Variational Data Assimilation : Optimization and Optimal Control

François-Xavier Le Dimet
Lab.Jean-Kuntzman
Université Grenoble-Alpes,
BP 53, 38041 GRENOBLE cedex 9, France

Ionel M. Navon
Department of Scientific Computing
Florida State University
483 Dirac Science Library, Tallahassee, Florida, USA

Răzvan Ştefănescu
Department of Mathematics
North Carolina State University
Box 8205, Raleigh, NC, USA

1 Introduction

In the last few years due to a constant increase in the need for more precise forecasting and nowcasting, several important developments have taken place in meteorology directed mainly in two different directions: modeling at either large scale or at smaller scales to include an ever increasing number of physical processes and parametrization of subgrid phenomena and adding new sources of data such as satellite data, radar, profilers, and other remote sensing devices. While this led to an abundance of widely distributed data it also created difficulties since most of the information is heterogeneous in space or time and present different levels of accuracy.

Therefore, a cardinal problem is how to link together the model and the data. This problem induces several questions: (i) How to retrieve meteorological fields from sparse and/or noisy data in such a way that the retrieved fields are in agreement with the general behavior of the atmosphere? (Data Analysis); (ii) How to insert pointwise data in a numerical forecasting model? This information is continuous in time, but localized in space (satellite data for instance)? (Data assimilation problem) (iii) How to validate or calibrate a model (or to invalidate it) from observational data? The dual question in this case being how to validate (invalidate) observed data when the behavior of the atmosphere is predicted by a numerical weather prediction model.

For these questions a global approach can be defined by using a variational formalism.

1.1 Historical perspective

Numerical weather prediction has started in the '50 with Charney, Fjörtoft, and Von Neumann [15] when an atmospheric forecast was obtained after an integration of a mathematical model staring from an initial condition. In the early years of numerical meteorology this initial condition was determined by optimal interpolation i.e. an interpolation weighted by statistics on the atmospheric fields. When Yoshi Sasaki arrived in Oklahoma, a state frequently devastated by tornadoes, he worked on mesoscale meteorology; the methods of large scale meteorology can not be directly applied at this scale due to the lack of observations and of statistics on rare events. Interpolation methods tend to regularize the fields at mesoscale level where the identification of discontinuities such as squall lines is of great importance. To retrieve these fields, Sasaki, in his pioneering basic papers has proposed to use the mathematical model itself as a constraint in order to retrieve the atmospheric fields: variational methods applied to meteorology were born. At the same period that the Optimal Control methods for Partial Differential Equation were developed, Lions [52] pioneered the theoretical basic support of these methods, consisting of the Calculus of Variations in adequate functional spaces.

Optimal Control makes the difference between a "State Variable" and a "Control Variable" and permits to alleviate the difficulties linked to the determination of boundary and/or initial conditions in numerical models. The '80 and '90 have witnessed important improvements in the computational tools with advent of high performance and parallel supercomputers, the development of more precise numerical weather prediction models along with a better understanding of the underlying atmospheric physics and the coverage of networks of observations especially with launching of dedicated satellites. A consequence was that of rendering obsolete the optimal interpolation methods, mainly because the retrieved fields where not in agreement neither with the physics of the models nor with their dynamics. At the end of this period variational methods were successfully introduced in many national operational centers.

1.2 Variational Methods in Meteorology: A Perspective

There are two main approaches employed when modeling a system described by a state variable, X. The first approach consists of finding a set of equations F such that X is the unique solution of the state equation

$$F(X) = 0. \tag{1.1}$$

In most cases system F must have as many equations as X has components in order to possess a unique solution – this is the problem of closure. In meteorology this problem has often been solved by using various artifacts such as adding supplementary equations. The second approach to the problem of closure is the variational one consisting in finding X as the solution of a problem of optimization i.e. by finding the extremum of some known functional J.

Such an approach was proposed in theoretical mechanics more than 250 years ago by Euler [24, 23] and by Lagrange [44, 43]. In the domain of numerical analysis Sobolev or Galerkin type methods are also based upon variational principles Ritz [71],Galerkin [28] .

In meteorology, using the most general terms, we assume the state of the atmosphere to be described by the set of equations (1.1).

As mentioned, if this system possesses fewer equations than unknowns, the system is said to be non-closed. However, one can still close it by introducing a variational approach.

If X_{obs} denotes an observation of a meteorological field, we will choose from among all the solutions of the system $F(W) = 0$, the solution closest to the observation X_{obs}. The resulting solution will be the optimal solution. In this manner a connection is established between the data and the observations.

In meteorology, the first application of variational methods has been pioneered by Sasaki, Gu, and Yan [79] and Sasaki [74]. Later on, Washington and Duquet [95], Stephens [84, 85] and Sasaki [75, 77, 76, 76] have given a great impetus towards the development of variational methods in meteorology.

In a series of basic papers Sasaki [75, 77, 76, 76] generalized the application of variational methods in meteorology to include time variations and dynamical equations in order to filter high-frequency noise and to obtain dynamically acceptable initial values in data void areas. In all these approaches, the Euler-Lagrange equations were used to calculate the optimal X.

Numerous other manuscripts applying these ideas appeared in the meteorological literature during the 70's using the variational formulation. In parallel with the introduction of variational methods in meteorology, starting in the 60's and 70's, mathematicians in coordination with other scientific disciplines have achieved significant advances in optimization theory and optimal control, both from the theoretical viewpoint as well as from the computational one. In particular significant advances have been achieved in the development of optimization algorithms (Gill, Murray, and Wright [31], Fletcher [27], Powell [68], Bertsekas [4], Lugenberger [56] to cite but a few).

Optimal control methods have been introduced by Pontryagin, Boltyanskii, Gamkrelidze, and Mishchenko [67], and they have been generalized for systems governed by partial differential equations (Lions [52]).

The application of an optimal control theory to meteorological problems has for the first time supplied the correct framework for a unified approach to analysis, data assimilation and initialization for meteorological problems.

Other techniques strongly related to variational and optimization theory, such as optimum interpolation, Kalman-Bucy filtering (Ghil, Cohn, Tavantzis, Bube, and Isaacson [30]), smoothing splines (Wahba [91, 93, 92]), Krieging, generalized cross-validation (GCV) (Wahba and Wendelberger [94]) have also emerged. For a unified approach Lorenc [54]) manuscript could be consulted.

1.2.1 Variational methods: for which purposes?

The first applications of variational methods targeted the objective analysis of meteorological fields, i.e. to retrieve fields from pointwise distributed data in space. In most of the important meteorological situations the temporal evolution of the fields is crucial, therefore, some attempts were carried out towards extending variational analysis to dynamic analysis. Introducing sparsity of data in time using variational tools has led to 4-D data assimilation for numerical weather prediction models.

To perform a forecast a meteorological model requires an initial condition. This initial condition must be as close as possible to the observations while remaining compatible with the model. The problem of initialization may be stated as a variational problem and solved in this way.

A general formalism of variational problems has to deal with observations but these observations may not necessarily be physical ones. For instance they may result out of a numerical model (output of a numerical model). Furthermore, the constraints imposed upon the analysis may have no physical origin and could only have been introduced for numerical purposes.

Many applications were carried out in similar situations as mentioned above resulting in a global approach of variational methods, such as for instance enforcing conservation of integral invariants in numerical models (Navon [62], Navon and De Villiers [63]), or design of discretization schemes (Sasaki [78]). A major difficulty for the classical approach to variational methods for meteorologically significant problems, in particular for those where dynamics play a prominent part, is the fact that the size of the discrete problem to be solved is prohibitive.

A way to circumvent this difficulty is to introduce optimal control methods permitting a significant reduction of the problem size. These techniques, upon which we will expand in a later section, introduce the adjoint of the numerical model. Knowledge of the adjoint of the model turns out to be particularly useful, because it can be applied towards a sensitivity analysis (Hall, Cacuci, and Schlesinger [36], Cacuci and Hall [13]) or for environmental studies such as the estimation of the impact of industrial pollution upon the environment (see Marchuk [58]).

In this review paper we will present the most important contributions concerning applications of variational methods using the general formalism of mathematical programming.

1.3 Variational Methods in Meteorology: The Optimization Theory View Point

Numerical weather prediction is based on the integration of a dynamic system of partial differential equations modeling the behavior of the atmosphere.

From a mathematical view point this approach is equivalent to the classical Cauchy problem. Therefore discrete initial conditions describing the state of the atmosphere have to be provided prior to the integration.

In order to retrieve a complete description of the atmosphere one can add information to the raw data using one of the following families of several methods: (i) Perform a simple interpolation, no information is added to the data. This procedure is purely algorithmic ; (ii) Add as information the statistical structure of the fields and use an optimal interpolation type method. Unfortunately this information is not always available or may be inadequate for instance as is the case with a paroxysmal event; (iii) variational method.

Variational methods are based on the fact that a given meteorological observation has not an intrinsic credibility. The same measurement of wind, to give just an example, may be used to study the flow around a hill, or may be inserted in a mesoscale model, or may be used in a global model of atmospheric circulation. According to the particular framework where the data will be used, variable trust will be attributed to the same data.

Variational methods try to achieve a best fit, with respect to some '*a priori*' criterion, of data to

a model by placing the data into the most adequate framework where it should be used, and permits us to link the data and the model.

In the first part of the paper we will show how variational methods can be defined and which are the ingredients necessary to build a variational method, all this in the perspective of the surveyed accumulated work. Then we will show how to solve related variational problems in the framework of a systematic classification of the reviewed work. This classification will permit us to review different variational methods as well as the context in which they were performed.

The last section will be devoted to future developments and potential applications of variational methods in meteorology.

2 Ingredients of a Variational Method

2.1 Definition of a Variational Method

In the most condensed way a variational method may be defined as a search, amongst all the possible solutions of a model, of the solution closest to a given observation. Therefore a variational method will be defined by the following ingredients:

- i) An atmospheric variable X, describing the state of the atmosphere.
- ii)A model which may be mathematically written as:

$$\frac{dX}{dt} + A(X) = 0 \tag{2.1}$$

  where A is a linear or non-linear operator.

  We suppose that system (2.1) is not closed by which we mean that in order to obtain an unique solution to (2.1) some additional information has to be provided.

- iii) A control variable U that may comprise the initial conditions, boundary conditions, or both, the vector X itself or a part of it. Once U is defined – a unique solution $X(U)$ of (2.1) will be associated with it. The vector control variable U must belong to some set of admissible control U_{ad}. The definition of U_{ad} may include physical information which can be stated in the form of inequalities.
- iv)An observation X_{obs} of the meteorological fields.
- v)J, a cost function measuring the difference between a solution of (2.1) associated with U and the observations X_{obs}.

The variational problem is determined in terms of these last five items and it can be stated as following problem:

$$\text{Determine } U^* \text{ which belongs to } U_{\text{ad}} \text{ and minimizes the cost function } J. \tag{2.2}$$

The second stage of the solution of the variational problem will be to determine, or at least to approximate U^* (and therefore the optimal associated state of the atmosphere $X(U^*)$).

In order to achieve this, we first have to set up an optimality condition and then to perform an algorithm for solving problem (2.2).

2.1.1 The optimality Condition

A general optimality condition is given by the variational inequality (see Lions (1968))

$$(\nabla J(U^*), V - U^*) \geq 0 \text{ for all } V \text{ belonging to } U_{\mathrm{ad}}, \tag{2.3}$$

where ∇J is the gradient of the functional J with respect to the variable U.

In the case where U_{ad} has the structure of a linear space, variational inequality (2.3) is reduced to the equality

$$\nabla J(U^*) = 0. \tag{2.4}$$

2.1.2 The Algorithm of Solution

As stated above – variational problems are problems of optimization with or without constraints. There exist standard procedures (Le Dimet and Talagrand [48] ,Navon and Legler [64]) to solve them.

A common requirement of these procedures is the need to explicitly supply the gradient of J with respect to U to the code.

Moreover, the basic problem to be solved is always a problem of unconstrained minimization for which the method of conjugate gradient may be used (see Navon and Legler [64]).

3 Variational Analysis

Basically, the problem of retrieving meteorological fields X from observations X_{obs}, in such a way that X verify some model:

$$F(X) = 0 \tag{3.1}$$

and are as close as possible, in the sense of a given functional J, to the observations X_{obs}, is a problem of optimization with constraints.

Sasaki (1970) in historical paper has introduced two formalisms. The *weak constraint formalism* consists in minimizing without constraint the functional J defined by

$$J_1(X) = J(X) + K\|F(X)\|^2. \tag{3.2}$$

It is easily seen that for large values of K, $F(X)$ has to be small for minimizing J_1, therefore, for a specified value of K, constraint (3.1) is only approximately verified. In what follows K is a generic constant used as a coefficient of a weak constraint. This is justified by the fact that equation (3.1) is not a perfect representation for the atmosphere and therefore should not be satisfied with a greater precision than its own accuracy.

The optimal condition, which in the Euler-Lagrange equation gives the optimal analyzed field X^*, is the solution of the equation

$$\nabla J_1(X^*) = \nabla J(X^*) + 2K \cdot \ F'(X^*) \cdot F(X^*) = 0. \tag{3.3}$$

In this equation ∇J_1(respectively ∇J) is the gradient of J_1(respectively J) with respect to X, while F' is the Jacobian matrix of F. No standard method exists for solving (3.3). As such a method of solution has to be chosen in agreement with the particular expressions for J and F. In the majority of cases, and even always when F is non-linear, an iterative algorithm has to be carried out.

The second formalism is called strong constrained where the model has to be exactly verified. In consequence we have to deal with a problem of optimization under constraint and the approach by optimal control permits to alleviate to some extent the difficulties linked to this formalism.

4 Optimal Control Techniques

4.1 General Results

Optimal control methods for distributed systems have been extensively studied and applied in many areas such as mechanics, economics, engineering, oceanography, etc.

Due to the fact that the formalism of optimal control problems includes the minimization of a functional, the cost function, they are variational methods and as such their numerical solution requires the computation of the gradient of the cost functional with respect to the state variable.

In many cases, the cost function is only an implicit function of the state variable which may be an initial condition or a boundary condition. Therefore, more sophisticated mathematical techniques must be used for estimating the gradient. One such particular method, the adjoint model technique, was specially developed for this purpose. A difficulty of this approach is the necessity to write well-posed problems and to carefully specify the functional framework of the variational problem.

We assume that the state of the atmosphere is described by a variable X belonging to some Hilbert space $\mathcal{H}$ (of finite or infinite dimension) and by a model written as

$$F(X) = 0 \tag{4.1}$$

We suppose that X may be split into two parts, Y and U, each part belonging to the Hilbert spaces $\mathcal{Y}$ and $\mathcal{U}$, respectively.

Therefore, (4.1) may be written as

$$F(Y,U) = 0 \tag{4.2}$$

where U is the control variable, chosen in such a way that for each given U, equation (4.2) has a unique solution $Y(U)$.

In this way we may define G by

$$G : \mathcal{Y} \to \mathcal{U} \tag{4.3}$$

for each U belonging to $\mathcal{U}$. Then

$$G(Y) = U \tag{4.4}$$

has a unique solution in $\mathcal{Y}$.

Furthermore, we will assume that for each Y belonging to $\mathcal{Y}$, $\frac{\partial F}{\partial Y}(Y)$ is an isomorphism from $\mathcal{Y}$ to $\mathcal{U}$.

Therefore, it is possible to define an inverse function Φ such that:

$$\begin{array}{rll} \Phi : & \mathcal{U} & \to \mathcal{Y} \\ & U & \to \Phi(U) = Y \end{array}$$

verifying :

$$\Phi(G(Y)) = Y$$
$$\Phi'(U) = \left[\frac{\partial F}{\partial Y}(\Phi(U))\right]^{-1}$$

Another Hilbert space has to be defined: the space of observations Θ in which an observation Z_{obs} is given. As pointed out, the observation is not necessarily a physical one, and it is not supposed to verify the equations of the model.

Let C be an operator from the space of the state variable to the space of observations; for each value of the control U we associate a state of the atmosphere $Y(U)$ and a model observation

$$Z(U) = C(Y(U)). \tag{4.5}$$

The cost function $J(U)$ is a measure of the distance between the model observation associated to the control U and the observation. It is defined by:

$$J(U) = \frac{1}{2} \left\| C(Y(U)) - Z_{obs} \right\|_{\Theta}^2 \tag{4.6}$$

Therefore, the problem is to determine the optimal control variable U^* defined by

$$U^* = \arg(MinJ(U)||u \in \mathcal{U}). \tag{4.7}$$

From a theoretical viewpoint, the system of optimality giving U^* is dependent upon the gradient of J with respect to U.

From a numerical viewpoint, U^* may be estimated by an iterative method starting from a first given U_0. In the same way, the numerical implementation of the iterative method requires the computation of the gradient of J with respect to U.

For deriving the gradient, a systematic method is the following:

- i)Let V be some variable belonging to $\mathcal{U}$; then the directional derivative of J in direction V will verify :

$$\begin{aligned} J'(U,V) = \nabla J(U)\cdot V &= \left(C'(Y)\cdot V, C(Y) - Z_{obs}\right)_\Theta \\ &= \langle C'(Y)V, \Lambda_\Theta\left(C(Y) - Z_{obs}\right)\rangle_{\Theta',\Theta} \end{aligned} \tag{4.8}$$

  where Λ_Θ is the canonical isomorphism between Θ and its dual space Θ', and $\langle\cdot,\cdot\rangle$ denotes the duality between Hilbert spaces.

- ii)Let R be a linear operator from $\mathcal{Y}$ to $\mathcal{U}$, we define its dual operator to be the operator R^* from $\mathcal{U}'$ to $\mathcal{Y}'$ defined by:

$$\left\langle R\cdot\mathcal{Y}, U'\right\rangle_{\mathcal{U}} = \langle Y, R^*\cdot U'\rangle_{\mathcal{Y}}$$

  Using the dual operator of C' in (4.8) gives:

$$\nabla J(U).V = \langle V, C'(Y)^*\Lambda_\Theta\left(C(Y) - Z_{obs}\right)\rangle_{\mathcal{U},\mathcal{U}'}$$

- iii) Let us now define the adjoint system by :

$$\left(\frac{\partial F}{\partial Y}\right)^* P = -C'(Y)^*\Lambda_\Theta\left(C(Y(U)) - Z_{obs}\right) \tag{4.9}$$

  Then :

$$\begin{aligned} \nabla J(U)\cdot V &= \langle V, (\frac{\partial F}{\partial Y})^*.P\rangle_{\mathcal{U},\mathcal{U}'} \\ \nabla J(Y)\cdot V &= \langle\frac{\partial F}{\partial Y}\cdot V, P\rangle_{\mathcal{Y},\mathcal{Y}'} \end{aligned} \tag{4.10}$$

J is a functional defined on the space $\mathcal{U}$, so its gradient belongs to the dual space $\mathcal{U}'$. Theoretically, it is always possible to identify a Hilbert space to its dual. However, in practical problems there exist inclusion relations between the spaces used here, and when a space has been identified to its dual, it is no longer possible to identify subspaces with their duals.

In the practical phase of optimal control methods we were always operating in finite-dimensional spaces where no such problems exist.

Therefore equation (4.10) permits us to compute the gradient of J, applied to the direction V by determining P, the adjoint variable, as the solution of the adjoint system (4.9).

From this abstract situation let us extract two more practical examples enabling us to see how the gradient is computed. For an initial condition problem we will consider the case where the control variable is the initial condition, while for a boundary value problem we will see how to compute the gradient when the control variable is the value on the boundary.

4.2 Control of the Initial Condition

After a spatial discretization, we will assume that the state of the atmosphere, modeled by a vector Θ is verifying for the time interval $[0, T]$ the equation:

$$\frac{d\Theta(t)}{dt} = H(\Theta(t)) \tag{4.11}$$

where $\Theta(t)$ belongs to a finite dimensional space.

With an initial condition $\Theta(0) = \mu$, equation (4.11) has a unique solution $\Theta(\mu, t)$.

For the sake of simplicity, we will assume that a continuous observation $\widetilde{\Theta}$, in time, is given on the time interval $[0, T]$. The distance between a solution of (4.11) and the observation is defined by

$$J(\mu) = \frac{1}{2} \int_0^T \left\| \Theta(\mu, t) - \widetilde{\Theta}(t) \right\|^2 dt \tag{4.12}$$

where $\| \cdot \|$ is the Euclidean norm in finite dimensional space. With respect to the general theory developed above the space of the state variable is the same as the space of the observations. In practice, the observations are pointwise in both space and in time, therefore, Dirac's measures have to be introduced in the definition of J.

The derivation of the gradient of J with respect to μ is obtained as follows:

Let ν be some element belonging to the space of the initial conditions. The directional derivative of Θ in direction ν is defined by :

$$\widehat{\Theta}(\mu, \nu) = \lim_{\alpha \to 0} \frac{\Theta[(\mu + \alpha), t] - \Theta(\mu, t)}{\alpha} \tag{4.13}$$

where $\widehat{\Theta}(\mu, \nu)$ is the solution of the differential system:

$$\begin{aligned} \frac{d\widehat{\Theta}(\mu, \nu)]}{dt} &= \frac{\partial H}{\partial \Theta}[\Theta(\mu, t)] \cdot \widehat{\Theta}(\mu, \nu) \\ \widehat{\Theta}(0) &= \nu \end{aligned} \tag{4.14}$$

obtained by writing (4.11) with initial condition μ, then with initial condition $\mu_\alpha \nu$ and by letting the scalar α tend to zero. In (4.14) the expression $\frac{\partial H}{\partial \Theta}$ denotes the Jacobian of H.

The directional derivative of J in direction ν is obtained by taking the derivative of (4.12) leading to:

$$J'(\mu,\nu) = \int_0^T \left(\widehat{\Theta}(\mu,\nu,t), \Theta(\mu,t) - \widetilde{\Theta}(t)\right) dt \tag{4.15}$$

Let ψ be the dual variable to Θ, ψ is defined as the solution of the adjoint system to (4.11) given by :

$$\begin{aligned} \frac{d\psi}{dt}(\mu,t) + \left[\frac{\partial H}{\partial \Theta}\Theta(\mu,t)\right]^T \cdot \psi(\mu,t) &= \left(\Theta(\mu,t) - \widetilde{\Theta}(t)\right) \\ \psi(T) &= 0 \end{aligned} \tag{4.16}$$

Let us write the scalar product of (4.15) with $\widehat{\Theta}$, then by integrating from 0 to T, we obtain:

$$J'(\mu,\nu) = \int_0^T \left(\frac{d\psi}{dt} + \left[\frac{\partial H}{\partial \Theta}\Theta(\mu,t)\right]^T \cdot \psi(\mu,t), \widehat{\Theta}(\mu,\nu,t)\right) dt)$$

The time derivative in (4.16) is integrated by parts and then by using (4.14) we obtain:

$$J'(\mu,\nu) = \nabla J(\mu)\cdot\nu = \psi(\mu,0)\cdot\nu \tag{4.17}$$

Therefore, the gradient of J is obtained as the value at time zero of the dual variable. The backward integration of the adjoint system from T to 0 permits us to estimate the gradient of the cost functional and to perform a descent-type method.

An important remark for potential applications of control methods is the fact that with a different cost function only the right hand side of (4.16) has to be changed. The main difficulty encountered for programming optimal control methods is to write the left hand side of (4.16). This one is independent of the cost function and is intrinsic for a given model. Once it has been written and derived it can be used for other purposes such as data assimilation, initialization, sensitivity analysis, etc.

4.3 Control of the Boundary

For the sake of simplicity, we will suppose that on a domain Ω, of boundary Γ, some field is verifying the Laplace equation

$$\Delta U = f. \tag{4.18}$$

Together with a boundary condition $U/\Gamma = V$, Laplace equation (4.18) has a unique solution, $U(V)$.

Let $\mathcal{T}$ be a set of points belonging to Ω, where some observations $\widetilde{U}$ of U are performed.

$$\mathcal{T} = \{Z_1, Z_2, \ldots, Z_N\}$$

The cost function is defined by

$$J(V) = \frac{1}{2}\sum_{i=1}^{N}\left(U\left(V, Z_i\right) - \widetilde{U}\left(Z_i\right)\right)^2, \tag{4.19}$$

while the directional derivative $\overline{U}$ of U in a direction H is the solution of

$$\begin{array}{ll} \Delta\overline{U}(H) & = 0 \\ \overline{U}(H)/\Gamma & = H. \end{array} \tag{4.20}$$

The directional derivative of J verifies

$$J'(V, H) = \sum_{i=1}^{N}\left(\overline{U}\left(Z_i\right), U\left(V, Z_i\right) - \widetilde{U}\left(Z_i\right)\right). \tag{4.21}$$

The adjoint system to (4.19) is introduced with P the dual variable to U.

$$\begin{array}{ll} \Delta P & = \sum_{i=1}^{N} U\left(V, Z_i\right) - \widetilde{U}\left(Z_i\right) \\ P/\Gamma & = 0 \end{array} \tag{4.22}$$

As above, (4.22) is multiplied by $\overline{U}\left(H, Z_i\right)$ integrated on Ω, and after an integration by parts we find

$$\nabla J(V) = \frac{\partial P}{\partial n}/\Gamma, \tag{4.23}$$

$\frac{\partial P}{\partial n}$ being the normal derivative of P on the boundary Γ. The estimation of the gradient for carrying out a descent-method requires the estimation of the gradient of J, which is obtained by solving the adjoint system (4.22).

Let us point out that this case is especially simple due to the fact that the Laplacian operator is self-adjoint. Therefore, a Laplace's equation solver may be used to solve both the direct and the adjoint problem.

This problem could have been solved using a classical variational formalism, for instance with a weak constraint formalism we would have to minimize the functional

$$J(U) = \frac{1}{2}\sum\left(U\left(Z_i\right) - \widetilde{U}\left(Z_i\right)\right)^2 + \frac{1}{C}\int_{\Omega}\left(\Delta U - f\right)^2 dy. \tag{4.24}$$

The Euler-Lagrange equation for (4.24) is a fourth order partial differential equation with complicated boundary conditions. From a numerical viewpoint the size of the discrete problem associated with (4.24) is equal to the number of grid points in the discrete point of view domain Ω.

By comparison, for the optimal control approach the dimension of the problem to be solved is only equal to the number of points on the discrete boundary. In this way we have obtained a significant reduction of the size of the problem.

5 Weak Constraints in Variational Data Assimilation

The canonical approach for variational data assimilation (VDA), based on Optimal Control, implicitly assumes that the model is without error. This is not true because of the physical errors due to approximation in the physics of the problem, for instance in the parametrization of non linear interactions and also in the physical processes and the mathematical error due to discretization of the equations and also to iterative processes carried out to solve non linear problems or subproblems.

To alleviate this problem Sasaki has introduced the concept of *weak constraint* permitting to have a model that is only approximately verified.

5.1 Three basic methods in Constrained Optimization

Let us consider the constrained optimization problem:

Minimize $J(X)$ subject to the constraint $G(X) = 0$, where X belongs to some space $\mathcal{X}$, J is a mapping from $\mathcal{X}$ to $\mathbb{R}$ and G is a mapping from $\mathcal{X}$ to some linear space $\mathcal{Y}$

Here we assume the differentiability of J and G. There are three basic algorithms to obtain a numerical solution to this problem:

5.1.1 Duality Methods

In this method we introduce a Lagrange multiplier Λ in the dual space $\mathcal{Y}$ of and the Lagrangian $\mathcal{L}$ defined by:

$$\mathcal{L}(X, \Lambda) = J(X) + (\Lambda, G(X)) \tag{5.1}$$

Then optimal solution of the constrained optimization problem is a saddle point of the Lagrangian and is characterized by:

$$\frac{\partial \mathcal{L}}{\partial X} = \nabla J + \left[\frac{\partial G}{\partial X}\right]^t .\Lambda = 0 \tag{5.2}$$

$$\frac{\partial \mathcal{L}}{\partial \Lambda} = G(X) = 0 \tag{5.3}$$

$$\tag{5.4}$$

In Sasaki's terminology this is the strong constraint formalism, it is worthwhile to point pout that X is the state variable of the problem, Sasaki doesn't make the difference between state variable and control variables that could be the initial condition and/or boundary conditions. The equation above are the Optimality System (O.S.) it can be solved by an iterative algorithm of the form :

$$X_{n+1} = X_n - \rho_n.D_n \tag{5.5}$$

$$\Lambda_{n+1} = \Lambda_n + \eta_n.W_n \tag{5.6}$$

Where D_n is a direction of descent, estimated from the gradient of J, X_{n+1} realizes the minimum of $\mathcal{L}$ along this direction. On the same token W_n is a direction of ascent and Λ_{n+1} realizes the

maximum of $\mathcal{L}$ along this direction. ρ_n and η_n are scalars. In practice some stopping criterion for the iterative algorithm has to be defined and therefore, at the end of the process, the constraint is not exactly satisfied and, by this way, an error is introduced, This error cannot be controlled. The Lagrange multiplier introduced in this method is nothing else than the adjoint variable used in the terminology of VDA problems stated as problems of Optimal Control. In practice the convergence of this type of algorithms is slow.

5.1.2 Penalty Methods

In this approach we define a *penalized functional* J_ϵ by:

$$J_\epsilon(X) = J(X) + \frac{1}{\epsilon}\|G(X)\|^2. \tag{5.7}$$

X_ϵ is the minimizer of J_ϵ, when $\epsilon- > 0$ then $X_\epsilon- > X^*$ solution of the original constrained optimization problem.

X_ϵ is the solution of the equation:

$$\nabla J_\epsilon(X) = \nabla J(X) + \frac{2}{\epsilon}\left[\frac{\partial G}{\partial X}\right]^t .G(X) = 0. \tag{5.8}$$

$$\tag{5.9}$$

As above, the minimization of the penalized functional is solved by an iterative algorithms of descent type. A major inconvenient of this method is to become quickly ill conditioned when ϵ is small. This is the basic *weak constraint formalism* of Sasaki with the difference that ϵ is fixed and doesn't change with the iterations. A consequence is that the constraint is not exactly verified, the choice of ϵ could permit some control on the amplitude of the error on the constraint.

5.1.3 Augmented Lagrangian Methods

This algorithm is a combination of duality and penalization, it is defined by an Augmented Lagrangian :

$$\mathcal{L}_\epsilon(X, \Lambda) = J(X) + (\Lambda, G(X)) + \frac{1}{\epsilon}\|G(X)\|^2. \tag{5.10}$$

$(X_\epsilon, \Lambda_\epsilon)$, saddle point of the Augmented Lagrangian is a solution of the constrained optimization problem. It is evaluated by a descent-ascent iterative method. The penalty term added to the Lagrangian can be considered as a regularization term in the sense of Tykhonov and make the problem well conditioned. This is exactly the sense of the background term in the usual terminology of VDA.

5.1.4 Remarks

For dynamical models if there is no difference between state variable and control variable then all the evolution of the model as to be considered as the state variable and we have to deal with huge numerical problems. At the present time, for operational models, the size of the variable, at a given time is of the order of 1 billion, therefore if we want to carry out an analysis on 1000 time step, the dimension of the variable will be of the order of 10^{12}, this is out of the scope of numerical optimization.
The *weak constraint formalism* permits to consider some error on the model but the algorithms doesn't permit neither to evaluate this error nor to identify its source. In the next sections we will see how to alleviate this inconvenient.

5.2 Direct Control of the error in VDA

5.2.1 General Formalism

Let's go back to the general formalism of VDA with a dynamical model. In this approach we introduce some state error Y in some space $\mathcal{Y}$ as state variable,and the model is now written as :

$$\frac{dX}{dt} = F(X) + \Pi.Y \tag{5.11}$$

$$X_0 = U. \tag{5.12}$$

Π is a linear operator from $\mathcal{X}$ to $\mathcal{Y}$, a priori Y depends on time but it can be steady state. The cost function, we want to minimize with respect to U and Y is defined by :

$$J(U,Y) = \frac{1}{2}\int_0^T \|H[X(U,Y,t)] - X_{obs}(t)\|^2 dt + \frac{1}{2}\|U - U_0\|^2 + \frac{1}{2}\|Y\|^2, \tag{5.13}$$

The last term in the definition of the cost function is to have the error Y as small as possible, while H is the linear observation operator. In order to simplify notations the covariances errors are the identity. Using more complex covariances is straightforward.

5.2.2 Optimality System

As usual, we introduce two directions to compute the directional derivatives. The gradient of J has two components:

$$\nabla J = \begin{pmatrix} \nabla_U J \\ \nabla_Y J \end{pmatrix}.$$

We introduce the adjoint variable P, the solution of the following system:

$$\begin{aligned} \frac{dP}{dt} &= \left[\frac{\partial P}{\partial X}\right]^T .P + H^t(HX - X_{obs}) \\ P(T) &= 0. \end{aligned} \tag{5.14}$$

Then we obtain :

$$\nabla J = \begin{pmatrix} -P(0) + U - U_0 \\ \\ \Pi^t P + Y \end{pmatrix}. \tag{5.15}$$

For practical purposes Y has to be located in a space with a dimension comparable to the dimension of the initial condition U, if the dimension of $\mathcal{Y}$ were too large then the problem would become numerically intractable. A way to alleviate this difficulty is to discretize the error spaces of test functions, we will write Y under the form:

$$Y = \sum_{i=1}^{n}\sum_{j=1}^{m} k_{ij}\phi_i\psi_j, \tag{5.16}$$

where k_{ij} are the elements of matrix K, Φ is a time dependent vector with elements ϕ_i and Ψ represents a vector of steady state elements ψ_i. Therefore we have $Y = \Phi^t K \Psi$ and the model becomes:

$$\frac{dX}{dt} = F(X) + \Pi.\Phi^t K \Psi \tag{5.17}$$

$$X_0 = U. \tag{5.18}$$

With a cost function:

$$J(U,K) = \frac{1}{2}\int_0^T \|H[X(U,K,t)] - X_{obs}(t)\|^2 dt + \frac{1}{2}\|U - U_0\|^2 + \frac{1}{2}\|K\|^2, \tag{5.19}$$

The adjoint model is the same than above but the second component of the gradient i.e. the gradient with respect to K becomes:

$$\nabla_U J = \int_0^T \Phi\Pi^t P\Psi dt + K. \tag{5.20}$$

5.2.3 Control of the error of observation

By the same token we can consider an error of observation and try to identify it, therefore we introduce a control variable Z belonging to space of observation. The model remains the same only the cost function is going to change and becomes:

$$J(U,Y) = \frac{1}{2}\int_0^T \|H[X(U,Y,t)] - X_{obs}(t) - Z(t)\|^2 dt + \frac{1}{2}\|U - U_0\|^2 + \frac{1}{2}\|Z\|^2, \tag{5.21}$$

It's easy to see that the adjoint model becomes:

$$\begin{aligned} \frac{dP}{dt} &= \left[\frac{\partial P}{\partial X}\right]^t .P + H^t(HX - X_{obs} - Z) \\ P(T) &= 0. \end{aligned} \tag{5.22}$$

and the gradient has a component with respect to U which is unchanged and we also have a component of the gradient with respect to Z which is :

$$\nabla_Z J = (2Z + X_{obs} - HX) \tag{5.23}$$

As we did above the error of observation can be discretized in an adequate base in order to reduce the dimension of the system. In theory both observation errors and model errors could be jointly controlled.

5.3 Weak constraint : control of systematic error.

The error of the model could be a random error and/or a systematic error. To identify a systematic error an additional term can be included in the model written as :

$$\frac{dX}{dt} = F(X) + E(K, t) \tag{5.24}$$

$$X_0 = U. \tag{5.25}$$

Where K is a low order parameter we want to identify. We assume that the error is governed by the system:

$$\frac{dE}{dt} = G(E, K) \tag{5.26}$$

$$E_0 = V. \tag{5.27}$$

In this case the assimilation will be the determination of (U^*, V^*, K^*) minimizing the cost function defined by:

$$J(U, V, K) = \frac{1}{2}\int_0^T \|H[X(U, Y, t)] - X_{obs}(t)\|^2 dt \tag{5.28}$$

$$+\frac{1}{2}\int_0^T \|E\|^2 dt + \frac{1}{2}\|U - U_0\|^2 + \frac{1}{2}\|K\|^2, \tag{5.29}$$

To get the Optimality System we need to introduce two adjoints variables P and Q as the solution of :

$$\frac{dP}{dt} = \left[\frac{\partial P}{\partial X}\right]^t + H^t(HX - X_{obs} - Z) \tag{5.30}$$

$$P(T) = 0 \tag{5.31}$$

$$\frac{dQ}{dt} = \left[\frac{\partial Q}{\partial E}\right]^t .Q + E \tag{5.32}$$

$$Q(T) = 0 \tag{5.33}$$

$$\tag{5.34}$$

Thanks to a backward integration of the adjoint model we get the three components of the gradient $\nabla J(U, V, K)$:

$$\nabla_U J = (U - U_0) - P(0) \tag{5.35}$$

$$\nabla_V J = -Q(0) + E \tag{5.36}$$

$$\nabla_K J = \left[\frac{\partial G}{\partial K}\right]^t .Q + \frac{\partial G}{\partial K} \tag{5.37}$$

The control of both errors of model and of data can be carried out simultaneously if a model of observation error were added.

5.4 Example: Saint-Venant's equations

Saint-Venant's equations, also known as shallow water equations, are used for an incompressible fluid for which the depth is small with respect to the horizontal dimensions. General equations of geophysical fluid dynamics are vertically integrated using the hydrostatic hypothesis, therefore vertical acceleration is neglected. In Cartesian coordinates they are :

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + \frac{\partial \phi}{\partial x} = 0 \tag{5.38}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + \frac{\partial \phi}{\partial y} = 0 \tag{5.39}$$

$$\frac{\partial \phi}{\partial t} + \frac{\partial u\phi}{\partial x} + \frac{\partial v\phi}{\partial y} = 0 \tag{5.40}$$

In this system $X = (u, v, \phi)^T$ is the state variable, u and v are the components of the horizontal velocity; ϕ is the geopotential (proportional to the height of the free surface) and f the Coriolis' parameter. For sake of simplicity, the following hypothesis are used:

a) The error of the model is neglected. Only the initial condition will be considered as a control variable.

b) Lateral boundary conditions are periodic. This is verified in global models.

c) Observations are supposed to be continuous with respect to time. Of course this is not the case in practice. The observation operators are taken identity matrices. If $X_0 = (u_0, v_0, \phi_0)^T$ is the initial condition and the cost function is given by :

$$J(X_0) = \frac{1}{2}\int_0^T [\|u - u_{obs}\|^2 + \|v - v_{obs}\|^2 + \gamma\|\phi - \phi_{obs}\|^2]dt, \tag{5.41}$$

where γ is a weight function, then the directional derivatives $\bar{X} = (\bar{u}, \bar{v}, \bar{\phi})^T$ in the direction $h = (h_u, h_v, h_\phi)^T$ (in the control space) will be solutions of the linear tangent model :

$$\frac{\partial \bar{u}}{\partial t} + u\frac{\partial \bar{u}}{\partial x} + \bar{u}\frac{\partial u}{\partial x} + v\frac{\partial \bar{u}}{\partial y} + \bar{v}\frac{\partial u}{\partial y} - f\bar{v} + \frac{\partial \bar{\phi}}{\partial x} = 0 \tag{5.42}$$

$$\frac{\partial \bar{v}}{\partial t} + u\frac{\partial \bar{v}}{\partial x} + \bar{u}\frac{\partial v}{\partial x} + v\frac{\partial \bar{v}}{\partial y} + \bar{v}\frac{\partial v}{\partial y} + f\bar{u} + \frac{\partial \bar{\phi}}{\partial y} = 0 \tag{5.43}$$

$$\frac{\partial \bar{\phi}}{\partial t} + \frac{\partial \bar{u}\phi}{\partial x} + \frac{\partial u\bar{\phi}}{\partial x} + \frac{\partial \bar{v}\phi}{\partial y} + \frac{\partial v\bar{\phi}}{\partial y} = 0 \tag{5.44}$$

Introducing three adjoint variables p, q, φ we can compute the adjoint system, it writes :

$$\frac{\partial p}{\partial t} + \frac{\partial pu}{\partial x} + v\frac{\partial p}{\partial y} - q\frac{\partial v}{\partial x} - fq + \frac{\partial \phi\varphi}{\partial x} = u_{obs} - u, \tag{5.45}$$

$$\frac{\partial q}{\partial t} - p\frac{\partial u}{\partial y} + u\frac{\partial q}{\partial x} + \frac{\partial qv}{\partial y} + fp + \frac{\partial \phi\varphi}{\partial y} = v_{obs} - v, \tag{5.46}$$

$$\frac{\partial \varphi}{\partial t} + \frac{\partial u\varphi}{\partial x} + \frac{\partial v\varphi}{\partial y} + \frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} = \gamma(\phi_{obs} - \phi). \tag{5.47}$$

5.4.1 Control of errors with dynamics: an example

Now we consider the Burgers equation with homogeneous boundary conditions and the state variable is u defined on $[0,1] \times [0,T]$ with an error $E(x,t)$ governed by a diffusion equation with an unknown parameter γ. The unknown initial conditions on u and E to be found are α and β. The model and the error are characterized by:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} - \frac{\partial}{\partial x}(\mu\frac{\partial u}{\partial x}) + E = F, \tag{5.48}$$

$$u(0) = \alpha. \tag{5.49}$$

$$\frac{\partial E}{\partial t} - \frac{\partial}{\partial x}(\gamma\frac{\partial E}{\partial x}) = 0, \tag{5.50}$$

$$E(0) = \beta. \tag{5.51}$$

The problem is to determine$\alpha^*, \beta^*, \gamma^*$ minimizing the cost function J defined by:

$$J(\alpha,\beta,\gamma) = \frac{1}{2}\int_0^T\int_0^1 (u-u_{obs})^2 dxdt + \frac{1}{2}\int_0^T\int_0^1 E^2 dxdt + \frac{1}{2}\int_0^1 \gamma^2 dx + \frac{1}{2}\alpha^2 + \frac{1}{2}\beta^2. \tag{5.52}$$

For sake of simplicity we have chosen the simplest form with a complete observation, identity observation operators and ignoring the statistical information. The forcing term is denoted by F and parameter μ is known. As usual the next step is to derive the Gâteaux derivatives u and E in some directions in the spaces of control variables then introducing p and q two adjoint variables as the solution of the equations:

$$\frac{\partial p}{\partial t} + u\frac{\partial p}{\partial t} + \frac{\partial}{\partial x}\Big(\mu\frac{\partial p}{\partial x}\Big) = u_{obs} - u \tag{5.53}$$

$$p(T) = 0 \tag{5.54}$$

$$\frac{\partial q}{\partial t} + \frac{\partial}{\partial x}(\gamma\frac{\partial q}{\partial x}) + E + p = 0 \tag{5.55}$$

$$q(T) = 0. \tag{5.56}$$

$$\tag{5.57}$$

Then the components of the gradient ∇J verify:

$$\nabla J_\alpha = -p(0) + \int_0^1 (u - u_{obs})dx \tag{5.58}$$

$$\nabla J_\beta = -q(0) + \beta \tag{5.59}$$

$$\nabla J_\gamma = \int_0^T \Big(\frac{\partial E}{\partial x}\frac{\partial q}{\partial x} + q\frac{\partial^2 E}{\partial x^2}\Big)dx \tag{5.60}$$

$$\tag{5.61}$$

6 Second order methods

The optimality system, the Euler-Lagrange equation, provides only a necessary condition for optimality. In the linear case, the solution is unique if the Hessian is positive definite. From a general point of view the information given by the Hessian is important for theoretical, numerical and practical issues. For operational models it is impossible to compute the Hessian itself, as it is a square matrix with around 10^{18} terms, nevertheless the most important information can be extracted from the spectrum of the Hessian which can be estimated without an explicit determination of this matrix. This information is of importance for estimating the condition number of the Hessian for preparing an efficient preconditioning.

A general method to get this information is to apply the techniques described above to the couple made by the direct and adjoint models (Le Dimet et al. **?**]), leading to a so called second order adjoint. The following steps are carried out :

- Linearization of the direct and adjoint models with respect to the state variable.
- Introducing second order adjoint variables.

- Transposition to put in light the linear dependence with respect to the directions.

The system obtained, the second order adjoint, is used to compute the product of the Hessian by any vector. Of course if we consider all the vectors of the canonical base then it will be possible to get the complete Hessian.

The determination of this product permits to access some information.

- By using Lanczos type methods and deflation, it is possible to compute the eigenvectors and eigenvalues of the Hessian.
- To carry out second order optimization methods of Newton-type are used for equations of the form:

  $$\nabla J(X) = 0$$

  The iterations are :

  $$X_{n+1} = X_n - H^{-1}(X_n).\nabla J(X_n)$$

  where H is the Hessian of J. At each iteration a linear system should be solved. This is done by carrying out some iterations of a conjugate gradient methods which requires computing the Hessian-vector product.

For the Saint-Venant equations the second order adjoint system is given by

$$\begin{aligned}
&\frac{\partial \hat{u}}{\partial t} + u\frac{\partial \hat{u}}{\partial x} + v\frac{\partial \hat{v}}{\partial y} + \hat{u}\frac{\partial v}{\partial y} - \hat{v}\frac{\partial v}{\partial y} - f\hat{v} + \phi\frac{\partial \hat{\phi}}{\partial x} \\
&\qquad = \tilde{v}\frac{\partial \bar{v}}{\partial x} - \bar{u}\frac{\partial \tilde{u}}{\partial x} - \bar{v}\frac{\partial \bar{u}}{\partial y} + \tilde{u}\frac{\partial \bar{v}}{\partial y} - \bar{\phi}\frac{\partial \tilde{\phi}}{\partial x} - \bar{u} \qquad (6.1)\\
&\frac{\partial \hat{v}}{\partial t} + \hat{u}\frac{\partial u}{\partial y} - u\frac{\partial \hat{v}}{\partial x} + \hat{v}\frac{\partial u}{\partial x} + v\frac{\partial \hat{v}}{\partial y} + f\hat{u} + \phi\frac{\partial \hat{\phi}}{\partial y} \\
&\qquad = \tilde{u}\frac{\partial \bar{u}}{\partial x} - \bar{u}\frac{\partial \tilde{v}}{\partial x} - \tilde{v}\frac{\partial \bar{u}}{\partial y} + \bar{u}\frac{\partial \tilde{v}}{\partial y} - \bar{\phi}\frac{\partial \tilde{\phi}}{\partial y} - \bar{v} \qquad (6.2)\\
&\frac{\partial \hat{\phi}}{\partial t} + \frac{\partial \hat{u}}{\partial x} + \frac{\partial \hat{v}}{\partial y} + u\frac{\partial \hat{\phi}}{\partial x} + v\frac{\partial \hat{\phi}}{\partial y} = -\bar{u}\frac{\partial \tilde{\phi}}{\partial x} - \bar{v}\frac{\partial \tilde{\phi}}{\partial x} - \gamma\bar{\phi}, \qquad (6.3)
\end{aligned}$$

where $Q = (\hat{u}, \hat{v}, \hat{\phi})^T$ and $R = (\bar{u}, \bar{v}, \bar{\phi})^T$ are the second and first order adjoint variables.

From the formal point of view we see that first and second order differ by second order terms which do not take into account the adjoint variable. The computation of second derivatives require storing both the trajectories of the direct and adjoint models. For very large models it could be more economical to recompute these trajectories.

6.1 Sensitivity analysis.

In the environmental sciences the mathematical models contain parameters which cannot be estimated very precisely either because they are difficult to measure or because they represent some subgrid phenomena. Therefore it is important to be able to estimate the impact of uncertainties on the outputs of the model. Sensitivity analysis is defined by:

- X is the state vector of the model, K a vectorial parameter of the model $F(X, K) = 0$.
- $G(X, K)$ is the response function : a real value function
- By definition the sensitivity of the model is the gradient of G with respect to K. The difficulty encountered comes from the implicit dependence of G on K through X, solution of the model.

Several methods can be used to estimate the sensitivity:

- By finite differences. We get:

$$\frac{\partial G}{\partial K_i} \simeq \quad \frac{G\left(X(K+\alpha e_i), K+\alpha e_i\right) - G\left(X(K), K\right)}{\alpha},$$

  where K_i is the i^{th} component of vector K.

  The main inconvenience of this method is its computational cost : it requires solving the model as many times as the dimension of the model. Furthermore the determination of the parameter α may be tricky . If it too large, the variation of G could be nonlinear, since small value round off errors may dominate the variation of G. The main advantage of this method is that it is very easy to implement.

- Sensitivity via an adjoint model. Let $F(X, K) = 0$ be the direct model. We introduce its adjoint:

$$\left[\frac{\partial F}{\partial X}\right]^* .P = \frac{\partial G}{\partial X},$$

  where $\left[\frac{\partial F}{\partial X}\right]^*$ is the adjoint operator of $\frac{\partial F}{\partial X}$. Then the gradient is given by:

$$\nabla G = \frac{\partial G}{\partial K} - \left[\frac{\partial F}{\partial K}\right]^* .P$$

The advantage of this method is that the sensitivity is obtained by only one run of the adjoint model. The price to be paid is the derivation of the adjoint code.

6.2 Sensitivity in the presence of data.

In geophysics a usual request is the estimation of the sensitivity with respect to observations. What will be the impact of an uncertainty on the prediction? It is clear that observations are not directly used in the direct model, since they can be found only among the forcing terms in the adjoint model. Therefore to apply the general formalism of sensitivity analysis we should apply it not to the model itself but to the optimality system, i.e. the model plus the adjoint model. A very simple example with a scalar ordinary differential equation is given in [49] showing that the direct model is not sufficient to carry out sensitivity analysis in the presence of data. Deriving the optimality system will introduce second order derivatives.

Consider a model of the form $F(X, I) = 0$ where I stands for the input of the model. Assuming F is a steady state or time dependent operator, $J(X, I)$ is the cost function in the variational data assimilation and P is the adjoint variable, then the optimality system is:

$$\begin{cases} F(X,I) = 0 \\ \left[\frac{\partial F}{\partial X}\right]^* \cdot P - \frac{\partial J}{\partial X} = 0 \\ \left[\frac{\partial F}{\partial I}\right]^* \cdot P - \frac{\partial J}{\partial I} = 0 \end{cases} \tag{6.4}$$

The optimality system can be considered as a generalized model: $\mathcal{F}$ with the state variable $Z = \begin{pmatrix} X \\ P \end{pmatrix}$.

If i is a perturbation on I, we get:

$$\frac{\partial F}{\partial X} \cdot \hat{X} + \frac{\partial F}{\partial I} \cdot i = 0 \tag{6.5}$$

$$\left[\frac{\partial^2 F}{\partial X^2} \cdot \hat{X} + \frac{\partial^2 F}{\partial X \partial I} \cdot i\right]^* \cdot P + \left[\frac{\partial F}{\partial X}\right]^* \cdot \hat{P} - \frac{\partial^2 J}{\partial X^2} \cdot \hat{X} - \frac{\partial^2 J}{\partial X \partial I} \cdot i = 0 \tag{6.6}$$

$$\left[\frac{\partial^2 F}{\partial I^2} \cdot i + \frac{\partial^2 F}{\partial X \partial I} \cdot \hat{X}\right]^* \cdot P + \left[\frac{\partial F}{\partial I}\right]^* \cdot \hat{P} - \frac{\partial^2 J}{\partial I^2} \cdot i - \frac{\partial^2 J}{\partial I \partial X} \cdot \hat{X} = 0 \tag{6.7}$$

$$\hat{\mathbf{G}}(X, I, i) = \frac{\partial \mathbf{G}}{\partial X} \cdot \hat{X} + \frac{\partial \mathbf{G}}{\partial I} \cdot i, \tag{6.8}$$

where $\widehat{X}$ and $\widehat{P}$ are the Gâteaux derivative of X and P in the direction i. Let us introduce the second order adjoint variables Q and R. Here G is the response function introduce in subsection 6.1 and we are looking to estimate the gradient of G with respect to I. By taking the inner product of (6.5) and (6.6) by Q, and of (6.7) by R, and adding it, we obtain:

$$\begin{aligned} &< \hat{X}, \left[\tfrac{\partial F}{\partial X}\right]^* \cdot Q + \left[\tfrac{\partial^2 F}{\partial X^2} \cdot Q\right]^* \cdot P - \left[\tfrac{\partial^2 J}{\partial X^2}\right]^* \cdot Q - \left[\tfrac{\partial^2 J}{\partial I \partial X}\right]^* \cdot R + \left[\tfrac{\partial^2 F}{\partial X \partial I} \cdot P\right]^* \cdot R > \\ &+ < \hat{P}, \left[\tfrac{\partial F}{\partial X}\right] \cdot Q + \left[\tfrac{\partial F}{\partial I}\right] \cdot R > + \\ &< i, \left[\tfrac{\partial F}{\partial I}\right]^* \cdot Q + \left[\tfrac{\partial^2 F}{\partial X \partial I} \cdot Q\right]^* \cdot P - \left[\tfrac{\partial^2 J}{\partial X \partial I}\right]^* \cdot Q - \left[\tfrac{\partial^2 J}{\partial I^2}\right]^* \cdot R + \left[\tfrac{\partial^2 F}{\partial I^2} \cdot R\right]^* \cdot P >= 0 \end{aligned}$$

Identifying in (6.8) it is seen that if Q and R are defined as solutions of :

$$\begin{cases} \left[\frac{\partial F}{\partial X}\right]^* \cdot Q + \left[\frac{\partial^2 F}{\partial X^2} \cdot Q\right]^* \cdot P - \left[\frac{\partial^2 J}{\partial X^2}\right]^* \cdot Q - \left[\frac{\partial^2 J}{\partial I \partial X}\right]^* \cdot R + \left[\frac{\partial^2 F}{\partial X \partial I} \cdot P\right]^* \cdot R = \frac{\partial \mathbf{G}}{\partial X} \\ \left[\frac{\partial F}{\partial X}\right] \cdot Q + \left[\frac{\partial F}{\partial I}\right] \cdot R = 0 \end{cases} \tag{6.9}$$

Then we get the gradient of G with respect to I (the sensitivity) by :

$$\mathcal{S} = \frac{d\mathbf{G}}{dI} = -\left[\frac{\partial F}{\partial I}\right]^* \cdot Q - \left[\frac{\partial^2 F}{\partial X \partial I} \cdot Q\right]^* \cdot P + \left[\frac{\partial^2 J}{\partial X \partial I}\right]^* \cdot Q + \left[\frac{\partial^2 J}{\partial I^2}\right]^* \cdot R - \left[\frac{\partial^2 F}{\partial I^2} \cdot R\right]^* \cdot P + \frac{\partial \mathbf{G}}{\partial I} \tag{6.10}$$

To summarize the algorithm is the following:

i. Solve the optimality system (6.4) to get X and P

ii. Solve the coupled system (6.9) to compute Q and R

iii. Then the sensitivity is given by (6.10).

7 Sensitivity with respect to sources

7.1 Identification of the fields

Let us assume that the flow described by the state variable X satisfies the following time dependent differential system between time 0 and final time T :

$$\begin{cases} \frac{dX}{dt} &= F(X), \\ X(0) &= U. \end{cases} \tag{7.1}$$

The pollutant, considered as a passive tracer, is described by its concentration whose evolution is given by the following equations :

$$\begin{cases} \frac{dC}{dt} &= G(X, C, S), \\ C(0) &= V, \end{cases} \tag{7.2}$$

where C is the pollutant's concentration and S is a function of space and time and represents the production of pollutant.

The first task is to retrieve the fields from observations $X_{obs} \in \mathcal{X}_{obs}$ corresponding to the state variable X and $C_{obs} \in \mathcal{C}_{obs}$ associated with the concentration C of the pollutant. We introduce a cost function J defined by

$$J(U, V) = \frac{1}{2} \int_0^T \|EX - X_{obs}\|^2 dt + \frac{1}{2} \int_0^T \|DC - C_{obs}\|^2 dt \tag{7.3}$$

where E is an operator from the space of the state variable toward the space of observations and D from the space of concentration toward the space of observations of concentration. For sake of simplicity, we do not introduce regularization terms in the cost function. In practice they are of crucial importance. For retrieving the state variable and the concentration, we have to determine U^* and V^* which minimize J. They are solutions of the Optimality System. If P and Q are defined as the solutions of the following adjoint models :

$$\left\{ \begin{array}{rcl} \dfrac{dP}{dt} + \left[\dfrac{\partial F}{\partial X}\right]^* .P + \left[\dfrac{\partial G}{\partial X}\right]^* .Q & = & E^*(EX - X_{obs}); \\ P(T) & = & 0; \end{array} \right. \tag{7.4}$$

$$\left\{ \begin{array}{rcl} \dfrac{dQ}{dt} + \left[\dfrac{\partial G}{\partial C}\right]^* .Q & = & D^*(DC - C_{obs}); \\ Q(T) & = & 0, \end{array} \right. \tag{7.5}$$

then from the backward integration of the system we get the gradient:

$$\nabla J_U = -P(0) \tag{7.6}$$

$$\nabla J_V = -Q(0) \tag{7.7}$$

7.2 Formulation of the sensitivity problem

Let Ω_A, a sub-domain of the physical space be the region of interest (response region) and φ the function giving the measure of the effect of interest. By effect of interest, we mean the "effect of the pollutant" and we want to evaluate the sensitivity of φ with respect to the source. Our quantity of interest φ is a function of the concentration C and of the source S. We define the response function as:

$$\Phi_A(C,S) = \int_0^T \int_{\Omega_A} \varphi(C,S) dx. \tag{7.8}$$

In the simplest case, φ can be defined as $\varphi = C$, in which case Φ_A is the mean over the space and the time of the concentration of the pollutant. By definition, the sensitivity with respect to the source S is the gradient of the response function Φ_A with respect to S. Following the guidelines of the derivation of the gradient as presented in section 7.1, we obtain the tangent linear and adjoint models. Their solutions are $\hat{X}$, $\hat{C}$, $\hat{P}$ and $\hat{Q}$, the Gâteaux derivatives with respect to S (in the direction s) of the variables of the optimality system given by equations (7.1), (7.2), (7.4) and (7.5). The models are elaborated again below:

$$\left\{ \begin{array}{rcl} \dfrac{d\hat{X}}{dt} & = & \dfrac{\partial F}{\partial X} \cdot \hat{X} \\ \hat{X}(0) & = & \hat{U} \end{array} \right. \tag{7.9}$$

$$\left\{\begin{array}{rcl} \dfrac{d\hat{C}}{dt} & = & \dfrac{\partial G}{\partial X}\cdot\hat{X}+\dfrac{\partial G}{\partial C}\cdot\hat{C}+\dfrac{\partial G}{\partial S}\cdot s \\ \hat{C}(0) & = & \hat{V} \end{array}\right. \tag{7.10}$$

$$\left\{\begin{array}{rcl} \dfrac{d\hat{P}}{dt}+\left[\dfrac{\partial F}{\partial X}\right]^*\cdot\hat{P}+\left[\dfrac{\partial^2 F}{\partial X^2}\hat{X}\right]^*\cdot P+\left[\dfrac{\partial G}{\partial X}\right]^*\cdot\hat{Q} & & \\ \quad+\left[\dfrac{\partial^2 G}{\partial X^2}\cdot\hat{X}+\dfrac{\partial^2 G}{\partial C\partial X}\cdot\hat{C}+\dfrac{\partial^2 G}{\partial S\partial X}\cdot s\right]^*\cdot Q & = & E^*E\hat{X} \\ \hat{P}(T) & = & 0, \end{array}\right. \tag{7.11}$$

$$\left\{\begin{array}{rcl} \dfrac{d\hat{Q}}{dt}+\left[\dfrac{\partial G}{\partial C}\right]^*\cdot\hat{Q}+\left[\dfrac{\partial^2 G}{\partial C^2}\cdot\hat{C}+\dfrac{\partial^2 G}{\partial X\partial C}\hat{X}+\dfrac{\partial^2 G}{\partial S\partial I}s\right]^*\cdot Q & = & D^*D\cdot\hat{C} \\ \hat{Q}(T) & = & 0, \end{array}\right. \tag{7.12}$$

where $\hat{U}$ in (7.9) and $\hat{V}$ in (7.10) are the Gâteaux derivatives of the optimal initial condition U and V, respectively. These terms are important because the information is propagated from the initial condition. For the response function, the Gâteaux derivatives with respect to S is given by:

$$\hat{\varphi}(S,C,s)=\frac{\partial\varphi}{\partial S}\cdot s+\frac{\partial\varphi}{\partial C}\cdot\hat{C} \tag{7.13}$$

To compute the gradient $\nabla_S\varphi(C,S)$, we introduce four second-order adjoint variables Γ, Λ, Φ and Ψ; the system of equations (7.9) is multiplied by Γ, (7.10) by Λ, (7.11) by Φ and (7.12) by Ψ, all the terms are added together and integrated by parts and we get:

$$\mathcal{Z}+\int_0^T\left\langle\hat{X},\mathcal{A}\right\rangle dt+\int_0^T\left\langle\hat{C},\mathcal{B}\right\rangle dt+\int_0^T\left\langle\hat{P},\mathcal{L}\right\rangle dt+\int_0^T\left\langle\hat{Q},\mathcal{W}\right\rangle dt+\int_0^T\langle s,\mathcal{K}\rangle\, dt=0 \tag{7.14}$$

with:

$$\begin{array}{rcl} \mathcal{Z} & = & \left\langle\Gamma(T),\hat{X}(T)\right\rangle-\left\langle\Gamma(0),\hat{X}(0)\right\rangle+\left\langle\Lambda(T),\hat{C}(T)\right\rangle-\left\langle\Lambda(0),\hat{C}(0)\right\rangle \\ & & +\left\langle\Phi(T),\hat{P}(T)\right\rangle-\left\langle\Phi(0),\hat{P}(0)\right\rangle+\left\langle\Psi(T),\hat{Q}(T)\right\rangle-\left\langle\Psi(0),\hat{Q}(0)\right\rangle, \end{array} \tag{7.15}$$

$$\begin{array}{rcl} \mathcal{A} & = & -\dfrac{d\Gamma}{dt}-\left[\dfrac{\partial F}{\partial X}\right]^*\cdot\Gamma-\left[\dfrac{\partial G}{\partial X}\right]^*\cdot\Lambda+\left[\dfrac{\partial^2 F}{\partial X^2}\Phi\right]^*\cdot P \\ & & +\left[\dfrac{\partial^2 G}{\partial X^2}\Phi\right]^*\cdot Q+\left[\dfrac{\partial^2 G}{\partial C\partial X}\Psi\right]^*\cdot Q-E^*E\Phi, \end{array} \tag{7.16}$$

$$\mathcal{B} = -\frac{d\Lambda}{dt} - \left[\frac{\partial G}{\partial C}\right]^* \cdot \Lambda + \left[\frac{\partial^2 G}{\partial C \partial X}\Phi\right]^* \cdot Q + \left[\frac{\partial^2 G}{\partial X^2}\Psi\right]^* \cdot Q - D^* D\Psi, \tag{7.17}$$

$$\mathcal{L} = -\frac{d\Phi}{dt} + \left[\frac{\partial F}{\partial X}\right] \cdot \Phi, \tag{7.18}$$

$$\mathcal{W} = -\frac{d\Psi}{dt} + \left[\frac{\partial G}{\partial C}\right] \cdot \Psi \tag{7.19}$$

and

$$\mathcal{K} = -\left[\frac{\partial G}{\partial S}\right]^* \cdot \Lambda + \left[\frac{\partial^2 G}{\partial X^2}\Phi\right]^* \cdot Q + \left[\frac{\partial^2 G}{\partial C \partial S}\Psi\right]^* \cdot Q. \tag{7.20}$$

If the known values are taken into account in the expression of $\mathcal{Z}$, it becomes:

$$\begin{aligned} \mathcal{Z} = & \left\langle \Gamma(T), \hat{X}(T) \right\rangle - \left\langle \Gamma(0), \hat{U} \right\rangle + \left\langle \Lambda(T), \hat{C}(T) \right\rangle - \left\langle \Lambda(0), \hat{V} \right\rangle \\ & - \left\langle \Phi(0), \hat{P}(0) \right\rangle - \left\langle \Psi(0), \hat{Q}(0) \right\rangle. \end{aligned} \tag{7.21}$$

If Γ, Λ, Φ and Ψ are the solution of the following second order adjoint systems of equations

$$\begin{cases} -\dfrac{d\Gamma}{dt} - \left[\dfrac{\partial F}{\partial X}\right]^* \cdot \Gamma - \left[\dfrac{\partial G}{\partial X}\right]^* \cdot \Lambda + \left[\dfrac{\partial^2 F}{\partial X^2}\Phi\right]^* \cdot P \\ \qquad + \left[\dfrac{\partial^2 G}{\partial X^2}\Phi\right]^* \cdot Q + \left[\dfrac{\partial^2 G}{\partial C \partial X}\Psi\right]^* \cdot Q - E^* E\Phi = 0; \\ \Gamma(0) = 0; \\ \Gamma(T) = 0, \end{cases} \tag{7.22}$$

$$\begin{cases} -\dfrac{d\Lambda}{dt} - \left[\dfrac{\partial G}{\partial C}\right]^* \cdot \Lambda + \left[\dfrac{\partial^2 G}{\partial C \partial X}\Phi\right]^* \cdot Q + \left[\dfrac{\partial^2 G}{\partial X^2}\Psi\right]^* \cdot Q - D^* D\Psi = \dfrac{\partial \varphi}{\partial C}; \\ \Lambda(0) = 0; \\ \Lambda(T) = 0, \end{cases} \tag{7.23}$$

$$-\frac{d\Phi}{dt} + \left[\frac{\partial F}{\partial X}\right] \cdot \Phi = 0, \tag{7.24}$$

$$-\frac{d\Psi}{dt} + \left[\frac{\partial G}{\partial C}\right] \cdot \Psi = 0, \tag{7.25}$$

then it comes that:

$$\nabla\varphi = \left[\frac{\partial G}{\partial S}\right]^* \cdot \Lambda - \left[\frac{\partial^2 G}{\partial X^2}\Phi\right]^* \cdot Q - \left[\frac{\partial^2 G}{\partial C \partial S}\Psi\right]^* \cdot Q + \frac{\partial \varphi}{\partial S}. \tag{7.26}$$

The conditions $\Gamma(0) = 0$ and $\Gamma(T) = 0$ are imposed because there is no information on $\hat{X}(T)$ and $\hat{U}$, respectively. For similar reasons, the conditions $\Lambda(0) = 0$ and $\Lambda(T) = 0$ are also imposed. Because the initial conditions are optimal, $\hat{P}(0)$ defined by

$$\hat{P}(0) = \lim_{\alpha \to 0} \frac{P(0)|_{S+\alpha s} - P(0)|_S}{\alpha} \tag{7.27}$$

is zero. Similarly, $\hat{Q}(0) = 0$.

We get a coupled system of four differential equations (7.22) to (7.25) of the first order with respect to time. Two equations have an initial and a final condition while the two others have no condition at all: that is a non-standard problem.

A theoretical question remains on the existence and the uniqueness of a solution. Some developments in this direction are underway.

7.2.1 Solving the non-standard problem

To make it simple, we consider a system of two differential equations, the extension to four equations being straightforward. The method proposed is based on the theory of optimal control [52]. We consider the generic system on the time interval $[0, T]$

$$\left\{ \begin{array}{rcl} \dfrac{dX}{dt} & = & K(X, Y),\ t \in [0, T]; \\ \dfrac{dY}{dt} & = & L(X, Y),\ t \in [0, T] \end{array} \right. \tag{7.28}$$

with

$$\left\{ \begin{array}{rcl} X(0) & = & 0; \\ X(T) & = & 0 \end{array} \right. \tag{7.29}$$

and no condition on Y. Let transform it into a problem of optimal control. We consider the problem (7.28) with the initial condition

$$\left\{ \begin{array}{rcl} X(0) & = & 0; \\ Y(0) & = & U. \end{array} \right. \tag{7.30}$$

We assume that under these conditions, (7.28) has a unique solution for $t \in [0, T]$. Let $X(T, U)$ be the value of X at time $t = T$ for the value U of $Y(0)$, we define the cost function

$$J_P(U) = \frac{1}{2} \|X(T, U)\|^2. \tag{7.31}$$

The problem becomes the determination of U^* by minimizing J_P. We can expect that at the optimum $X(T, U^*) = 0$. The Gâteaux derivatives with respect to U in the direction u are given

by:

$$\frac{d\hat{X}}{dt} = \frac{\partial K}{\partial X}\cdot\hat{X} + \frac{\partial K}{\partial Y}\cdot\hat{Y} \tag{7.32}$$

$$\frac{d\hat{Y}}{dt} = \frac{\partial L}{\partial X}\cdot\hat{X} + \frac{\partial L}{\partial Y}\cdot\hat{Y} \tag{7.33}$$

$$\hat{X}(0) = 0 \tag{7.34}$$

$$\hat{Y}(0) = u \tag{7.35}$$

$$\hat{J}_p(U) = \left\langle X(T), \hat{X}(T)\right\rangle \tag{7.36}$$

Let us introduce the adjoint variables W and Z and proceed to the integration by part. We get:

$$\begin{array}{l}\left\langle \hat{X}(T), W(T)\right\rangle + \left\langle \hat{Y}(T), Z(T)\right\rangle - \left\langle \hat{Y}(0), Z(0)\right\rangle \\ -\displaystyle\int_0^T \left\langle \hat{X}, \frac{dW}{dt} + \left[\frac{\partial K}{\partial X}\right]^*\cdot W + \left[\frac{\partial L}{\partial X}\right]^*\cdot Z\right\rangle dt \\ -\displaystyle\int_0^T \left\langle \hat{Y}, \frac{dZ}{dt} + \left[\frac{\partial K}{\partial Y}\right]^*\cdot W + \left[\frac{\partial L}{\partial Y}\right]^*\cdot Z\right\rangle dt = 0\end{array} \tag{7.37}$$

If Z and W are defined as the solution of:

$$\frac{dW}{dt} + \left[\frac{\partial K}{\partial X}\right]^*\cdot W + \left[\frac{\partial L}{\partial X}\right]^*\cdot Z = 0; \tag{7.38}$$

$$\frac{dZ}{dt} + \left[\frac{\partial K}{\partial Y}\right]^*\cdot W + \left[\frac{\partial L}{\partial Y}\right]^*\cdot Z = 0; \tag{7.39}$$

$$Z(T) = 0;\ W(T) = X(T), \tag{7.40}$$

then we get

$$\nabla J_P(U) = Z(0). \tag{7.41}$$

A theoretical question remains on the existence and the uniqueness of a solution. Some development in this direction are underway.

7.2.2 Extension and potential development

Without any theoretical difficulty, the development carried out above can be extended to:

- the case of several sources of pollution; in which case the source S becomes a vector $S = (S_1, S_2, \cdots, S_m)$, where S_i is the i^{th} source and m is the total number of sources;
- the case of several pollutant with kinetic chemistry. The same methodology can be applied with a vector of concentration in the place of a single concentration; $C = (C_1, C_2, \cdots, C_m)$. Some numerical difficulties may arise if the characteristic time of the kinetic are very heterogeneous.

8 Incremental Methods

During the 1990s, the 4D VAR methodology matured and was adopted at several important international Numerical Weather Prediction centers. However, although 4D VAR cost function and gradient can be evaluated at the cost of one integration of the forecast model followed by one integration of the adjoint model, the computational cost to implement it was still prohibitive since a typical minimization requires between 10 and 100 evaluations of the gradient. The cost of the adjoint model is typically 3 times that of the forward model [34]. The analysis window in a typical operational model such as the ECMWF system is 12-hours. Thus, the cost of the analysis is roughly equivalent to between 20 and 200 days of model integration with 10^8 variables, making it computationally prohibitive for NWP centers that have to deliver timely forecasts to the public. Amongst the methods that greatly facilitated the adoption, application and implementation of 4D VAR data assimilation at major operational centers and contributed to advance of the technique, the incremental 4D VAR method is of paramount importance. Courtier, Thépaut, and Hollingsworth [17] (based on an idea of Derber) introduced the incremental formulation of the 4D VAR. The incremental algorithm reduces the cost of 4D VAR mainly by reducing the resolution of the model, thus allowing the 4D VAR method to become computationally feasible.

The incremental 4D VAR algorithm reduces the resolution of the model and eliminates most of the time-consuming physical packages, thereby enabling the 4D VAR method to become computationally feasible. Furthermore, the incremental 4D VAR algorithm removes the nonlinearities in the cost minimization by using a forward integration of the linear model instead of a nonlinear one. The minimization procedure is identical to the usual 4D VAR algorithm except that the increment trajectory is obtained by integration of the linear model.

The reference trajectory (which is needed for integrating the linear and adjoint models and which starts from the background integration) is not updated at every iteration. This simplified iterative procedure for minimizing the incremental cost function is called the inner loop, and is much cheaper computationally to implement by comparison to the full 4D VAR algorithm. However, when the quadratic cost function is approximated in this way, the incremental 4D VAR algorithm no longer converges to the solution of the original problem. Furthermore, the analysis increments are calculated at reduced resolution and must be interpolated to conform to the high-resolution models grid. Consequently, after performing a user-defined number of inner- loops, one outer-loop is performed to update the high-resolution reference trajectory and the observation departures. After each outer-loop update, it is possible to use progressively higher resolutions for the inner-loop Other simplifications introduced by the incremental 4D VAR method will be briefly described below. The nonlinearity of the model and/or of the observation operator can produce multiple minima in the cost function, which will impact the convergence of the minimization algorithm. The incremental 4D VAR algorithm removes the nonlinearities in the cost minimization by using a forward integration of the linear model instead of a nonlinear one. It also uses a coarser resolution model and eliminates most of the time-consuming physical packages. In this section we will address several algorithmic aspects of incremental 4D VAR that are used in present day implementations of 4D VAR data assimilation. Some aspects related to the incremental method versus the full 4D VAR

were addressed by Li, Navon, and Zhu [51]. They conducted a set of four-dimensional variational assimilation (4D VAR) experiments using both a standard method and an incremental method and compared the corresponding performances

8.1 Description of the Method

Courtier, Thépaut, and Hollingsworth [17] devised an incremental 4D VAR algorithm which removes nonlinearities in the minimization by using a forward integration of a linear model instead of a nonlinear one. The minimization of the cost functional is carried out at a reduced model resolution which leads to an effective reduction of computational cost and memory requirements. The 4D VAR problem consists in finding the state at time t0 that minimizes the cost function:

$$J(X_0) = \frac{1}{2}(X_0 - X_b)^T B^{-1}(X_0 - X_b) + \tag{8.1}$$

$$\frac{1}{2}\sum_{i=1}^{n}(H_i(X_i) - Y_i)^T R_{i-1}(H_i(X_i) - Y_i) \tag{8.2}$$

subject to the statesX_i satisfying the NWP model (8.3) as a strong constraint. In optimal control language this is referred to as Partial Differential Equations (PDE) constrained optimization. We consider a discrete nonlinear dynamical system given by the equation:

$$X_{i+1} = M_i(X_i), \tag{8.3}$$

where $X_i \in \mathbb{R}^n$ is the state vector at time t_i and M_i represents the nonlinear model operator that propagates the state vector at time t_i to time t_{i+1} for $i = 0, 1, \cdots, n-1$. We assume that we have imperfect observations $Y_i \in \mathbb{R}^{p_i}$ at t_i. Here $H_i : \mathbb{R}^n \to \mathbb{R}^{p_i}$ is known as being the observation operator and maps the state vector toward the observation space. The matrix B contains the background error covariances and R_i are the observation error covariances matrices. In the incremental formulation the solution to the nonlinear minimization problem is approximated by a sequence of minimizations of linear quadratic cost functions. We define $X_0^{(k)}$ as being the k^{th} estimate to the solution and we linearize the equation around the model trajectory from this estimate. The incremental approach is designed to find the analysis increment $\delta x_0^{(k)} = X_0^{(k)} - X_b$, that minimizes the following cost function

$$\widehat{J}(\delta x_0^{(k)}) = \frac{1}{2}(\delta x_0^{(k)})^T B^{-1}\delta x_0^{(k)} + \tag{8.4}$$

$$\frac{1}{2}\sum_{i=1}^{n}(d_k - \mathbf{H}_i\mathbf{M}_i\delta x_0^{(k)})^T R_{i-1}(d_k - \mathbf{H}_i\mathbf{M}_i\delta x_0^{(k)}), \tag{8.5}$$

where $d_i = Y_i - H_i(M_{0\to i}(X_b))$ are the innovation vectors at time step i and $\mathbf{M}_i$, and $\mathbf{H}_i$ denote the tangent linear versions of the forecast model and observation operators. Now the constraint is

given by the tangent linear model

$$\delta x_{i+1}^{(k)} = \mathbf{M}_i \delta x_i^{(k)}.$$

The process of minimization is similar to the usual 4-D VAR algorithm except that the control variable is the increment at time t_0 and and $\mathbf{M}_i$ depicts the linear tangent model operator evaluated at the current estimate of the nonlinear trajectory usually referred to as the linearized state. The optimization process is described in figure 1.

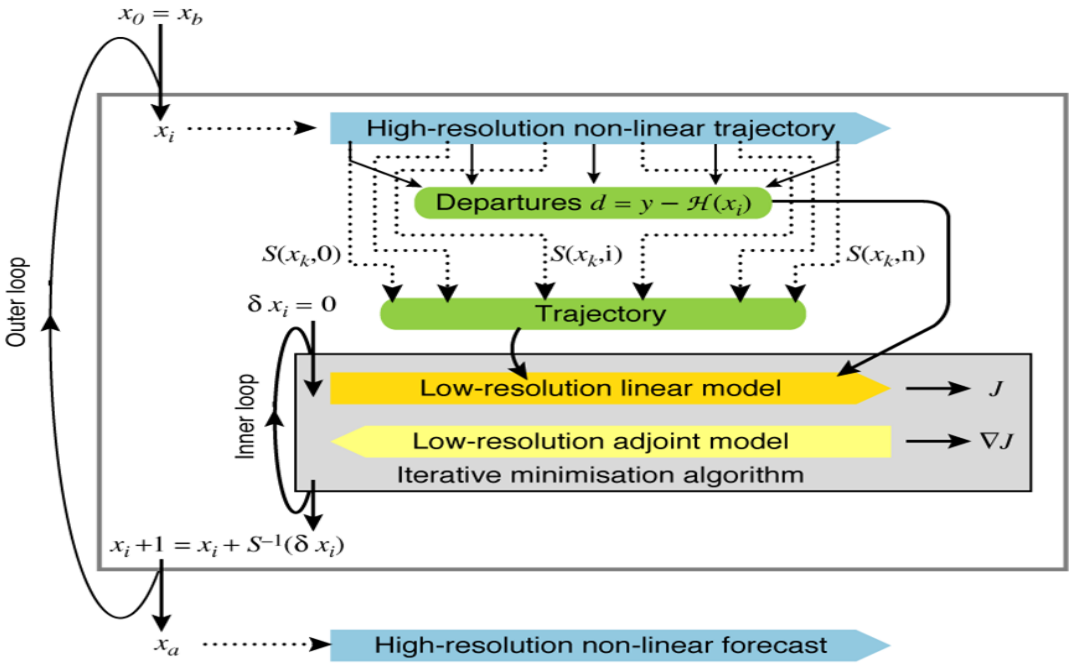


Figure 1: Incremental 4D-Var

The trajectory is obtained by integration of the linear model. The reference trajectory required by the linear and adjoint models comes from the background integration and is not updated at every iteration. Correspondingly, the iterative procedure of minimizing the incremental cost function is called the inner-loop which is much cheaper computationally to implement, due to the incremental 4-D VAR simplifications. When the quadratic cost function is approximated in this way, the 4-D VAR algorithm no longer converges to the solution of the original problem. The analysis increments are calculated at a reduced resolution and must be interpolated to the high-resolution model's grid. This drawback is partially overcome by executing after a number of inner-loops, a single outer-loop iteration which is updating the high-resolution reference trajectory and the observation departures. Correspondingly, the iterative procedure of minimizing the incremental cost function is called the outer-loop. After each outer-loop update, it is possible to use a progressively higher resolution for the inner-loops. Such a procedure was carried out in a multi-incremental algo-

rithm proposed by Veersé and Thepaut [89]. The incremental method was shown by Lawless [46] to be equivalent to an inexact Gauss-Newton method applied to the original nonlinear cost function. The outer-loop iterations can be shown to be locally convergent under certain conditions, provided that the inner-loop minimization is solved with sufficient accuracy (see, e.g., Gratton, Lawless, and Nichols [33]). In practice, however, very few outer-loop steps are performed, typically three. The inclusion of full physics in the adjoint model requires the 4-D VAR algorithm to overcome the negative effect of strong nonlinearities present in physics parametrization packages while being able to take advantage of the positive aspects resulting from the consistency between the forecasting nonlinear model and adjoint model. Several approaches have been proposed for mitigating the negative effect of strong nonlinearities in physical processes included in the adjoint model. These approaches involved either direct modifications or simplifications to physical parameterizations. Županski and Mesinger [98] and Tsuyuki [88] showed beneficial effects when smoothing formulas are used to replace those with discontinuities. The ECMWF system uses simplified physics in the adjoint model, although modifications or simplifications may lead to inconsistencies between the nonlinear forecasting model and the corresponding adjoint model.

In a further, multi-incremental, extension of the incremental 4D VAR method, the inner-loop resolution is increased after each iteration of the outer- loop. In particular, the information about the shape of the cost-function obtained during the early low-resolution iterations provides a very effective preconditioner for subsequent iterations at higher resolution, thus reducing the number of costly iterations. The inner-loops can be efficiently minimized using the conjugate gradient method, provided the cost-function is quadratic, i.e. when the operators involved in the definition of the cost function (the model and the observation operators) are linear. For this reason, the inner-loops have been completely linearized; the non-linear effects are all gathered at the outer-loop level.

9 Developments in variational data assimilation in last 2 decades

Starting with the advent of the incremental model a tremendous amount of research efforts focused on implementation of 4-D Var at operational centers. The principle of four-dimensional variational (4D-Var) assimilation usually assumes implicitly that the forecast model is ”perfect” within the assimilation window an approach referred to as strong constraint and looks for the model trajectory which best fits the data (background and observations) over the time window. Such a data assimilation method has been implemented in the last few years at various NWP centres with substantial benefit(Rabier [69]). To name a few, Meteo-France, UK Met-Office and Canadian Environment service, led all by pioneering work at ECMWF. However the next step was to consider observation bias, observation error correlation and model error (bias and random)in weak constraint 4-D VAR. See work of Trémolet [86, 87] and Akella and Navon [1].

For Gaussian, temporally-uncorrelated model error, the weak-constrained 4D-Var cost function is The cost function assumes the following form for Gaussian, temporally uncorrelated model error.

$$
\begin{aligned}
J(X) = & \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) \\
& + \frac{1}{2}\sum_{i=0}^{n}[\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i]^T \mathbf{R}^{-1}[\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i] \\
& + \frac{1}{2}\sum_{i=0}^{n}[\mathbf{x}_i) - \mathcal{M}_i(\mathbf{x}_{i-1})]^T \mathbf{Q}^{-1}[\mathbf{x}_i) - \mathcal{M}_i(\mathbf{x}_{i-1})].
\end{aligned} \tag{9.1}
$$

The matrix $\mathbf{Q}$ is taken usually to be proportional to $\mathbf{B}$. The pioneering work in weak constraint 4-D VAR is considered to be the one of Derber [21]. This continues to be an area of active research.

9.1 Estimation of background and observation error covariances

Modelling and specification of the covariance matrix of background error constitute important components of any data assimilation system. The main attributes of the background error covariance matrix $\mathbf{B}$ are:

- To spread out the information from the observations; correlations in the background covariance matrix will perform spatial spreading of information from observation points to a finite domain surrounding them;
- To provide statistically consistent increments at the neighboring grid points and levels of the model;
- To ensure that observations of one model variable (e.g., temperature) produce dynamically consistent increments in the other model variables (e.g. vorticity and divergence).For operational models, a typical background covariance matrix contains 10^7 x 10^7 elements. Therefore, non-essential components of this important covariance matrix may need to be neglected in order to produce a computationally feasible algorithm.

Construction of background error covariances has been addressed in the literature by the so-called "innovation method", in which the background errors are assumed to be independent of observation errors. The so-called NMC method was introduced by Parrish and Derber [66] as a surrogate for samples of background error using differences between forecasts of different length that verify at the same time. The ensemble method for constructing background covariances was proposed by Fisher [26], while Ingleby [38] proposed using statistical structures of forecast errors. One can attempt to disentangle information about the statistics of background error from the available information (innovation statistics), or one can try to find a surrogate quantity whose error statistics can be argued to be similar to those of the unknown background errors.

9.2 Observation error covariance

The problem of variational data assimilation for nonlinear evolution model can be formulated as an optimal control problem to find the initial condition, boundary conditions and/or model parameters. The input data contain observation and background errors, hence there is an error in the optimal solution. For mildly nonlinear dynamics, the covariance matrix of the optimal solution error can be approximated by the inverse Hessian of the cost function w.r.t control variables. For problems characterized by strongly nonlinear dynamics, a new statistical method based on the computation of a sample of inverse Hessians was suggested. This method relies on the efficient computation of the inverse Hessian by means of iterative methods (Lanczos and quasi-Newton BFGS) with preconditioning (Shutyaev, Gejadze, Copeland, and Dimet [81], Le Dimet, Navon, and Daescu [50].

Adjoint-based methods makes forecast sensitivity to data assimilation system input parameters $[\mathbf{y}, \mathbf{R}, \mathbf{x}_b, \mathbf{B}]$ possible. Forecast sensitivity to observations (FSO) – is used to monitor the impact of observations to reduce short-range forecast errors. In particular, forecast R-sensitivity (Daescu and Todling [20], Daescu and Langland [19]) may be used to provide guidance to error covariance tuning procedures. The sensitivity of a scalar measure of forecast error is computed with respect to changes to a set of covariance parameters (Lupu et al. [57]). Forecast R- and B-sensitivities can provide guidance toward the real covariance matrices. The method may show if background information is being over (or under) weighted. In this case it appears the Ensemble Data Assimilation (EDA) based background errors are overweighting the background.

10 Hybrid Data assimilation

Hybrid Data assimilation is a practical feasible way to introduce flow dependence in the background error covariances required for sequential or variational data assimilation. Starting with Lorenc [55], Whitaker and Hamill [96], Buehner, Houtekamer, Charette, Mitchell, and He [10] it was shown that combining the time-varying background error covariance derived from an ensemble of forecasts with stationary, climatological background error covariance leads to improvements. The resulting procedure is so-called, hybrid data assimilation system. Several operational numerical weather prediction centers use three- or four-dimensional variational (3D/4DVar) techniques and have implemented hybrid approaches in these contexts. The hybrid data assimilation involves developing hybrid covariance models, i.e. a linear combination of a static B matrix (built from climatology and typically used in 4D-Var applications) with a flow-dependent B matrix (described using an ensemble). This hybrid approach has been operational at ECMWF for some time(Buizza, Leutbecher, and Isaksen [11];, Isaksen, Bonavita, Buizza, Fisher, Haseler, Leutbecher, and Raynaud [39], Bonavita, Isaksen, and Hólm [7]), and is now operational at the Met Office, UK, for their global model (Clayton, Lorenc, and Barker [16]) and at Environment Canada (Buehner, Houtekamer, Charette, Mitchell, and He [9]). A theoretical basis for the construction of the hybrid covariances, in particular how to weigh static and flow-dependent components, is described by Bishop and Satterfield [5] and Bishop, Satterfield, and Shanley [6].

11 Numerical Experiments

This section focuses on non-linear strong constraint 4D-Var experiments and it is divided in two parts. The first one centers on numerical simulations using 1D-Burgers model while the second part concentrates on computer simulations of shallow water equations model. As discussed, the adjoint models are required to alleviate the computational complexity of estimating the gradient during the optimization routines. It is known that such models have validity regions depending on the amount of perturbation considered as input. To asses the quality of adjoint models, tangent linear and adjoint tests are performed. Then the potential of 4D-Var method is discussed, its objective function and associated gradient as well as the analysis errors with respect to the observations being examined.

11.1 Burgers Model

Burgers' equation is an important partial differential equation from fluid mechanics [12]. The evolution of the velocity u of a fluid evolves according to

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \mu\frac{\partial^2 u}{\partial x^2}, \quad x \in [-1, 1], \quad t \in (0, 0.2]. \tag{11.1}$$

Here μ denotes the viscosity coefficient. The model has homogeneous Dirichlet boundary conditions $u(-1, t) = u(1, t) = 0$, and the integration time is $t \in (0, 0.2]$. An Euler explicit scheme is implemented using a spatial mesh of $nx = 41$ equidistant points on [-1, 1], with $\Delta x = 0.05$. A uniform temporal mesh with $nt = 21$ points covers the interval $[0, 0.02]$, with $\Delta t = 0.001$. A set of initial conditions is depicted in Figure 2 together with the final solution obtained after integrating the discrete Burgers model (11.1) in time.

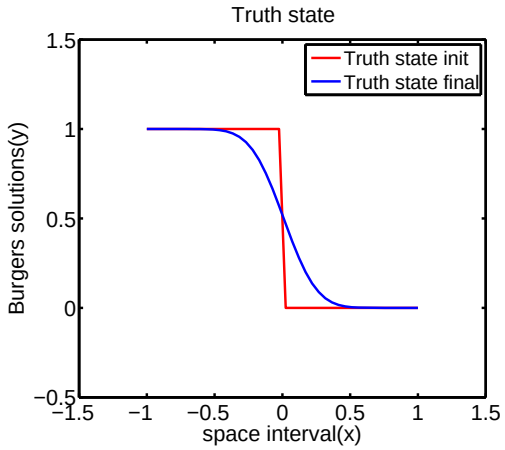


Figure 2: Initial and final true states

For our data assimilation experiments, we add uniform random perturbations $\varepsilon \in U(-0.5, 0.5)$ to the above truth state initial conditions and generate twin-experiment observations at every grid space point location and every time step. The background state or the first guess for the 4D-Var simulations is shown in figure 3 along with the final time solution. The background and observation error covariance matrices are taken to be identity matrices.

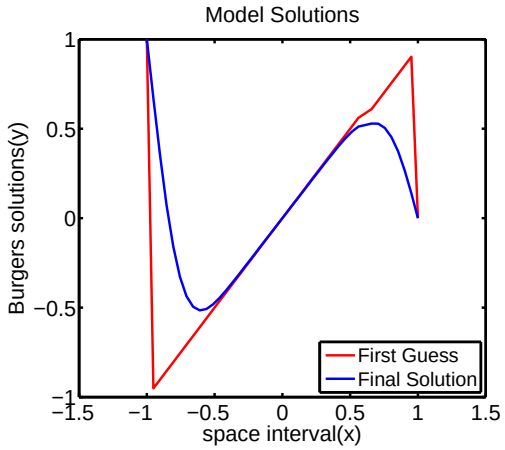


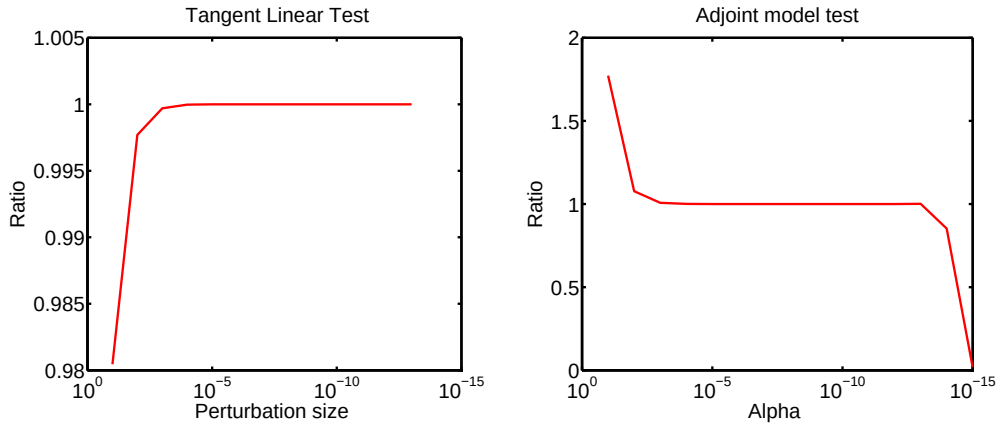
Figure 3: Initial and Final Solutions of Burgers Model

The objective function of lack of fit between model forecast and observations is minimized using the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (LBFGS) introduced by Liu and Nocedal [53]. This optimization algorithm is a member of quasi-Newton methods and requires the gradient estimation of the 4D-Var objective function. As mentioned earlier, the fastest approach for computing the associated gradient is to employ an adjoint model. Before using it inside the LBFGS algorithm, it is necessary to verify the accuracy of the tangent linear and adjoint models by checking their output agreement with the finite difference approximations. These linearity tests used in this study are derived from the alpha test described by Navon, Zou, Derber, and Sela [61, eq. (2.20)].). The depicted values from figure 4 are obtained using

$$\mathrm{adj}_{\mathrm{test}} = \frac{\mathrm{J}(\mathbf{x}_0 + \alpha\delta\mathbf{x}_0) - \mathrm{J}(\mathbf{x}_0)}{< \nabla \mathrm{J}(\mathbf{x}_0), \alpha\delta\mathbf{x}_0 >_2}, \quad \mathrm{tl}_{\mathrm{test}} = \frac{\|\mathcal{M}_{0,\mathrm{t_f}}(\mathbf{x}_0 + \alpha\delta\mathbf{x}_0)(\mathrm{t_f}) - \mathcal{M}_{0,\mathrm{t_f}}(\mathbf{x}_0)(\mathrm{t_f})\|_2}{\|\mathbf{M}_{0,\mathrm{t_f}}(\alpha\delta\tilde{\mathbf{x}}_0)(\mathrm{t_f})\|_2}, \tag{11.2}$$

where α represented on the x axis in figure 4 controls the magnitude of the perturbation. The forward and tangent linear models denoted by $\mathcal{M}$ and $\mathbf{M}$ are integrated along the entire time interval. The test formulations (11.2) uses t_f to represent the final time which in our case was set to $t_f = 0.2$ and we employed the Euclidian scalar product and norm.

Figure 5 depicts the minimization performances of the 4D-Var Burgers data assimilation system. The cost function and gradient values are normalized by dividing them with their initial values. After no more than 50 iterations the optimization routine stopped reaching a gradient value smaller than 10^{-4}.



(a) Tangent Linear

(b) Adjoint Model

Figure 4: Linearization tests validating the usage of adjoint and tangent linear models for certain perturbations ranges

The analysis is obtained and its quality is measured in terms of Euclidian distance away from the observations. In comparison with the first guess, we notice a significant improvement in the error magnitude as shown in figure 6. This underlines the performance of the variational 4D-Var algorithm justifying its large scale usage at numerical weather prediction centers around the globe in the form of incremental 4D-Var.

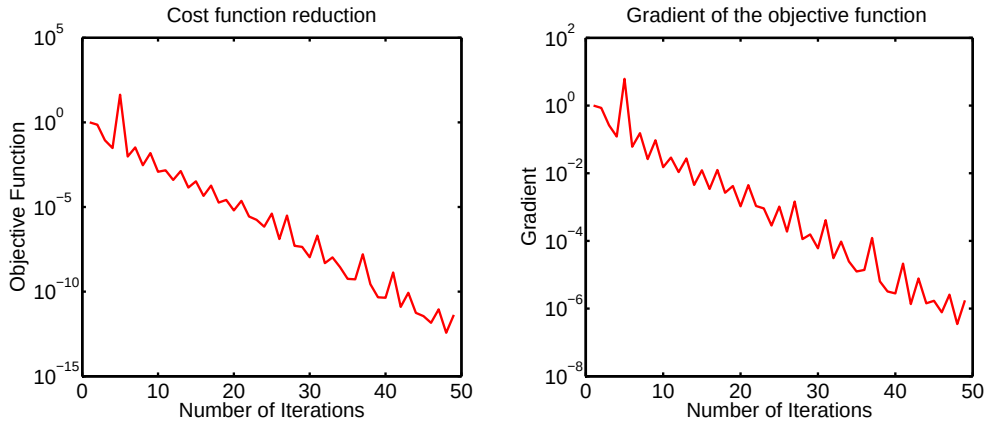
11.2 Shallow Water Equations Model

SWE has proved its capabilities in modeling propagation of Rossby waves in the atmosphere, rivers, lakes and oceans as well as gravity waves in a smaller domain. The alternating direction fully implicit finite difference scheme Gustafsson [35] is considered in this chapter and it is stable for large CFL condition numbers (we tested the stability of the scheme for a CFL condition number equal up to 8.9301). We also refer to Fairweather and Navon [25], Navon and Villiers [60] for other research work on this topic.

The SWE model using the β-plane approximation on a rectangular domain is introduced (see Gustafsson [35])

$$\frac{\partial w}{\partial t} = A(w)\frac{\partial w}{\partial x} + B(w)\frac{\partial w}{\partial y} + C(y)w, \quad (x,y) \in [0,L] \times [0,D], \quad t \in (0, t_{\mathrm{f}}], \tag{11.3}$$

where $w = (u, v, \phi)^T$ is a vector function, u, v are the velocity components in the x and y directions, respectively, h is the depth of the fluid, g is the acceleration due to gravity, and $\phi = 2\sqrt{gh}$.



(a) Cost Function (b) Gradient of the Cost Function

Figure 5: Reduction in the cost function and its associated gradient

The matrices A, B and C are assuming the form

$$A = -\begin{pmatrix} u & 0 & \phi/2 \\ 0 & u & 0 \\ \phi/2 & 0 & u \end{pmatrix}, \quad B = -\begin{pmatrix} v & 0 & 0 \\ 0 & v & \phi/2 \\ 0 & \phi/2 & v \end{pmatrix}, \quad C = \begin{pmatrix} 0 & f & 0 \\ -f & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where f is the Coriolis term.

$$f = \widehat{f} + \beta(y - D/2), \quad \beta = \frac{\partial f}{\partial y}, \quad \forall\, y,$$

with $\widehat{f}$ and β constants.

We assume periodic solutions in the x direction for all three state variables while in the y direction

$$v(x, 0, t) = v(x, D, t) = 0, \quad x \in [0, L], \quad t \in (0, t_{\mathrm{f}}]$$

and Neumann boundary conditions are considered for u and ϕ.

Initially $w(x, y, 0) = \psi(x, y), \quad \psi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \quad (x, y) \in [0, L] \times [0, D]$. Now we introduce a mesh of $n = N_x \cdot N_y$ equidistant grid points on $[0, L] \times [0, D]$, with $\Delta x = L/(N_x - 1), \quad \Delta y = D/(N_y - 1)$. We also discretize the time interval $[0, t_{\mathrm{f}}]$ using N_t equally distributed points and $\Delta t = t_{\mathrm{f}}/(N_t - 1)$. Next we define vectors of the unknown variables of dimension n containing approximate solutions such as

$$\boldsymbol{w}(t_N) \approx [w(x_i, y_j, t_N)]_{i=1,2,..,N_x, \quad j=1,2,..,N_y} \in \mathbb{R}^n, \quad N = 1, 2, ..N_t.$$

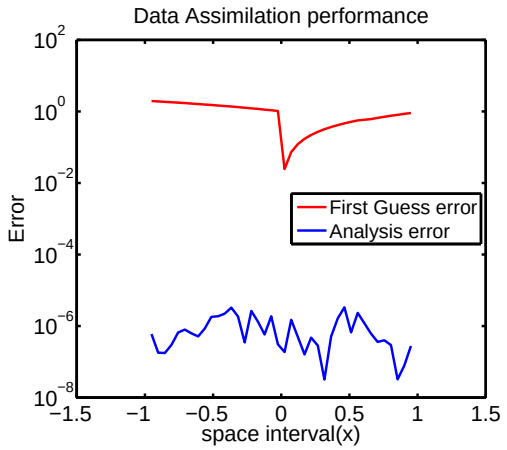


Figure 6: First guess and optimal initial conditions errors

The semi-discrete equations of SWE (11.3) are:

$$\mathbf{u}' = -F_{11}(\mathbf{u}) - F_{12}(\boldsymbol{\phi}) - F_{13}(\mathbf{u},\mathbf{v}) + \mathbf{F}\odot\mathbf{v}, \tag{11.4}$$

$$\mathbf{v}' = -F_{21}(\mathbf{u}) - F_{22}(\mathbf{v}) - F_{23}(\boldsymbol{\phi}) - \mathbf{F}\odot\mathbf{u}, \tag{11.5}$$

$$\boldsymbol{\phi}' = -F_{31}(\mathbf{u},\boldsymbol{\phi}) - F_{32}(\mathbf{u},\boldsymbol{\phi}) - F_{33}(\mathbf{v},\boldsymbol{\phi}) - F_{34}(\mathbf{v},\boldsymbol{\phi}), \tag{11.6}$$

where $\mathbf{u}'$, $\mathbf{v}'$, $\boldsymbol{\phi}'$ denote semi-discrete time derivatives, $\mathbf{F}\in\mathbb{R}^n$ stores Coriolis components, $\odot$ is the component-wise multiplication operator, while the nonlinear terms $F_{i,j}$ are defined as follows:

$$\begin{aligned}
&F_{11},F_{12},F_{21},F_{23}:\mathbb{R}^n\to\mathbb{R}^n,\quad F_{13},F_{22},F_{3i}:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}^n,\quad i=1,2,3,4,\\
&F_{11}(\mathbf{u})=\boldsymbol{u}\odot A_x\boldsymbol{u},\quad F_{12}(\boldsymbol{\phi})=\frac{1}{2}\boldsymbol{\phi}\odot A_x\boldsymbol{\phi},\quad F_{13}(\boldsymbol{u},\boldsymbol{v})=\boldsymbol{v}\odot A_y\boldsymbol{u},\\
&F_{21}(\boldsymbol{u},\boldsymbol{v})=\boldsymbol{u}\odot A_x\boldsymbol{v},\quad F_{22}(\boldsymbol{v})=\boldsymbol{v}\odot A_y\boldsymbol{v};\quad F_{23}(\boldsymbol{\phi})=\frac{1}{2}\boldsymbol{\phi}\odot A_y\boldsymbol{\phi},\\
&F_{31}(\boldsymbol{u},\boldsymbol{\phi})=\frac{1}{2}\boldsymbol{\phi}\odot A_x\boldsymbol{u},\quad F_{32}(\boldsymbol{u},\boldsymbol{\phi})=\boldsymbol{u}\odot A_x\boldsymbol{\phi},\\
&F_{33}(\boldsymbol{v},\boldsymbol{\phi})=\frac{1}{2}\boldsymbol{\phi}\odot A_y\boldsymbol{v},\quad F_{34}(\boldsymbol{v},\boldsymbol{\phi})=\boldsymbol{v}\odot A_y\boldsymbol{\phi},
\end{aligned} \tag{11.7}$$

where $A_x, A_y\in\mathbb{R}^{n\times n}$ are constant coefficient matrices for discrete first-order and second-order differential operators which incorporate the boundary conditions. The numerical scheme was implemented in Fortran and uses a sparse matrix environment. For operations with sparse matrices we employed SPARSEKIT library Saad [72]and the sparse linear systems obtained during the quasi-Newton iterations were solved using MGMRES library Barrett et al. [3], Kelley [41], Saad [73]. Here we did not decouple the model equations as in Stefanescu and Navon [18] where the Jacobian is either block cyclic tridiagonal orblock tridiagonal. By keeping all discrete equations together the

corresponding SWE adjoint model can be solved with the same implicit scheme used for forward model.

For the 4D-Var numerical experiments we use the following constants $L = 6000km,\ D = 4400km,\ t_f = 3h,\ \widehat{f} = 10^{-4}s^{-1},\ \beta = 1.5 \cdot 10^{-11}s^{-1}m^{-1},\ g = 10ms^{-2},\ H_0 = 2000m,\ H_1 = 220m,\ H_2 = 133m$. The mesh coordinates are $N_x = 31$, $N_y = 21$ and $N_t = 91$. Next we derived the initial conditions from the initial height condition No. 1 of Grammeltvedt [32] i.e.

$$h(x,y) = H_0 + H_1 + \tanh\left(9\frac{D/2-y}{2D}\right) + H_2\text{sech}^2\left(9\frac{D/2-y}{2D}\right)\sin\left(\frac{2\pi x}{L}\right),$$
$$0 \le x \le L,\ 0 \le y \le D.$$

The initial velocity fields were then obtained from the initial height field using the geostrophic relationship

$$u = \left(\frac{-g}{f}\right)\frac{\partial h}{\partial y},\ v = \left(\frac{g}{f}\right)\frac{\partial h}{\partial x}.$$

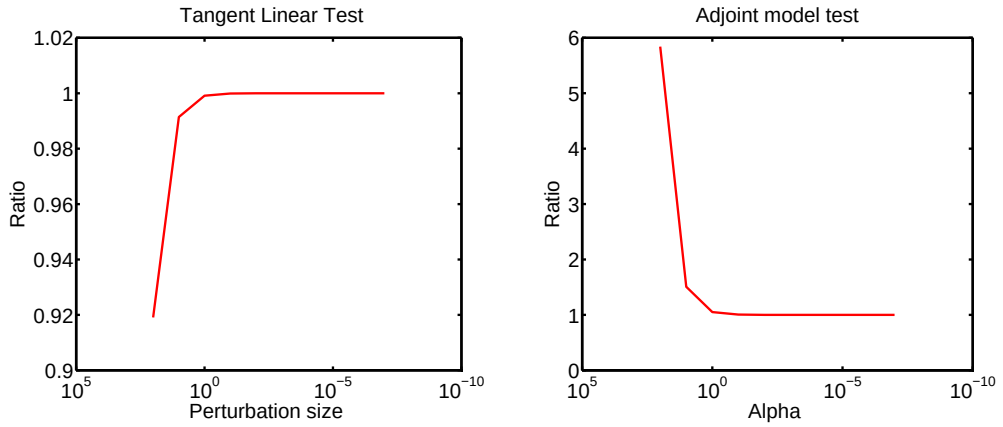
The background and observation error covariance matrices are taken diagonal. The associated background variances are set to $4,\ 6$ and 55 for velocity components u and v and geopotential height ϕ. The observations variances are taken constant for the entire time interval with 3 and 5 associated to velocity components while 50 is used in the case of geopotential height. The corresponding objective function has the following form:

$$J(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{w}_0^{\text{b}} - \mathbf{w}_0\right)^T\mathbf{B}_0^{-1}\left(\mathbf{w}_0^{\text{b}} - \mathbf{w}_0\right) + \frac{1}{2}\sum_{i=0}^{N_{obs}}\left(\mathbf{y}_i - H_i\left(\mathbf{w}_i\right)\right)^T\mathbf{R}^{-1}\left(\mathbf{y}_i - H\left(\mathbf{w}_i\right)\right), \quad (11.8)$$

where $\mathbf{w}_0^{\text{b}}$ is the background state. The observations are not available at every time step but only at 10 locations inside the time interval, i.e. $1,\ 10,\ 19,\ 28,\ 37,\ 46,\ 55,\ 64,\ 73,\ 91$. The observation operators H_i are taken as identities meaning that we are observing the state variables only. These observations are obtained by perturbing the trajectory associated with Grammeltvedt conditions. The additive observation noise is set to $0.1 \cdot \varepsilon_o$, where ε_o is sampled from a multivariate distribution $\mathcal{N}(\mathbf{0}, \mathbf{R})$. The background state is obtained by adding normal noise $3 \cdot \varepsilon_b$, where $\varepsilon_b \in \mathcal{N}(\mathbf{0}, \mathbf{B})$. The length of the assimilation window is selected to be $3h$. The implicit scheme allowed us to integrate in time using a larger time step and select $N_t = 91$ time steps.

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization method option contained in the CONMIN software (Shanno and Phua [80]) is employed for the SWE 4-D VAR. BFGS uses a line search method which is globally convergent in the sense that $lim_{k\to\infty}\|\nabla J^{(k)}\| = 0$ and utilizes approximate Hessians to include convergence to a local minimum. The discrete tangent linear and adjoint models were derived by hand and their accuracy was verified using Navon et al. [61] techniques and the results are depicted in figure 7.

The performances of SWE 4D-Var data assimilation system are presented in Figure 8. The square norm of the gradient is shown in panel (b). The optimization routine stopped after 60 iterations when the local criterium $\nabla J^{(k)} < 10^{-4}$ was satisfied.



(a) Tangent Linear

(b) Adjoint Model

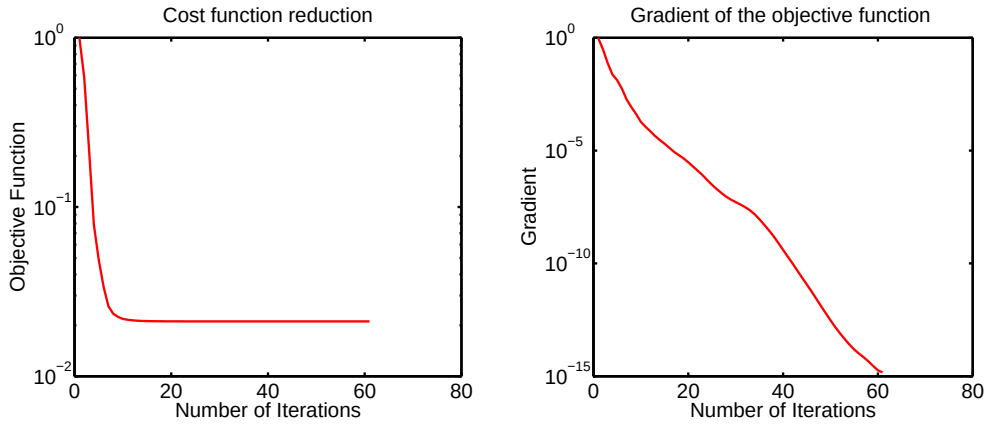
Figure 7: Swallow Water Equations tangent linear and adjoint models tests

Figure 9 depicts the analysis geostrophic wind field in comparison with the first guess field. Clearly the algorithm was able to recover more accurate state variables from the observations as evidenced from the truth wind field shown in panel (a). Similar results are obtained for the geopotential analysis as shown in figure 10.

The applications of reduced order modeling techniques have the potential to significantly speed up the solution of variational data assimilation problems with nonlinear dynamical models as shown by Ştefănescu, Sandu, and Navon [83]. This could represent a very important step for obtaining analyses faster than real time at numerical weather prediction centers. Moreover it has been proved that the solution of reduced optimization system equipped with a trust region algorithm converges to the high-fidelity analyses [2] which is not the case with the solution of incremental 4D-Var [86].

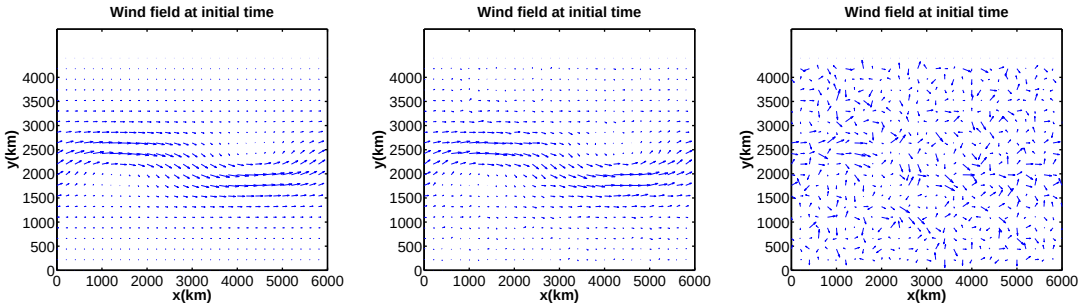
12 Outlook of Modern Data Assimilation topics

The 4D-VAR data assimilation is available and implemented today at several operational numerical weather prediction centers starting with European Centre for Medium- Range Weather Forecasts (ECMWF), (Rabier, Järvinen, Klinker, Mahfouf, and Simmons [70]; Klinker, Rabier, Kelly, and Mahfouf [42]) while a similar system was operational at Meteo-France in 2000 (Janisková, Thépaut, and Geleyn [40]; Gauthier and Thepaut [29]; Desroziers, Hello, and Thepaut [22]). More recently 4-D VAR was implemented at UK Met office, Japan and Canada. Park and Županski [65] survey the status and progress of the four-dimensional variational data assimilation with emphasis on application to prediction of mesoscale/ storm-scale atmospheric phenomena. See also Zupanski, Zupanski, Rogers, Parrish, and DiMego [99]. The impact of adopting 4-D VAR was qualified as substantial, resulting in an improvement in NWP quality and accuracy (see Rabier [69] in special Issue of QJRMS 2005). 4-D VAR combined with improvement in error specifications and with a



(a) Cost Function (b) Gradient of the Cost Function

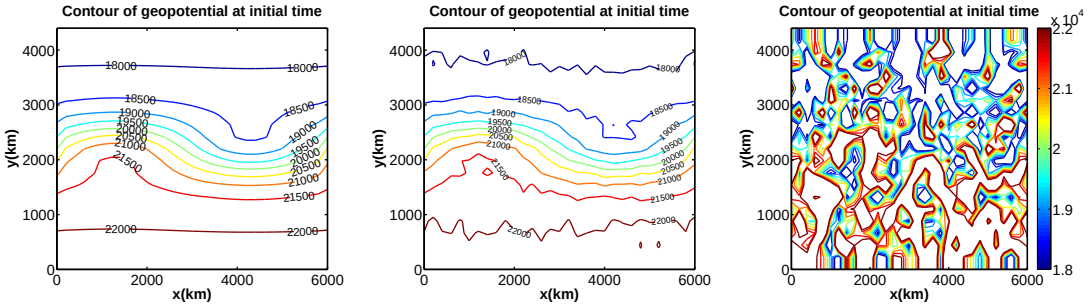
Figure 8: Reduction in the cost function and its associated gradient



(a) Reference (b) Analysis (c) First Guess

Figure 9: Initial conditions of wind before and after data assimilation

large increase in a variety of observations has led to improvements in NWP accuracy (Simmons and Hollingsworth [82]). Hollingsworth et al. [37] show how observing system improvements led to improvements of forecast scores while Bouttier and Kelly [8] show that the improvement of forecast scores for the southern hemisphere are due to satellite data. Also, error statistics for different sources of observation constitutes an active field of research aimed mainly at obtaining better representation of the specific observation operators. It has become amply evident that in the last 15 years major improvements in NWP are due to large extent to development of sources of observations and that 4-D VAR and sequential data assimilation can take advantage of them due to major research efforts at universities, federal laboratories and operational centers. For new opportunities for research see the article by McLaughlin, O'Neill, Derber, and Kamachi [59] that illuminates and outlines possibilities for enhanced collaboration within the data assimilation community. It is certain that data assimilation concepts have become widely applied in all the geosciences as more



(a) Reference (b) Analysis (c) First Guess

Figure 10: Geopotential initial conditions before and after data assimilation

geoscience scientific disciplines gain access to larger amounts of data, from satellite remote sensing and from sensor networks, and as Earth system models increase in both accuracy and sophistication.

12.1 Data assimilation applied to other fields

Data assimilation methods are currently also used in other environmental forecasting problems, e.g. in hydrological forecasting. Basically, the same types of data assimilation methods as those described above are in use there. An example of chemical data assimilation using AUTOCHEM can be found at CDA Central. Given the abundance of spacecraft data for other planets in the solar system, data assimilation is now also applied beyond the Earth to obtain re-analyses of the atmospheric state of extra-terrestrial planets. Mars is the only extraterrestrial planet to which data assimilation has been applied so far. Available spacecraft data include, in particular, retrievals of temperature and dust/water ice optical thicknesses from the Thermal Emission Spectrometer on board NASA's Mars Global Surveyor and the Mars Climate Sounder on board NASA's Mars Reconnaissance Orbiter. Two methods of data assimilation have been applied to these datasets: an Analysis Correction scheme and two Ensemble Kalman Filter schemes. Both are using a global circulation model of the Martian atmosphere as forward model. The Mars Analysis Correction Data Assimilation (MACDA) dataset is publicly available from the British Atmospheric Data Centre. Data assimilation is now a part of the challenge for every forecasting problem encompassing multi-physics multi-scale systems. Dealing with biased data is a serious challenge in data assimilation. Further development of methods to deal with biases will be of particular use. If there are several instruments observing the same variable then intercomparing them using probability distribution functions can be useful. Other uses include trajectory estimation for the Apollo program, GPS, and atmospheric chemistry. A particular application is the prediction of future oil production. Data assimilation is extensively used in petroleum reservoir engineering, where it is usually referred to as "history matching". Data assimilation methods are used for uncertainty assessment of performance predictions of wells in oil reservoirs and for generating computational models used for optimizing

decision parameters that would improve oil recovery. Recently data assimilation has been extended to blood circulation in hemodynamics to determine and analyze the blood flow patterns in the aortic root - since flow reconstruction by image processing is not accurate enough

12.2 Further applications of Variational Data Assimilation

At the beginning VDA methods were applied to meteorology, first to mesoscale and then to global models. In a second phase VDA methods were applied to oceanography when optimal interpolation methods became unable to retrieve physically consistent fields'

During this period the observation of the Earth has experienced a major improvement due the advent of satellites.

Neither for the atmosphere nor for the ocean are satellites observing the state variables of the mathematical models, rather what is basically observed and measured are radiances , which are indirectly linked to the state variables, by solving inverse problems and this is the purpose of variational methods, therefore spatial observations were naturally introduced in VDA.

12.2.1 Data assimilation for continental waters

The evolution of rivers is of great importance especially for flood prediction. Going back to the ingredients of VDA, what are they in the framework of continental waters:

- Models: Basically we have to deal with the equations of conservation derived from fluids dynamics. For some rivers such as the Yangtze river,the content of the sediment flow has to be taken into account. There are difficulties to define the geometry of the domain both on the lateral boundaries and also on the bottom of the river which is poorly known in practical applications.
- Sink and source terms, these terms are of various nature : rain, infiltration, sources, they influence the geometry of the domain and most of the time they are poorly known and subject to nonlinearities like for instance the saturation of the soil.
- Data: many rivers are not equipped with sensors, the cost of operational observations is very expensive and furthermore they not accessible to satellite measurements.
- Statistics : Very poor for the case of extreme events

Therefore the framework of VDA as it is applied in meteorology is very difficult to transpose to hydrology (Vieux, LeDimet, and Armand [90]), nevertheless VDA is useful for model calibration and sensitivity analysis.

12.2.2 Data assimilation in agronomy

In the last few years, encouraging results using radiative transfer model inversion techniques were obtained for land biophysical variables retrieval. However, the inversion of radiative transfer models is a severely ill-posed problem, that may lead to significant uncertainties in the biophysical variables estimates. Improvement of performance of the inversion process requires additional information to be exploited by including better radiative transfer models, exploitation of proper prior information on the distribution of the canopy and atmosphere variables, knowledge of uncertainties in satellite measurements, as well as possible spatial and temporal constraints.

In their paper, Lauvernet, Baret, Hascoët, Buis, and Le Dimet [45] focus on the use of coupled atmosphere-surface radiative transfer models (SMAC+SAIL+PROSPECT) to estimate some key biophysical variables from top of atmosphere canopy reflectance data. The inversion is achieved over an ensemble of pixels belonging to a spatial window where aerosol properties are assumed to be constant, and over a temporal window of few days where the vegetation state is assumed not to vary. The ensemble inversion scheme accounting for the spatial and temporal constraints is described. Top of the atmosphere reflectance observations are simulated for 13 spectral bands within the visible and near infrared domains. The coupled model is inverted with a variational method implementation aimed at solving very large inverse problems. It is based on the use of the adjoint model and a Quasi-Newton optimisation technique with a BFGS update.

12.2.3 Data Assimilation for plant growth

Functional-Structural plant models (FSPM) combine process-based models and architectural models for a better description of plant growth. The process-based models characterize plants mechanics like photosynthesis for agronomic applications.

By contrast, the architectural models were originally developed to analyze botanic patterns and/or topological structures . A typical FSPM can thus simulate not only plant organogenesis but also biomass production and partition at organ level (leaf, fruit, internode). Wu, Le Dimet, De Reffye, Hu, Cournède, and Kang [97] have used the FSPM GreenLab to model and optimize plant growth thanks to a dynamical system. This growth algorithm is based on a minimal set of physiological knowledge, such as the empirical rules of plant-environment interactions for biomass acquisition and the source-sink relations among organs that compete for assimilates.

Consequently the plant morphological plasticity can be described by a small set of endogenous parameters , thus reducing the complexity of parameters calibration. The model takes the form of a discrete non linear dynamical system, a difficulty arises from the fact that the nonlinearity is partly due to the biological thresholding or saturation effects. An optimal control method has been applied for plant functional-structural growth. Using variational methods based on optimal control two problems have been investigated, the calibration of models by minimizing the discrepancy between observations and computed solutions of the Greenlab model and also by solving an optimal water supply problem applied to growth of sunflower plants. The classical tools of Variational Data Assimilation, such as the adjoint model and optimization algorithms have been employed .

12.2.4 Assimilation of Images

The observation of the earth by satellites is an important source of information if we consider the dynamics of the flows: the evolution of fronts and /or storms provides an intuitive overview of the weather to come. This remark is also true for the ocean with temperature, salinity or the color of the ocean due to biological (algae) activities. From the dynamical point of view the information comes from the evolution of the discontinuities in the images. An important question is what is the nature of these images, basically we have two different phenomena:

- Lagrangian images : this is the case of small cumulus clouds under the tropics. In operational meteorology they are considered as lagrangian tracers , their velocity is estimated then plugged in a classical data assimilation scheme.
- Eulerian images : this is the case of lenticular clouds over a mountain, they looks almost steady state and nevertheless they are the signature of a strong meteorological feature. Therefore an estimation of the wind velocity based on the evolution of this clouds would give a wrong evaluation.

Therefore the problem is how to couple the dynamics of images with numerical models ? A functional space has to be defined for the images and also an operator from the space of solutions of the model towards the space of the images. Because the information is in the discontinuities of the images, the metrics should not be too much regularizing; the choice of adequate metrics remains an open problem (Le Dimet, Souopgui, Titaud, and Shutyaev [47]).

12.2.5 Data Assimilation in medicine

For a living person some direct measurement cannot be direclty carried out : this is the case of the heart. There exist several mathematical models for the heart counting on several parameters specific of the heart. The methods of data assimilation and the assimilation of images have been used for heart models calibration (Chapelle, Fragu, Mallet, and Moireau [14]).

References

[1] S Akella and IM Navon. Different approaches to model error formulation in 4D-Var: a study with high-resolution advection schemes. *Tellus A*, 61(1):112–128, 2009.

[2] Eyal Arian, Marco Fahl, and Ekkehard W Sachs. Trust-region proper orthogonal decomposition for flow control. Technical report, DTIC Document, 2000.

[3] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition*. SIAM, Philadelphia, PA, 1994.

[4] DP Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic Press, New York, 1982.

[5] Craig H Bishop and Elizabeth A Satterfield. Hidden error variance theory. Part I: Exposition and analytic model. *Monthly Weather Review*, 141(5):1454–1468, 2013.

[6] Craig H Bishop, Elizabeth A Satterfield, and Kevin T Shanley. Hidden error variance theory. Part II: An instrument that reveals hidden error variance distributions from ensemble forecasts and observations. *Monthly Weather Review*, 141(5):1469–1483, 2013.

[7] Massimo Bonavita, Lars Isaksen, and Elías Hólm. On the use of EDA background error variances in the ECMWF 4D-Var. *Quarterly journal of the royal meteorological society*, 138 (667):1540–1559, 2012.

[8] François Bouttier and Graeme Kelly. Observing-system experiments in the ecmwf 4d-var data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 127(574): 1469–1488, 2001.

[9] Mark Buehner, PL Houtekamer, Cecilien Charette, Herschel L Mitchell, and Bin He. 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(5):1550–1566, 2010.

[10] Mark Buehner, PL Houtekamer, Cecilien Charette, Herschel L Mitchell, and Bin He. 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(5):1567–1586, 2010.

[11] Roberto Buizza, Martin Leutbecher, and Lars Isaksen. Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 134(637):2051–2066, 2008.

[12] Johannes Martinus Burgers. A mathematical model illustrating the theory of turbulence. *Adv. in Appl. Mech.*, 1:171–199, 1948.

[13] Dan G Cacuci and Matthew CG Hall. Efficient estimation of feedback effects with application to climate models. *Journal of the atmospheric sciences*, 41(13):2063–2068, 1984.

[14] Dominique Chapelle, Marc Fragu, Vivien Mallet, and Philippe Moireau. Fundamental principles of data assimilation underlying the Verdandi library: applications to biophysical model personalization within euHeart. *Medical & biological engineering & computing*, 51(11): 1221–1233, 2013.

[15] Jules G Charney, Ragnar Fjörtoft, and John Von Neumann. Numerical integration of the barotropic vorticity equation. *Tellus A*, 2(4), 1950.

[16] AM Clayton, Andrew C Lorenc, and Dale M Barker. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, 139(675):1445–1461, 2013.

[17] PHILIPPE Courtier, J-N Thépaut, and A Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387, 1994.

[18] R. Ştefănescu and I.M. Navon. POD/DEIM Nonlinear model order reduction of an ADI implicit shallow water equations model. *J. Comput. Phys.*, 237:95–114, 2013.

[19] Dacian N Daescu and Rolf H Langland. Error covariance sensitivity and impact estimation with adjoint 4d-var: theoretical aspects and first applications to navdas-ar. *Quarterly Journal of the Royal Meteorological Society*, 139(670):226–241, 2013.

[20] Dacian N Daescu and Ricardo Todling. Adjoint sensitivity of the model forecast to data assimilation system error covariance parameters. *Quarterly Journal of the Royal Meteorological Society*, 136(653):2000–2012, 2010.

[21] John C Derber. A variational continuous assimilation technique. *Monthly Weather Review*, 117(11):2437–2446, 1989.

[22] Gerald Desroziers, Gwenaelle Hello, and Jean-Noel Thepaut. A 4d-var re-analysis of fastex. *Quarterly Journal of the Royal Meteorological Society*, 129(589):1301–1315, 2003.

[23] Leonhard Euler. Elementa calculi variationum. *Originally published in Novi Commentarii academiae scientiarum Petropolitanae*, 10(1766):51–93, 1766.

[24] Leonhard Euler. 1744 methodus inveniendi lineas curvas maximi minimive proprietate gaudentes sive solutio problematis isoperimetrici latissimo sensu accepti. *Bousquet, Lausannae e Genevae, E65A. OO Ser. I*, 24, 1952.

[25] G. Fairweather and I.M. Navon. A linear ADI method for the shallow water equations. *Journal of Computational Physics*, 37:1–18, 1980.

[26] Mike Fisher. Background error covariance modelling. In *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*, pages 45–63, 2003.

[27] Roger Fletcher. *Practical methods of optimization*. John Wiley & Sons, 2013.

[28] B. G. Galerkin. Series occurring in various questions concerning the elastic equilibrium of rods and plates. *Vestnik Inzhenerov Tech.*, 19:897–908, 1915.

[29] Pierre Gauthier and Jean-Noel Thepaut. Impact of the digital filter as a weak constraint in the preoperational 4dvar assimilation system of météo-france. *Monthly Weather Review*, 129(8): 2089–2102, 2001.

[30] Michael Ghil, S Cohn, John Tavantzis, K Bube, and Eugene Isaacson. Applications of estimation theory to numerical weather prediction. In *Dynamic meteorology: Data assimilation methods*, pages 139–224. Springer, 1981.

[31] Philip E Gill, Walter Murray, and Margaret H Wright. Practical optimization. 1981.

[32] A. Grammeltvedt. A survey of finite difference schemes for the primitive equations for a barotropic fluid. *Monthly Weather Review*, 97(5):384–404, 1969.

[33] Serge Gratton, Amos S Lawless, and Nancy K Nichols. Approximate gauss-newton methods for nonlinear least squares problems. *SIAM Journal on Optimization*, 18(1):106–132, 2007.

[34] Andreas Griewank and Andrea Walther. *Evaluating derivatives: principles and techniques of algorithmic differentiation*. Siam, 2008.

[35] B. Gustafsson. An alternating direction implicit method for solving the shallow water equations. *Journal of Computational Physics*, 7:239–254, 1971.

[36] Matthew CG Hall, Dan G Cacuci, and Michael E Schlesinger. Sensitivity analysis of a radiative-convective model by the adjoint method. *Journal of the Atmospheric Sciences*, 39 (9):2038–2050, 1982.

[37] Anthony Hollingsworth, Sakari Uppala, Ernst Klinker, David Burridge, Frederic Vitart, Jeanette Onvlee, JW De Vries, Ad De Roo, and Christian Pfrang. The transformation of earth-system observations into information of socio-economic value in geoss. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3493–3512, 2005.

[38] N Bruce Ingleby. The statistical structure of forecast errors and its representation in the met. office global 3-d variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 127(571):209–231, 2001.

[39] Lars Isaksen, M Bonavita, R Buizza, M Fisher, J Haseler, M Leutbecher, and Laure Raynaud. *Ensemble of data assimilations at ECMWF*. European Centre for Medium-Range Weather Forecasts, 2010.

[40] Marta Janisková, Jean-Noël Thépaut, and Jean-François Geleyn. Simplified and regular physical parameterizations for incremental four-dimensional variational assimilation. *Monthly Weather Review*, 127(1):26–45, 1999.

[41] C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. Number 16 in Frontiers in Applied Mathematics. SIAM, 1995.

[42] Ernst Klinker, F Rabier, G Kelly, and J-F Mahfouf. The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration. *Quarterly Journal of the Royal Meteorological Society*, 126(564): 1191–1215, 2000.

[43] JL Lagrange. Application de la méthode exposée dans le mémoire précédent à la solution des problèmes de dynamique différents. *Œuvres de Lagrange (1867–1892)*, 1:151–316, 1762.

[44] Joseph Louis Lagrange. *Essai d'une nouvelle methode pour de'terminer les maxima, et les minima des formules integrales indefinies*. 1761.

[45] Claire Lauvernet, Frédéric Baret, Laurent Hascoët, Samuel Buis, and François-Xavier Le Dimet. Multitemporal-patch ensemble inversion of coupled surface–atmosphere radiative transfer models for land surface characterization. *Remote Sensing of Environment*, 112 (3):851–861, 2008.

[46] AS Lawless. A note on the analysis error associated with 3d-fgat. *Quarterly Journal of the Royal Meteorological Society*, 136(649):1094–1098, 2010.

[47] F-X Le Dimet, I Souopgui, O Titaud, and V Shutyaev. Toward the assimilation of images. *Nonlinear Processes in Geophysics Discussions*, 1:1381–1430, 2014.

[48] François-Xavier Le Dimet and Olivier Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, 38(2):97–110, 1986.

[49] François-Xavier Le Dimet, HE Ngodock, B Luong, and J Verron. Sensitivity analysis in variational data assimilation. *Journal-Meteorological Society of Japan Series 2*, 75:135–145, 1997.

[50] Francois-Xavier Le Dimet, IM Navon, and Dacian N Daescu. Second-Order Information in Data Assimilation. *Monthly Weather Review*, 130(3):629–648, 2002.

[51] Zhijin Li, I Michael Navon, and Yanqiu Zhu. Performance of 4D-Var with different strategies for the use of adjoint physics with the FSU global spectral model. *Monthly Weather Review*, 128(3):668–688, 2000.

[52] Jacques Louis Lions. *Optimal control of systems governed by partial differential equations*, volume 170. Springer Verlag, 1971.

[53] Dong C Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.

[54] Andrew C Lorenc. Analysis methods for numerical weather prediction. *Royal Meteorological Society, Quarterly Journal*, 112:1177–1194, 1986.

[55] Andrew C Lorenc. The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *QJR Meteorol. Soc*, 129:3183–3203, 2003.

[56] DG Lugenberger. *Linear and nonlinear programming*. Addison-Wesley, New York, 1984.

[57] Cristina Lupu, Carla Cardinali, and Tony McNally. Evaluation of observation impact and observation error covariance retuning. Presented at The World Weather Open Science Conference, Montreal Canada, 2014.

[58] GI Marchuk. Mathematical issues of industrial effluent optimization. *Journal of the Meteorological Society of Japan*, 60(1):481–485, 1982.

[59] Dennis McLaughlin, Alan O'Neill, John Derber, and Masafumi Kamachi. Opportunities for enhanced collaboration within the data assimilation community. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3683–3694, 2005.

[60] I. M. Navon and R. De Villiers. GUSTAF: A Quasi-Newton Nonlinear ADI FORTRAN IV Program for Solving the Shallow-Water Equations with Augmented Lagrangians. *Computers and Geosciences*, 12(2):151–173, 1986.

[61] I Michael Navon, Xiaolei Zou, J Derber, and J Sela. Variational data assimilation with an adiabatic version of the nmc spectral model. *Monthly Weather Review*, 120(7):1433–1446, 1992.

[62] IM Navon. Implementation of a posteriori methods for enforcing conservation of potential enstrophy and mass in discretized shallow-water equations models. *Monthly Weather Review*, 109(5):946–958, 1981.

[63] IM Navon and R De Villiers. Combined penalty multiplier optimization methods to enforce integral invariants conservation. *Monthly Weather Review*, 111(6):1228–1243, 1983.

[64] IM Navon and David M Legler. Conjugate-gradient methods for large-scale minimization in meteorology. *Monthly Weather Review*, 115(8):1479–1502, 1987.

[65] Seon K Park and D Županski. Four-dimensional variational data assimilation for mesoscale and storm-scale applications. *Meteorology and Atmospheric Physics*, 82(1-4):173–208, 2003.

[66] David F Parrish and John C Derber. The National Meteorological Center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120(8):1747–1763, 1992.

[67] LS Pontryagin, VG Boltyanskii, RV Gamkrelidze, and E Mishchenko. The mathematical theory of optimal processes (international series of monographs in pure and applied mathematics). *Interscience, New York*, 1962.

[68] Michael James David Powell. Nonlinear optimization 1981. In *NATO Conference Series. Series II: Systems Science, Proceedings of the NATO Advanced Research Institute, held at Cambridge (UK), July, 1981, London: Academic Press, 1982, edited by Powell, MJD*, volume 1, 1982.

[69] Florence Rabier. Overview of global data assimilation developments in numerical weather-prediction centres. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3215–3233, 2005.

[70] Florence Rabier, H Järvinen, E Klinker, J-F Mahfouf, and A Simmons. The ecmwf operational implementation of four-dimensional variational assimilation. i: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, 126(564): 1143–1170, 2000.

[71] W. Ritz. Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. *J. Reine Phys. Math.*, 135:1–61, 1908.

[72] Y. Saad. Sparsekit: a basic tool kit for sparse matrix computations. Technical Report, Computer Science Department, University of Minnesota, 1994.

[73] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2nd edition, 2003.

[74] Yoshikazu Sasaki. An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, 36(3):77–88, 1958.

[75] Yoshikazu Sasaki. Proposed inclusion of time variation terms, observational and theoretical, in numerical variational objective analysis. *J. Meteor. Soc. Japan*, 47(2):ll5–124, 1969.

[76] Yoshikazu Sasaki. Numerical variational analysis formulated under the constraints as determined by longwave equations and a low-pass filter. *Mon. Wea. Rev*, 98(12):884–898, 1970.

[77] Yoshikazu Sasaki. Some basic formalisms in numerical variational analysis. *Monthly Weather Review*, 98(12):875–883, 1970.

[78] Yoshikazu Sasaki. Variational design of finite-difference schemes for initial value problems with an integral invariant. *Journal of Computational Physics*, 21(3):270–278, 1976.

[79] Yoshikazu Sasaki, P Gu, and L Yan. A fundamental study of the numerical prediction based on the variational principle. *J. Meteor. Soc. Japan*, 33(6):262–275, 1955.

[80] D.F. Shanno and K.H. Phua. Remark on algorithm 500 - a variable method subroutine for unconstrained nonlinear minimization. *ACM Transactions on Mathematical Software*, 6:618–622, 1980.

[81] V Shutyaev, I Gejadze, GJM Copeland, and F-X Le Dimet. Optimal solution error covariance in highly nonlinear problems of variational data assimilation. *Nonlinear Processes in Geophysics*, 19(2):177–184, 2012.

[82] AJ Simmons and A Hollingsworth. Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 128(580):647–677, 2002.

[83] Răzvan Ştefănescu, Adrian Sandu, and Ionel Michael Navon. POD/DEIM reduced-order strategies for efficient four dimensional variational data assimilation. *Journal of Computational Physics*, 295:569–595, 2015.

[84] Jesse Jerald Stephens. *A variational approach to numerical weather analysis and prediction*. PhD thesis, Texas A& M University, College-Station, TX, 77863, 1966.

[85] JJ Stephens. Variational resolution of wind components. *Mon. Weather Rev*, 96:229–231, 1968.

[86] Yannick Trémolet. Incremental 4d-var convergence study. *Tellus A*, 59(5):706–718, 2007.

[87] Yannick Trémolet. Computation of observation sensitivity and observation impact in incremental variational data assimilation. *Tellus A*, 60(5):964–978, 2008.

[88] Tadashi Tsuyuki. Variational data assimilation in the Tropics using precipitation data. Part III: Assimilation of SSM/I precipitation rates. *Monthly Weather Review*, 125(7):1447–1464, 1997.

[89] F Veersé and J-N Thepaut. Multiple-truncation incremental approach for four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 124 (550):1889–1908, 1998.

[90] BE Vieux, F LeDimet, and D Armand. Inverse problem formulation for spatially distributed river basin model calibration using the adjoint method. *EGS, Annales Geophysicae, Part II, Hydrology, Oceans and Atmosphere, Supplement II to*, 16:C501, 1998.

[91] Grace Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 24(5): 383–393, 1975.

[92] Grace Wahba. Some new techniques for variational objective analysis on the sphere using splines. hough-functions, and sample spectral data. In *Proc. 7th Conf. on Probability and Statistics in the Atmos. Sci. Monterey Cal*, pages 213–216, 1981.

[93] Grace Wahba. Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing*, 2(1):5–16, 1981.

[94] Grace Wahba and James Wendelberger. Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly Weather Review*, 108(8):1122–1143, 1980.

[95] WM Washington and RJ Duquet. An objective analysis of stratospheric data by Sasaki's method. *Dept. of Meteorology, Penn. State University, University Park, PA*, 16802:23, 1963.

[96] Jeffrey S Whitaker and Thomas M Hamill. Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7):1913–1924, 2002.

[97] Lin Wu, François-Xavier Le Dimet, Philippe De Reffye, Bao-Gang Hu, Paul-Henry Cournède, and Meng-Zhen Kang. An optimal control methodology for plant growth—Case study of a water supply problem of sunflower. *Mathematics and Computers in Simulation*, 82 (5):909–923, 2012.

[98] Dušanka Županski and Fedor Mesinger. Four-dimensional variational assimilation of precipitation data. *Monthly Weather Review*, 123(4):1112–1127, 1995.

[99] Dusanka Zupanski, Milija Zupanski, Eric Rogers, David F Parrish, and Geoffrey J DiMego. Fine-resolution 4DVAR data assimilation for the Great Plains tornado outbreak of 3 May 1999. *Weather and Forecasting*, 17(3):506–525, 2002.