, 2002. A dual data assimilation method for a layered quasi-geostrophic ocean model. *RACSAM*, **96**, 315-320.

Auroux, D. and J. Blum, 2004. *Data assimilation methods for an oceanographic problem*. In Multidisciplinary Methods for Analysis, Optimization and Control of Complex Systems, Mathematics in Industry, V. Capasso and J. Periaux, Eds , **6,** 179-194, Springer-Verlag, Berlin.

Bergot, T. and A. Doerenbecher, 2002. A study on the optimization of the deployment of targeted observations using adjoint-based methods. *Q. J. R. Meteorol. Soc*., **128**, 1689-1712.

Bonnans, J.-F., J.-C. Gilbert, C. Lemaréchal and C. Sagatizabal, 2003. *Numerical Optimization – Theoretical and Practical Aspects*. Springer-Verlag, Berlin, 485 pp.

Buizza, R., C. Cardinali, G. Kelly and J.-N. Thépaut, 2007. The value of observations. II: The value of observations located in singular-vector-based target areas. *Q. J. R. Meteorol. Soc*., **133**, 1817-1832.

Carrassi, A., A. Trevisan, L. Descamps, O. Talagrand and F. Uboldi, 2008. Controlling instabilities along a 3DVar analysis cycle by assimilating in the unstable subspace: a comparison with the EnKF. *Non. Proc. Geophys.* **15**, 503-521.

Courtier, P., 1997. Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc*., **123**, 2449-2461.

Courtier, P, and O. Talagrand, 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Q. J. R. Meteorol. Soc.* **113**, 1329-1347.

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

Derber, J., 1989. A Variational Continuous Assimilation Technique, *Mon. Weather Rev*., **117**, 2437-2446.

Egbert, G.D., A.F. Bennett and M.G.C. Foreman, 1994.Topex/Poseidon tides estimated using a global inverse model. *J. Geophys. Res*., **99**, 24,821-24,852.

Ehrendorfer, M. and R. M. Errico, Eds, 2007. *Meteorologische Zeitschrift*, **16**, no 6, 591-818.

Elbern, H., A. Strunk, H. Schmidt and O. Talagrand, 2007. Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmos. Chem. Phys*., **7**, 3749-3769

Errera, Q. and D. Fonteyn, 2001. Four-dimensional variational chemical assimilation of CRISTA stratospheric measurements. *J. Geophys. Res*., **106**, 12,253-12,265.

Errico, R. M. and T., Vukisevic, 1992. Sensitivity Analysis Using an Adjoint of the PSU-NCAR Mesoscale Model. *Mon. Weather Rev.* **120**, 1644-1660.

Farrell, B. F., 1989. Optimal Excitation of Baroclinic Waves. *J. Atmos. Sci.* **46**, 1193-1206.

Fisher, M. and D.J. Lary, 1995. Lagrangian four-dimensional variational data assimilation of chemical species. *Q. J. R. Meteorol. Soc*., **121**, 1681-1704.

Fletcher, S. J. and M. Zupanski, 2006. A hybrid Normal and lognormal distribution for data assimilation. *Atmos. Sci. Lett*., **7**, 43-46.

Fletcher, S. J., 2007. Implications and impacts of transforming lognormal variables into normal variables in VAR. *Meteorologische Zeitschrift*, **16**, 755-765.

Fournier, A., C. Eymin and T. Alboussière, 2007. A case for variational geomagnetic data assimilation: insights from a one-dimensional, nonlinear, and sparsely observed MHD system. *Non. Proc. Geophys*., **14**, 163-180.

Giering, R. and T. Kaminski, 1998. Recipes for Adjoint Code Construction. *Trans. Math. Software*, **24**, 437-474.

Gustafsson, N., E. Källen and S. Thorsteinsson, 1998. Sensitivity of forecast errors to initial and lateral boundary conditions. *Tellus* **50A**, 167-185.

Hall, M. C. G., D. G. Cacuci and M. E. Schlesinger, 1982. Sensitivity analysis of a radiative-convective model by the adjoint method. *J. Atmos. Sci*. **39**, 2038-2050.

Hascoët, L. and V. Pascual, 2004. *TAPENADE 2.1 user's guide*, available at the address http://www.inria.fr/rrrt/rt-0300.html.

Hoffman, R. N., 1986. A four-dimensional analysis exactly satisfying equations of motion. *Mon. Weather Rev*., **114**, 388-397.

Järvinen, H., E. Andersson and F. Bouttier, 1999. Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus*, **51A**, 469-488.

Kalnay, E., 2002. *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, United Kingdom, 341 pp.

Klinker, E., F. Rabier, G. Kelly and J.-F. Mahfouf, 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.

Lacarra, J.-F., and O. Talagrand, 1988. Short-range evolution of small perturbations in a barotropic model. *Tellus*, **40A**, 81-95.

Lahoz, W.A., A.J. Geer, S. Bekki, N. Bormann, S. Ceccherini, H. Elbern, Q. Errera, H.J. Eskes, D. Fonteyn, D.R. Jackson, B. Khattatov, M. Marchand, S. Massart, V.-H. Peuch, S. Rharmili, M. Ridolfi, A. Segers, O. Talagrand, H.E. Thornton, A.F. Vik and T. von Clarmann, 2007. The Assimilation of Envisat data (ASSET) project. *Atmos. Chem. Phys*., **7**, 1773-1796.

Langland, R.H., and N. L. Baker, 2004. Estimation of observation impact using the NRL variational data assimilation adjoint system. *Tellus*, **56A**, 189-201.

Langland, R. H., R. Gelaro, G. D. Rohaly and M. A. Shapiro, 1999. Targeted observations in FASTEX: Adjoint-based targeting procedures and data impact experiments in IOP17 and IOP18. *Q. J. R. Meteorol. Soc*., **125**, 3241-3270.

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

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

Lions, J.-L., 1971. *Optimal control of systems governed by partial differential equations* (translated from the French). Springer-Verlag, Berlin.

Liu, C., Q. Xiao and B. Wang, 2008. An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part I: Technical Formulation and Preliminary Test. *Mon. Weather Rev*.,**136**, 3363-3373.

Lorenc, A.C. and F. Rawlins, 2005. Why does 4D-Var beat 3D-Var? *Q. J. R. Meteorol. Soc*., **131**, 3247–3257.

Lorenz, E.N., 1963. Deterministic Nonperiodic Flow. *J. Atmos. Sci*., **20**, 130-141.

Louvel, S., 2001. Implementation of a dual variational algorithm for assimilation of synthetic altimeter data in the oceanic primitive equation model MICOM. *J. Geophys. Res*., **106**, 9199-9212.

Marchuk, G.I., 1974. *Numerical solution of the problems of dynamics of the atmosphere and the ocean* (in Russian). Gidrometeoizdat, Leningrad.

Marotzke, J., R. Giering, K. Q. Zhang, D. Stammer, C. Hill and T. Lee, 1999. Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity. *J. Geophys. Res.*, **104**, 29,529-29,547.

Moore, A.M., and B.F. Farrell, 1993. Rapid Perturbation Growth on Spatially and Temporally Varying Oceanic Flows Determined Using an Adjoint Method: Application to the Gulf Stream. *J. Phys. Oceanogr.*, **23**, 1682-1702.

Mu M., 2000. Nonlinear singular vectors and nonlinear singular values. *Science in China* (ser. D) **43**, 375-385.

Mu, M., W. S. Duan and B. Wang, 2003. Conditional nonlinear optimal perturbation and its applications. *Non. Proc. Geophys*., **10**, 493-501.

Mu, M. and Z. Zhang, 2006. Conditional Nonlinear Optimal Perturbations of a Two-Dimensional Quasigeostrophic Model. *J. Atmos. Sci*., **63**, 1587-1604.

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

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

Rabier, F., P. Courtier and O. Talagrand, 1992. An Application of Adjoint Models to Sensitivity Analysis. *Beitr. Phys. Atmosph.* **65**, 177-192.

Reichle, R.H., 2000. *Variational Assimilation of Remote Sensing Data for land Surface Hydrologic Applications*, Doctoral thesis, Massachusetts Institute of Technology, Cambridge, Mass., USA, 192 pp.

Ricci, S., A.T. Weaver, J. Vialard and P. Rogel, 2005. Incorporating state-dependent Temperature-Salinity constraints in the background error covariance of variational ocean data assimilation. *Mon. Weather Rev*., **133**, 317-338.

Rivière, O., G. Lapeyre and O. Talagrand, 2008. Nonlinear Generalization of Singular Vectors: Behavior in a Baroclinic Unstable Flow. *J. Atmos. Sci*., **65**, 1896-1911.

Sasaki, Y., 1970a. Some basic formalisms in numerical variational analysis. *Mon. Weather Rev*., **98**, 875-883.

Sasaki, Y., 1970b. Numerical variational analysis formulated from the constraints as determined by longwave equations and a low-pass filter. *Mon. Weather Rev*., **98**, 884-898.

Sasaki, Y., 1970c. Numerical variational analysis formulated with weak constraint and application to surface analysis of severe storm gust. *Mon. Weather Rev*., **98**, 899–910.

Sun, Z., A. Tangborn and W. Kuang, 2007. Data assimilation in a sparsely observed one-dimensional modeled MHD system. *Non. Proc. Geophys*., **14**, 181-192.

Swanson, K., R. Vautard and C. Pires, 1998. Four-dimensional variational assimilation and predictability in a quasi-geostrophic model. *Tellus*, **50A**, 369-390.

Talagrand, O., 1991. The Use of Adjoint Equations in Numerical Modeling of the Atmospheric Circulation. In *Automatic Differentiation of Algorithms : Theory, Implementation, and Application*, A. Griewank and G. F. Corliss, eds, Society for Industrial and Applied Mathematics, Philadelphia, United States, 169-180.

Talagrand, O., 2003. Variational Assimilation. *Adjoint Equations*, in R. Swinbank, V. Shutyaev and W. A. Lahoz (editors), Data Assimilation for the Earth System, NATO Science Series: IV: Earth and Environmental Sciences Sciences 26, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 37-53.

Talagrand, O. and P. Courtier, 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Q. J. R. Meteorol. Soc*., **113**, 1311-1328.
Thacker, W.C. and R.B. Long, 1988. Fitting dynamics to data. *J. Geophys. Res*., **93**, 1227-1240.
Thépaut, J.-N. and P. Courtier, 1991. Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Q. J. R. Meteorol. Soc*., **117**, 1225-1254.
Thépaut, J.-N., R.N. Hoffman and P. Courtier, 1993. Interactions of Dynamics and Observations in a Four-Dimensional Variational Assimilation. *Mon. Weather Rev*., **121**, 3393-3414.
Trémolet, Y., 2006. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc*., **132**, 2483-2504.
Trémolet, Y., 2007. Incremental 4D-Var convergence study. *Tellus*, **59A**, 706-718.
Tromp, J., C. Tape and Q. Liu, 2005. Seismic tomography, adjoint methods, time reversal, and banana-donut kernels. *Geophys. J. Int*., **160**, 195-216.
Urban, B., 1993. A method to determine the theoretical maximum error growth in atmospheric models. *Tellus* **45A**, 270-280.
Vialard, J., A.T. Weaver, D.L.T. Anderson and P. Delecluse, 2003. Three- and Four-Dimensional Variational Assimilation with a General Circulation Model of the Tropical Pacific Ocean. Part II: Physical Validation. *Mon. Weather Rev*., **131**, 1379-1395.
Vossepoel, F., A.T. Weaver, J. Vialard and P. Delecluse, 2004. Adjustment of near-equatorial wind stress with four-dimensional variational data assimilation in a model of the Pacific Ocean. *Mon. Weather Rev.*, **132**, 2070-2083.
Waelbroeck, C. and J.-F. Louis, 1995, Sensitivity analysis of a model of CO2 exchange in tundra ecosystems by the adjoint method. *J. Geophys. Res.*, **100**, 2801-2816.
Weaver, A.T., and D.L.T. Anderson, 1997. Variational Assimilation of Altimeter Data in a Multilayer Model of the Tropical Pacific Ocean. *J. Phys. Oceanogr*., **27**, 664-682.
Zhang, Y., C. H. Bischof, R. C. Easter and P.-T. Wu, 1998. Sensitivity analysis of a mixed-phase chemical mechanism using automatic differentiation. *J. Geophys. Res.*, **103**, 18,953-18,979.
Zupanski, D., 1997. A General Weak Constraint Applicable to Operational 4DVAR Data Assimilation Systems. *Mon. Weather Rev*., **125**, 2274-2292.