Data Assimilation for Geophysical Fluids

Jacques Blum

Laboratoire Jean-Alexandre Dieudonné, Université de Nice Sophia-Antipolis, Parc Valrose, 06108 Nice cedex 2, France

François-Xavier Le Dimet

INRIA and Laboratoire Jean-Kuntzmann, Université de Grenoble, 38041, Grenoble Cedex 9, France

I. Michael Navon

Department of Scientific Computing, The Florida State University, Tallahassee, Florida 32306-4120, USA

Abstract

The ultimate purpose of environmental studies is the forecast of its natural evolution. A prerequisite before a prediction is to retrieve at best the state of the environment. Data assimilation is the ensemble of techniques which, starting from heterogeneous information, permit to retrieve the initial state of a flow. In the first part, the mathematical models governing geophysical flows are presented together with the networks of observations of the atmosphere and of the ocean. In variational methods, we seek for the minimum of a functional estimating the discrepancy between the solution of the model and the observation. The derivation of the optimality system, using the adjoint state, permits to compute a gradient which is used in the optimization. The definition of the cost function permits to take into account the available statistical information through the choice of metrics in the space of observation and in the space of the initial condition. Some examples are presented on simplified models, especially an application in oceanography. Among the tools of optimal control, the adjoint model permits to carry out sensitivity studies, but if we look for the sensitivity of the prediction with respect to the observations, then a second-order analysis should be considered. One of the first methods used for assimilating data in oceanography is the nudging method, adding a forcing term in the equations. A variational variant of nudging method is described and also a so-called

Computational Methods for the Atmosphere and the Oceans
Special Volume (Roger M. Temam and Joseph J. Tribbia, Guest Editors) of
HANDBOOK OF NUMERICAL ANALYSIS, VOL. XIV
P.G. Ciarlet (Editor)

ISSN 1570-8659
DOI 10.1016/S1570-8659(08)00209-3

"back and forth" nudging method. The proper orthogonal decomposition method is introduced in order to reduce the cost of the variational method. For assimilating data, stochastic methods can be considered, being based on the Kalman filter extended to nonlinear problems, but the inconvenience of this method consists in the difficulty of handling huge covariance matrices. The dimension of the systems used for operational purposes (several hundred of millions of variables) requires to work with reduced variable techniques. The ensemble Kalman filter method, which is a Monte-Carlo implementation of the Bayesian update problem, is described. A considerable amount of information on geophysical flows is provided by satellites displaying images of their evolution, the assimilation of images into numerical models is a challenge for the future: variational methods are successfully considered in this perspective.

1. Introduction: specificity of geophysical flows

The mathematical modeling of geophysical flows has experienced a tremendous development during the last decades, mainly due to the growth of the available computing resources and to the development of networks of remote or in situ observations. The domain of modeling has been extended to complex flows such as the atmosphere with some chemical species or coupled media such as the atmosphere and the ocean. A tentative extension of the domain of prediction is also under way with seasonal prediction on one hand and short term and very accurate prediction on the other hand: nowcasting mainly devoted to extreme events. Geophysical fluids such as air, atmosphere, ocean, surface, or underground water are governed by the general equations of fluid dynamics: mass and energy conservation, behavior laws. Nevertheless some specific factors must be taken into account such as follows:

- Uniqueness of a situation. Each geophysical episode is unique. A given situation has never existed before and will not exist in the future. A field experiment cannot be duplicated. It means that environmental sciences are not strictu sensu experimental sciences: an hypothesis cannot be validated by repetitions of an experiment. Geophysical models should be tested and validated with data associated to distinct episodes.
- Nonlinearity. Geophysical processes are nonlinear due to their fluid component, and furthermore, they include some other nonlinear processes such as radiative transfer. Nonlinearity implies interactions and energy cascades between spatial and temporal scales. Seeking a numerical solution to the equations requires discretizing these equations and therefore cutting off some of the scales. A major problem comes from the fact that subscale processes could be associated with large fluxes of energy. For example, a cumulonimbus cloud has a characteristic size of 10 km in the horizontal and vertical directions. The typical grid size of a general circulation model (GCM) is of the order of 40 km, therefore, larger than the characteristic dimension of a cumulonimbus, a thunderstorm cloud. The total energy (thermal and mechanical) of such a cloud is considerable. By the same token typical vertical velocities of a GCM are of the order of some centimeters or decimeters per second,

in a cumulonimbus cloud there exist observations of vertical velocities of the order of 100 meters per second. Therefore, it will be crucial to represent the fluxes of energy associated to subgrid processes by some additional terms in the equations. Parametrization of subgrid effects will include some empirical coefficients that should be tuned in such a way that the model produces a good forecast.

- Initial and boundary conditions. The general equations are not sufficient to carry out a prediction. Some additional information, such as initial and boundary conditions, should be provided. Most of the geophysical fluids do not have any natural boundaries. In the same way, there are no natural initial conditions. Nevertheless, these conditions are essential for carrying out a prediction, more especially as the system is turbulent and hence very much dependent on the initial condition.

Therefore, it is crystal clear that modeling will have to take into account observations. If, for instance, we consider a measurement of the wind at a given site, the same data can be used either in a GCM or in a very local model. According to the context, i.e., the model, a different confidence will be attributed to the observation. It does not make to have a model without data or data without model (otherwise known as "Lions' theorem").

Data assimilation is the ensemble of techniques combining in an optimal way (in a sense to be defined) the mathematical information provided by the equations and the physical information given by the observation in order to retrieve the state of a flow. The concept of data assimilation can be extended to other sources of information, e.g., statistics of error on the observations and/or on the error of prediction. Another source of information is provided by images originating from space observations, which as of the present time are not optimally used. The goal of data assimilation is to link together these heterogeneous (in nature, quality, and density) sources of information in order to retrieve a coherent state of the environment at a given date.

The equations of the model (shallow water, quasi-geostrophic (QG), or general primitive equations) are of the first order with respect to time. In a GCM, there is no lateral boundary condition. Assuming that all the regularity conditions are fulfil (if they were known), an initial condition will be sufficient to integrate the model and to get the forecast.

Originally, the problem of data assimilation was to determine the initial condition from observations. Since the same mathematical tools are used, data assimilation also includes the estimation of some model parameters or of some boundary conditions.

As a first approximation, three types of methods are considered:

- Interpolation methods. These methods interpolate the measurement from the points of observation toward the grid points. The interpolation can be weighted by the statistics of the observations (covariance matrices). The method is simple to implement, but it is not justified by physical arguments: the retrieved fields will not be necessarily coherent from the physical viewpoint, e.g., the initial condition may be located outside of the attractor, therefore the associated solution will contain gravity waves until it reaches the attractor. Until a recent date, these methods were the most commonly used in operational meteorology. For a review on optimal interpolation, one can refer to KALNAY [2003].

- Variational methods. Data assimilation is set as being a problem of constrained optimization, then the tools of optimal control are carried out to solve it. At present, these methods are operationally used at the European Center for Medium Range Weather Forecasting (ECMWF) (Reading, UK), Météo-France, the National Center for Environmental Prediction (USA), Japan and Canada (see RABIER [2005], KALNAY [2003]).
- Stochastic methods. The basic idea is to consider the fields as the realization of a stochastic process and carry out Kalman filtering (KF) methods. The main difficulty stems from the fact that the covariance matrices of the state variables have huge dimensions for operational models. The ensemble Kalman filtering (EnKF) methods were devised to address this issue and are presently seeing a major development at different research centers.

Section 2 will be devoted to the presentation of a certain number of simplified models for the geophysical flows and to a description, with appropriate figures, of data available for the atmosphere or the ocean. In Section 3, the variational method, mentioned above and often called Four-Dimensional Variational Data Assimilation (4D-VAR) is explained in detail and two examples of solutions for shallow water or QG models will be given. Section 4 is devoted to a second-order adjoint analysis, which enables in particular to perform a sensitivity analysis on the results of the variational method. In Section 5, the so-called nudging method is explained, with a special emphasis on the optimal nudging method, which uses the variational technique. In order to reduce the cost of these 4D-VAR methods, the Proper Orthogonal Decomposition (POD) is introduced in Section 6, and the application of this reduced-space basis to variational methods is presented. In Section 7, KF is introduced and a special emphasis is given to the EnKF, which is widely used in operational data assimilation. Finally, the recent problem of assimilation of images in meteorology or oceanography is tackled in Section 8 by a variational technique.

2. Models and data for geophysical flows

2.1. Models

The equations governing the geophysical flows are derived from the general equations of fluid dynamics. The main variables used to describe the fluids are as follows:

- The components of the velocity
- Pressure
- Temperature
- Humidity in the atmosphere, salinity in the ocean
- Concentrations for chemical species

The constraints applied to these variables are as follows:

- Equations of mass conservation
- Momentum equation containing the Coriolis acceleration term, which is essential in the dynamic of flows at extra tropical latitudes

- Equation of energy conservation including law of thermodynamics
- Law of behavior (e.g., Mariotte's Law)
- Equations of chemical kinetics if a pollution type problem is considered

These equations are complex; therefore, we cannot expect to obtain an analytical solution. Before performing a numerical analysis of the system, it will be necessary to

- Simplify the equations. This task will be carried out on physical basis. For example, three-dimensional (3D) fields could be vertically integrated using hydrostatic assumptions in order to obtain a two-dimensional (2D) horizontal field which is more tractable for numerical purposes.
- Discretize the equations. The usual discretization methods are considered: finite differences, finite elements, or spectral methods.

Several of these techniques may be simultaneously used. For instance, in the ARPEGE model designed by Meteo-France, the horizontal discretization is spectral in longitude, a truncated development in Legendre's polynomial along latitude, while the vertical one is based on a finite difference scheme. The horizontal nonlinear terms (advection) are computed using finite differences and then transformed onto a spectral base. As mentioned above, it will be necessary to estimate the subgrid fluxes of energy and matter. A parametrization of these phenomena will be included in the model, which will contain some empirical parameters, difficult to estimate and to adjust because they are associated to some complex physical processes which cannot be directly measured.

In mathematics, it is usual to study the convergence of discrete models toward a continuous one when the discretization parameter goes to zero. Does this approach make sense for these problems? According to the value of this parameter, the physics of the problem will change and another parametrization will be necessary. With a grid size of 100 km, the cumulonimbus clouds will not be explicit in the model. With a grid size of 100 m, the physics of convective clouds, including the water cycle under gaseous, liquid, and solid phases, should be explicitly taken into account. According to the value of the discretization parameter, the domain of validity of the approximation will change. In this framework, the convergence of discretization schemes is beyond the scope of actual problems.

In meteorology as well as in oceanography or hydrology, it is usual to use a basic "toy" model for numerical experiments: the Saint-Venant's equations (or shallow water equations). These 2D horizontal equations are obtained after a vertical integration of the 3D fields, assuming the hydrostatic approximation, this is equivalent to neglecting the vertical acceleration. The density is supposed to be constant; therefore, there is no thermodynamic effect. They assume the form:

$$\begin{aligned}
&\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + \frac{\partial \phi}{\partial x} = 0,\\
&\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + \frac{\partial \phi}{\partial y} = 0,\\
&\frac{\partial \phi}{\partial t} + \frac{\partial u\phi}{\partial x} + \frac{\partial v\phi}{\partial y} = 0,
\end{aligned} \tag{2.1}$$

where u and v are the horizontal components of the velocity, ϕ is the geopotential (proportional to the pressure), and f is the Coriolis parameter for earth's rotation.

An important property of these equations is the existence of an attractor, on which the solution will orbit. For very simplified models such as Lorenz' equations of 3Ds in the space phase, the structure of this attractor is already topologically complex. For more realistic models, it is known that the attractor is, at midlatitudes, in the vicinity of the geostrophic equilibrium, which is defined by the equality of the gradient of pressure and the Coriolis force. From a practical point of view, the attractor is characterized by a weak development of gravity waves. Therefore, if the initial condition does not belong to a close neighborhood of the attractor, then the integration of the model will give rise to undue gravity waves until the solution reaches the attractor.

Another "toy" model, which filters the gravity waves, is the QG model where the dominant terms are the pressure gradient and the Coriolis force, which cancel each other in the geostrophic balance. It consists in a first-order expansion of the Navier-Stokes equation with respect to the Rossby number. It is an approximate model with respect to the full primitive equation model, in particular because thermodynamics are discarded. However, it has been shown to be able to realistically reproduce the statistical properties of midlatitude ocean circulation including the very energetic jet and mesoscale features, typical of regions like the Gulf Stream. The equations of this model will be given in the next section.

Operational models in meteorology and oceanography are of very large dimension with 10^7 to 10^9 variables, hence the implementation of efficient numerical methods is a challenge for high performance computing. Future developments of coupled models ocean-atmosphere will dramatically increase the need for efficient numerical methods for coupled models.

2.2. *Data*

At the present time, many sources of data are used. Around 300 millions of data are screened every day by the ECMWF located in Reading (UK). An exhaustive information can be found on the ECMWF site http://www.ecmwf.int/products/forecasts/d/charts/monitoring/coverage/.

In meteorology, the data collected for operational use are as follows:

- Ground observations: wind, temperature, pressure, humidity. These data are collected on a dedicated network, on ships, and also in airports (see Fig. 2.1). The number of observations varies from day to day. To give an order of magnitude, about 200,000 data are measured daily, but after a quality control process, only 30,000 are used in the assimilation.
- Pilot balloons provide information on the wind.
- Radiosondes provide data on the vertical structure of the atmosphere: wind, temperature, pressure, and humidity. This network is displayed in Fig. 2.2. The figures of synoptic measurements and radiosonde clearly show the heterogeneity of the data density: North America and Europe have a good coverage, while information is very sparse on the oceans. This lack of information is compensated by drifting buoys.

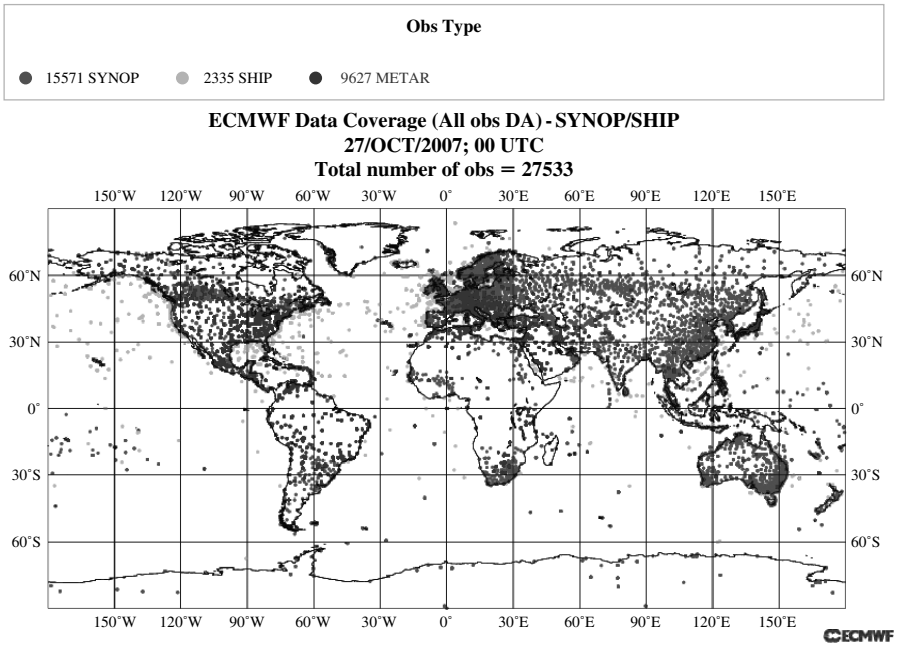


FIG. 2.1 SYNOP/SHIP data: synoptic networks in red, airport data in blue, and ship data in green.

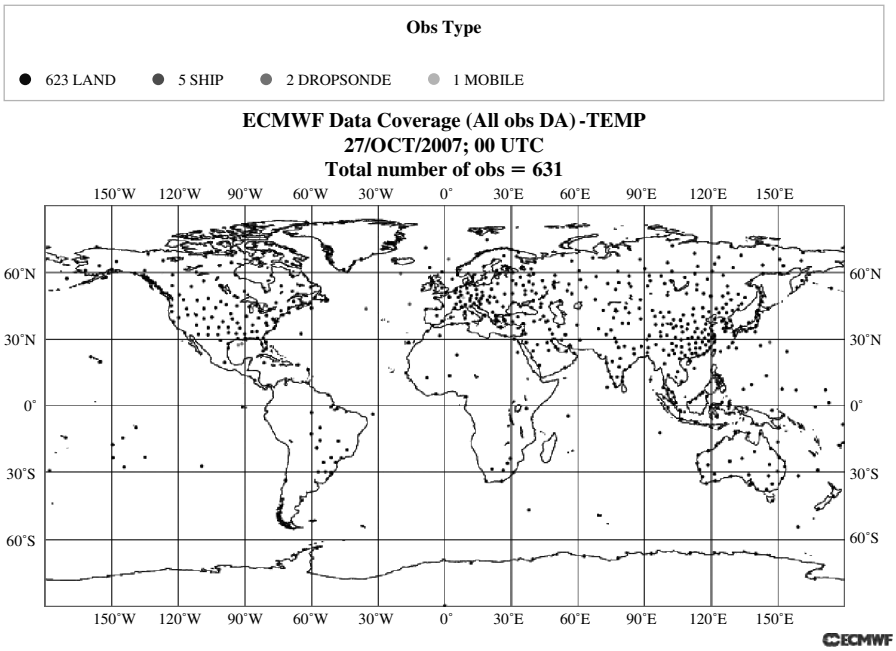


FIG. 2.2 Radiosonde measurements.

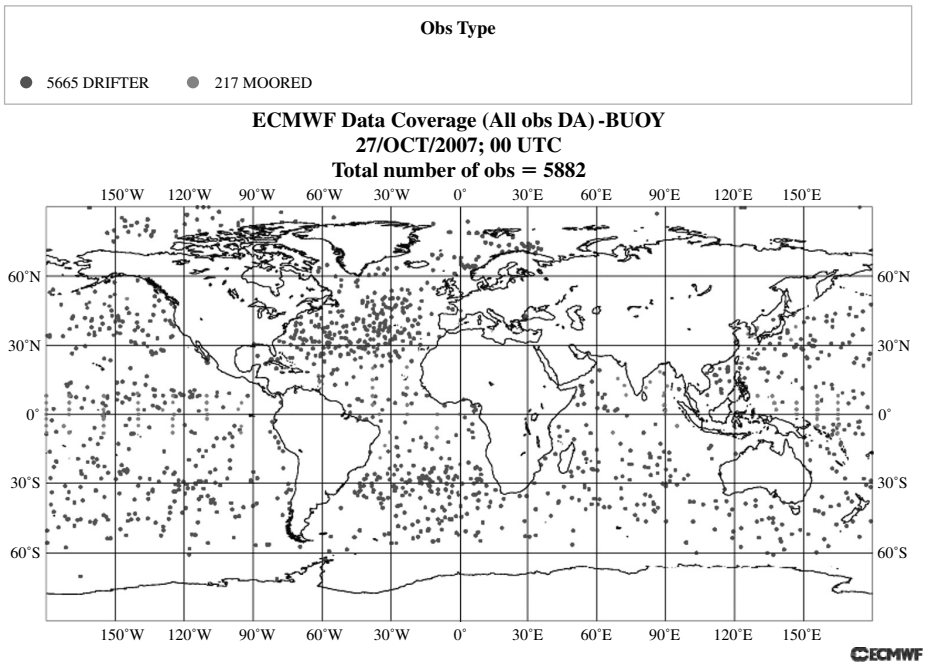


FIG. 2.3 Drifting and moored buoys.

- Drifting buoys measure the temperature of the air and of the ocean, salinity, and wind. The network is completed by moored buoys, mainly located in the most energetic part of the ocean (see Fig. 2.3). These data are also used for oceanic models. Around 6000 data provided by buoys are daily assimilated at ECMWF.

At the present time (2007), around 99 percent of screened data originates from 45 satellites, but only 94 percent of assimilated data comes from satellites. Two main categories of satellites are used:

- Geostationary satellites provide information on the wind by estimating the shifting of clouds considered as Lagrangian tracers. To make this measurement useful, the altitude of the clouds must be known (it is derived, by solving an inverse problem, from the vertical temperature profiles). Figure 2.4 displays the areas of observation covered by ten geostationary satellites. Around 300,000 observations are assimilated.
- Polar-orbiting satellites (NOAA, EUMETSAT) are used for the estimation of the vertical temperature profiles, basically radiances are measured, then temperatures are estimated as the solution of an inverse problem. Figure 2.5 displays the trajectories of six satellites on October 27th 2007. Around 400,000 observations are assimilated.

In oceanography, data are much scarcer than in meteorology. Electromagnetic waves do not penetrate deeply in the ocean; therefore, remote sensing is much more difficult. The

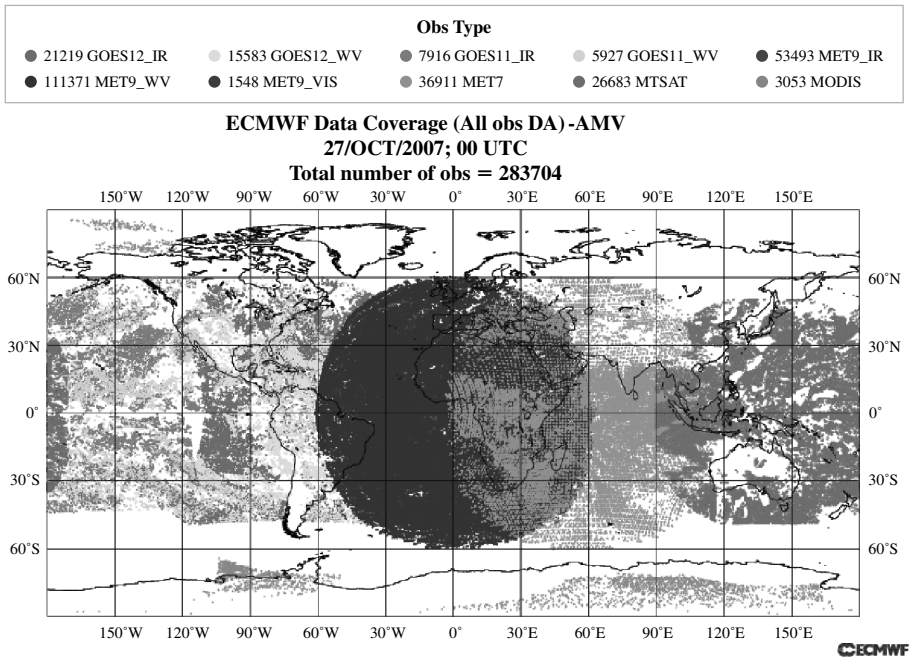


FIG. 2.4 Observations from ten geostationary satellites.

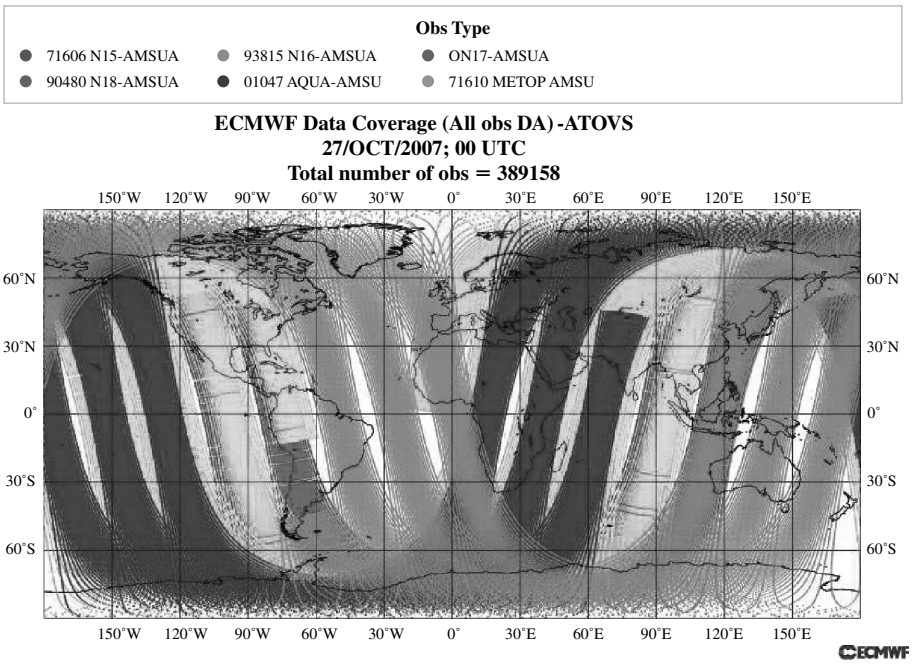


FIG. 2.5 Trajectories of six polar-orbiting satellites.

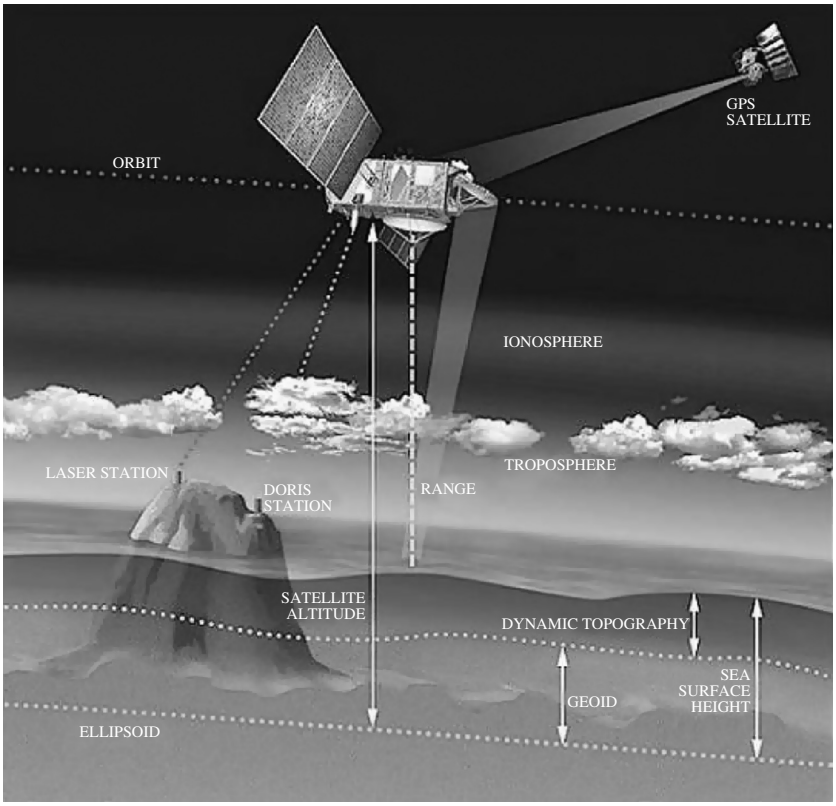


FIG. 2.6 Satellite altimetry (from Aviso web site).

development of operational oceanography is based on in situ measurements (temperature, salinity) and on altimetric satellites (ERS, Topex-Poseidon, Jason) measuring the surface elevation of the ocean with precision on the order of some centimeter (see Fig. 2.6). Lagrangian floats measure also the position of the drifting buoys with a given time periodicity (NODET [2006]).

Both in meteorology and oceanography, observations are heterogeneous in nature, density, and quality. A rough estimation of the number of screened data is 300 millions and of assimilated data is around 18 millions. This number has to be compared with the 800 millions of variables of the ECMWF operational model in 2007. Therefore, retrieving the state of the atmosphere from observations is clearly an ill-posed problem. Some a priori information has to be provided to estimate an initial state.

3. Variational methods

Variational methods were introduced by SASAKI [1958]. These methods consider the equations governing the flow as constraints, and the problem is closed by using a

variational principle, e.g., the minimization of the discrepancy between the model and the observations. In the following, we will consider that the model is discrete with respect to space variables. Optimal Control Techniques (LIONS [1971]) were proposed by LE DIMET [1980], LE DIMET and TALAGRAND [1986], and COURTIER and TALAGRAND [1987].

3.1. Ingredients

The various ingredients of a variational method are as follows:

- A state variable $X \in \mathcal{X}$ which describes the evolution of the medium at the grid points. X depends on time and is for operational models of large dimension (3.10^7 for the ECMWF model).
- A model describing the evolution of the fluid. Basically, it is a system of nonlinear differential equations which is written as

$$\begin{cases} \dfrac{\mathrm{d}X}{\mathrm{d}t} = F(X, U) \\ X(0) = V \end{cases}. \tag{3.1}$$

- A control variable $(U, V) \in \mathcal{P}$ space of control. Most of the time the control is the initial condition or/and some internal variables of the model: parameters or boundary conditions. We will assume that when a value has been fixed for the parameter, then the model has a unique solution. For sake of simplicity, we will not consider constraints on the state variable. Nevertheless, humidity and salinity cannot be negative; therefore, the set of controls does not necessarily have the structure of a vector space.
- Observations $X_{obs} \in \mathcal{O}_{obs}$. They are discrete and depend on time and space and are not, from either geographical or physical point of view, in the same space as the state variable. Therefore, we will introduce some operator C mapping the space of state into the state of observations. In practice, this operator can be complex.
- A cost function J measuring the discrepancy of the solution of the model associated to (U, V) and the observations.

$$J(U, V) = \frac{1}{2} \int_0^T \|C.X(U, V) - X_{obs}\|^2 \mathrm{d}t \tag{3.2}$$

The choice of the norm is important because it allows introduction of some a priori information like the statistics of the fields through the covariance matrix which is positive definite. In practice, some additional term is added to the cost function, e.g., the so-called background term which is the quadratic difference between the initial optimal variable and the last prediction. This term acts like a regularization term in the sense of Tikhonov (TIKHONOV and ARSENIN [1977]).

Then the problem of variational data assimilation (VDA) can be set as

$$\begin{cases} \text{Find} \quad U^*, V^* \in \mathcal{P} \quad \text{such that} \\ J(U^*, V^*) = \underset{(U,V)\in\mathcal{P}}{\text{Inf}} J(U, V). \end{cases} \tag{3.3}$$

3.2. *Optimality system*

With respect to (U, V), we have a problem of unconstrained optimization. Problem in Eq. (3.3) will have a unique solution if J is strictly convex, lower semicontinuous and if

$$\lim_{||(U,V)||\to+\infty} J(U, V) \to +\infty.$$

When J is differentiable, a necessary condition for (U^*, V^*) to be a solution is given by the Euler-Lagrange equation:

$$\nabla J(U^*, V^*) = 0,$$

where ∇J is the gradient of J with respect to (U, V).

Furthermore, the determination of ∇J permits one to implement optimization methods of gradient type.

Let $(u, v) \in \mathcal{P}$, $\widehat{X}$ be the Gâteaux-derivative (directional derivative) of X in the direction (u, v) that is the solution of

$$\begin{cases} \dfrac{\mathrm{d}\widehat{X}}{\mathrm{d}t} = \left[\dfrac{\partial F}{\partial X}\right].\widehat{X} + \left[\dfrac{\partial F}{\partial U}\right].u, \\ \widehat{X}(0) = v \end{cases} \tag{3.4}$$

where $\left[\dfrac{\partial F}{\partial X}\right]$ is the Jacobian of the model with respect to the state variable. This Eq. (3.4) is known as the *linear tangent model*.

By the same token, we get the directional derivative of J:

$$\widehat{J}(U, V, u, v) = \int_0^T \left(C.X - X_{obs}, C\widehat{X}\right)\mathrm{d}t.$$

We will get the gradient by exhibiting the linear dependence of $\widehat{J}$ with respect to (u, v). For this purpose, we introduce $P \in \mathcal{X}$ the so-called *adjoint variable*, to be defined later. Let us take the inner product of Eq. (3.4) with P, then integrate between 0 and T. An integration by part shows that if P is defined as the solution of

$$\begin{cases} \dfrac{\mathrm{d}P}{\mathrm{d}t} + \left[\dfrac{\partial F}{\partial X}\right]^T.P = C^t\left(C.X - X_{obs}\right) \\ P(T) = 0, \end{cases} \tag{3.5}$$

then the gradient is given by

$$\nabla J = \begin{pmatrix} \nabla_U J \\ \nabla_V J \end{pmatrix} = \begin{pmatrix} -\left[\dfrac{\partial F}{\partial U}\right]^t . P \\ -P\,(0) \end{pmatrix}.$$

Therefore, the gradient is obtained by a backward in time integration of the adjoint model.

3.3. Optimization

The determination of $\left(U^*, V^*\right)$ is carried out by performing a descent-type unconstrained optimization method. Given a first guess (U_0, V_0), we define a sequence by

$$\begin{pmatrix} U_n \\ V_n \end{pmatrix} = \begin{pmatrix} U_{n-1} \\ V_{n-1} \end{pmatrix} + \rho_n D_n.$$

D_n is the direction of descent. Usually conjugate gradient or Newton type methods are used. ρ_n is the step size defined by

$$\rho_n = ArgMin J \left(\begin{pmatrix} U_{n-1} \\ V_{n-1} \end{pmatrix} + \rho D_n \right),$$

This problem looks simple: it is the minimization of a function of one variable. For a nonlinear problem, it entails a high computational cost since several integrations of the model are required for the evaluation of J. Optimization libraries, e.g., MODULOPT (Gilbert and Lemarechal [1989]), are widely used and are efficient.

For a comprehensive test of powerful large-scale unconstrained minimization methods applied to VDA, see Zou, Navon, Berger, Phua, Schlick and Le Dimet [1993].

3.4. Implementation

A major difficulty encountered in the implementation of this method is the derivation of the adjoint model. A bad solution would be to derive the adjoint model from the continuous direct model, then to discretize it. The convergence of the optimization algorithm requires having the gradient of the cost function with a precision of the order of the computer's roundoff error.
Two steps are carried out for the derivation of the adjoint:

- Differentiation of the direct model. This step serves to determine the linear tangent model. This task is easily carried out by differentiating the direct code line by line.
- Transposition of the linear tangent model. Transposition with respect to time is simply the backward integration. To carry out the transposition, one should start from the last statement of the linear tangent code and transpose each statement. The difficulty stems from the hidden dependencies.

If some rules in the direct code are adhered to, then the derivation of the adjoint model can be made simpler, otherwise it is a long and painful task. Nevertheless, we

can use some automatic differentiation code such as Odyssee (ROSTAING-SCHMIDT and HASSOLD [1994]) (see also TAPENADE, TAMC, FASTOPT). Recent developments on these techniques can be found in MOHAMMADI and PIRONNEAU [2001].

3.5. Remarks

- If the model is nonlinear, then the cost function is not necessarily convex, and the optimization algorithm may converge toward a local minimum. In this case, one can expect convergence toward a global optimum only if the first guess is in the vicinity of the solution. This may occur in meteorology where the former forecast is supposed to be close to the actual state of the atmosphere. In practice, a so-called background term is added to the cost function, measuring the quadratic discrepancy with the prediction. In terms of control, this term could be considered as a regularization in the sense of Tikhonov.
- The optimization algorithm could converge to a correct mathematical solution but would be physically incorrect (e.g., negative humidity).The solution may be far away from the attractor, the regularization term will force the model to verify some additional constraints, e.g., for the solution to remain in the vicinity of the geostrophic equilibrium.
- Regularization terms permit to take into account the statistical information on the error by an adequate choice of the quadratic norm including an error covariance matrix.
- If the control variable U is time dependent, which is the case if boundary conditions are controlled, then we may get problems with a huge dimension. In this case, it will be important to choose an appropriate discretization of the control variable in order to reduce its dimension.
- PUEL [2002] has proposed a new approach to determine the final state, instead of the initial condition, that makes the inverse problem well-posed, thanks to Carleman inequalities.

3.6. Example 1: 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

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + \frac{\partial \phi}{\partial x} = 0, \tag{3.6}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + \frac{\partial \phi}{\partial y} = 0, \tag{3.7}$$

$$\frac{\partial \phi}{\partial t} + \frac{\partial u\phi}{\partial x} + \frac{\partial v\phi}{\partial y} = 0. \tag{3.8}$$

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 hypotheses 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. $\mathbf{C} \equiv \mathbf{I}$, where $\mathbf{I}$ is the identity operator. If $U_0 = (u_0, v_0, \phi_0)^T$ is the initial condition and if the cost function is given by

$$J(U_0) = \frac{1}{2}\int_0^T [\|u - u_{obs}\|^2 + \|v - v_{obs}\|^2 + \gamma\|\phi - \phi_{obs}\|^2]\mathrm{d}t, \tag{3.9}$$

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

$$\frac{\partial \bar{u}}{\partial t} + u\frac{\partial \bar{u}}{\partial x} + \bar{u}\frac{\partial u}{\partial x} + v\frac{\partial \bar{u}}{\partial y} + \bar{v}\frac{\partial u}{\partial y} - f\bar{v} + \frac{\partial \bar{\phi}}{\partial x} = 0, \tag{3.10}$$

$$\frac{\partial \bar{v}}{\partial t} + u\frac{\partial \bar{v}}{\partial x} + \bar{u}\frac{\partial v}{\partial x} + v\frac{\partial \bar{v}}{\partial y} + \bar{v}\frac{\partial v}{\partial y} + f\bar{u} + \frac{\partial \bar{\phi}}{\partial y} = 0, \tag{3.11}$$

$$\frac{\partial \bar{\phi}}{\partial t} + \frac{\partial \bar{u}\phi}{\partial x} + \frac{\partial u\bar{\phi}}{\partial x} + \frac{\partial \bar{v}\phi}{\partial y} + \frac{\partial v\bar{\phi}}{\partial y} = 0. \tag{3.12}$$

The adjoint model is obtained by transposition of the linear tangent model. Let $P = (\tilde{u}, \tilde{v}, \tilde{\phi})^T$ be the adjoint variable and after some integrations by parts both in time and space, we see that the adjoint model is defined as being the solution of

$$\frac{\partial \tilde{u}}{\partial t} + u\frac{\partial \tilde{u}}{\partial x} + v\frac{\partial \tilde{u}}{\partial y} + \tilde{u}\frac{\partial v}{\partial y} - \tilde{v}\frac{\partial v}{\partial x} - f\tilde{v} + \phi\frac{\partial \tilde{\phi}}{\partial x} = u - u_{obs}, \tag{3.13}$$

$$\frac{\partial \tilde{v}}{\partial t} - \tilde{u}\frac{\partial u}{\partial y} + u\frac{\partial \tilde{v}}{\partial x} + \tilde{v}\frac{\partial u}{\partial x} + v\frac{\partial \tilde{v}}{\partial y} + f\tilde{u} + \phi\frac{\partial \tilde{\phi}}{\partial y} = v - v_{obs}, \tag{3.14}$$

$$\frac{\partial \tilde{\phi}}{\partial t} + \frac{\partial \tilde{u}}{\partial x} + \frac{\partial \tilde{v}}{\partial y} + u\frac{\partial \tilde{\phi}}{\partial x} + v\frac{\partial \tilde{\phi}}{\partial y} = \gamma(\phi - \phi_{obs}), \tag{3.15}$$

with final conditions equal to 0.

Then the gradient of J is given by

$$\nabla J(U_0) = -P(0) = -\begin{pmatrix} \tilde{u}(0) \\ \tilde{v}(0) \\ \tilde{\phi}(0) \end{pmatrix}. \tag{3.16}$$

In VIDARD [2001], a square oceanic shallow water model (2000 km) is studied, it is discretized with a grid size of 25 km. The period of assimilation lasts one month, and the time-step is 90 mn. Fictitious data provided by the true solution are used after a random perturbation. The optimization code M1QN3 is issued from the MODULOPT optimization library (GILBERT and LEMARECHAL [1989]): it is a quasi Newton algorithm.

3.7. Example 2: a QG model

The oceanic model used in this study is based on the QG approximation obtained by writing the conservation of the potential vorticity (HOLLAND [1978]). The vertical structure of the ocean is divided into N layers. Each one has a constant density ρ_k with a depth H_k ($k = 1, \ldots, N$). We get a coupled system of N equations:

$$\frac{D_k(\theta_k(\Psi) + f)}{Dt} + \delta_{k,N}\, C_1 \Delta \Psi_N - C_3 \Delta^3 \Psi_k \;=\; F_k \quad \begin{array}{l} \text{dans } \Omega \times [0, T], \\ \forall k = 1, \ldots, N, \end{array} \tag{3.17}$$

where

- $\Omega \subset I\!R^2$ is the oceanic basin, and $[0, T]$ the time interval for the study;
- Ψ_k is the stream function in the layer k;
- $\theta_k(\Psi)$ is the potential vorticity in the layer k, given by

$$\begin{pmatrix} \theta_1(\Psi) \\ \vdots \\ \theta_N(\Psi) \end{pmatrix} = [\Delta - [W]] \begin{pmatrix} \Psi_1 \\ \vdots \\ \Psi_N \end{pmatrix},$$

where $[W]$ is a $N \times N$ tridiagonal matrix, whose entries depend on physical parameters:

$$W_{k,k-1} = -\frac{f_0^2}{H_k g'_{k-\frac{1}{2}}},\; W_{k,k+1} = -\frac{f_0^2}{H_k g'_{k+\frac{1}{2}}},\; W_{k,k} = \frac{f_0^2}{H_k}\Big(\frac{1}{g'_{k-\frac{1}{2}}} + \frac{1}{g'_{k+\frac{1}{2}}}\Big),$$

where f_0 is the value of the Coriolis parameter at the middle latitude of Ω, $g'_{k+\frac{1}{2}} = g\,(\rho_{k+1} - \rho_k)/\rho$ is the reduced gravity at the interface k-$k+1$ (g is the earth gravity and ρ the mean density of the fluid).

- f is the Coriolis force. According to the β-plane approximation, it linearly varies with latitude: $f(x, y) = f_0 + \beta.y$, where (x, y) are the Cartesian coordinates in Ω;
- $\dfrac{D_k.}{Dt}$ is the Lagrangian derivative in layer k, given by

$$\frac{D_k.}{Dt} = \frac{\partial.}{\partial t} - \frac{\partial \Psi_k}{\partial y}\frac{\partial.}{\partial x} + \frac{\partial \Psi_k}{\partial x}\frac{\partial.}{\partial y} = \frac{\partial.}{\partial t} + J(\Psi_k, .),$$

where $J(.,.)$ is the Jacobian operator $J(\varphi, \xi) = \dfrac{\partial \varphi}{\partial x}\dfrac{\partial \xi}{\partial y} - \dfrac{\partial \varphi}{\partial y}\dfrac{\partial \xi}{\partial x}$;

- $C_1\Delta\Psi_N$ is the dissipation on the bottom of the ocean;
- $C_3\Delta^3\Psi_k$ is the parametrization of internal and subgrid dissipation;
- F_k is a forcing term. In this model, only the tension due to the wind, denoted τ, is taken into account. Therefore, we get

$$F_1 = \mathrm{Rot}\tau \text{ and } F_k = 0, \quad \forall k \geq 2.$$

The Eq. (3.17) is written in vector form:

$$\frac{\partial}{\partial t}(\Delta - [W]) \begin{pmatrix} \Psi_1 \\ \vdots \\ \Psi_N \end{pmatrix} = \begin{pmatrix} G_1 \\ \vdots \\ G_N \end{pmatrix}, \tag{3.18}$$

with $G_k = F_k - J(\Psi_k, \theta_k(\Psi) + f) - \delta_{k,N} C_1 \Delta\Psi_N + C_3\Delta^3\Psi_k$.

We are going to consider altimetric measurement of the surface of the ocean given by satellite observations (Topex-Poseidon, Jason). The observed data is the change in the surface of the ocean. According to the QG approximation, it is proportional to the stream function in the surface layer:

$$h^{obs} = \frac{f_0}{g}\Psi_1^{obs}.$$

Therefore, we will assimilate surface data in order to retrieve the fluid circulation especially in the deep ocean layers.

The control vector is the initial state on the N layers:

$$u = \Big(\Psi_k(t=0)\Big)_{k=1,\ldots,N} \in \mathcal{U}_{\mathrm{ad}}.$$

The state vector is

$$\Big(\Psi_k(t)\Big)_{k=1,\ldots,\dot{N}}.$$

We assume that the stream function is observed at each point of the surface layer at discrete times t_j. Then the cost function is defined by

$$\mathcal{J}_\varepsilon(u) = \frac{1}{2}\sum_{j=1}^{n}\int_\Omega \Big(\Psi_1(t_j) - \Psi_1^{obs}(t_j)\Big)^2 \mathrm{d}s + \frac{\varepsilon}{2} \parallel R(u) \parallel_{\mathcal{T}}^2 .$$

The second term in the cost function is the regularization term in the sense of Tikhonov. It renders the inverse problem well posed, by taking into account the square of the potential vorticity of the initial state:

$$\parallel R(u) \parallel_{\mathcal{T}}^2 = \sum_{k=1}^{N} H_k\left[\int_\Omega \Big((\Delta\Psi_k)(0) - [W]_k.(\Psi)(0)\Big)^2 \mathrm{d}s\right].$$

The parameter ε in the cost function is the relative weight of the regularization with respect to the quadratic distance between the observations and the computed state.

The direct and the adjoint models are discretized using finite difference discretization for space and a leap-frog scheme for time.

As above, minimization is carried out with M1QN3. An important point is the choice of the inner product for the space of control. A systematic study was carried out in LUONG, BLUM and VERRON [1998] showing that the best choice is the inner product associated to the natural norm corresponding to the square root of the energy of the system, providing a good preconditioner for the optimization algorithm. Figures 3.1 and 3.2, showing the exact and the identified flow at the beginning and at the end of the assimilation period, prove that the reconstruction method is satisfactory.

The management of various time intervals, in order to improve the penetration of information in the deep layer, is presented in BLUM, LUONG and VERRON [1998]. A main difficulty comes from the dimension of the control space making this method costly from a computational viewpoint. Starting from a statistical analysis of the trajectory of the model, BLAYO, BLUM and VERRON [1998] proposed a method for the reduction of this space using POD vectors, which take into account the dynamics of the system (see Section 6).

For operational purpose, 4D-VAR methods have been implemented on the primitive equations model, by using an incremental method in order to reduce the cost of the resolution of the variational problem (THEPAUT and COURTIER [1987], COURTIER, THEPAUT and HOLLINGSWORTH [1994]).

4. Second-order methods

4.1. Hessian

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 definite positive. 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 10^{16} 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 composed of the direct and adjoint models (LE DIMET, NAVON and DAESCU [2002], WANG, NAVON, LE DIMET and ZOU [1992]), 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. Since the system is linear with respect to the adjoint variable, no linearization is necessary.

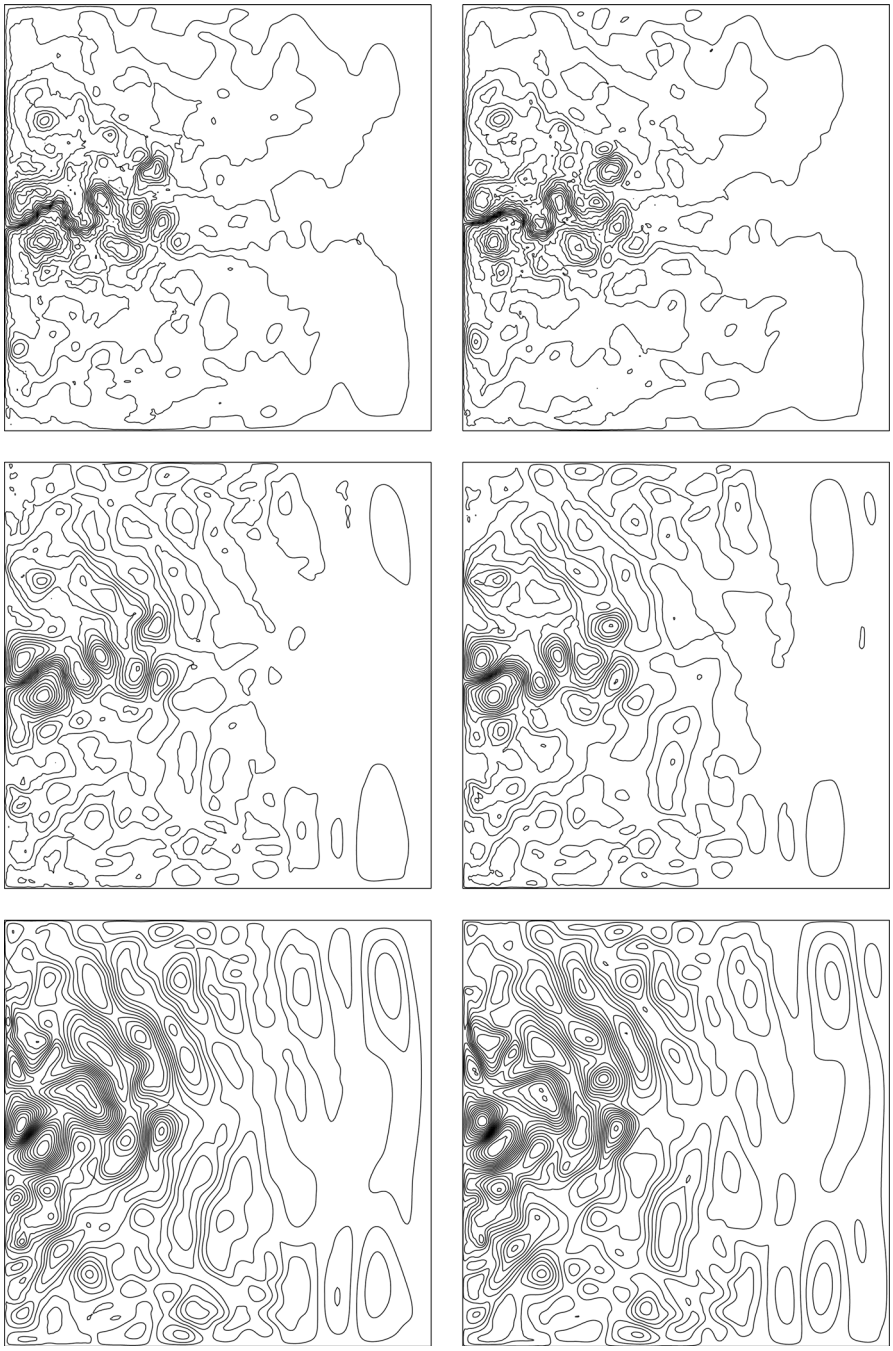


FIG. 3.1 True initial condition (left) and exact solution at the end of the assimilation period (right) for the three (from top to bottom) layers of a quasi-geostrophic model.

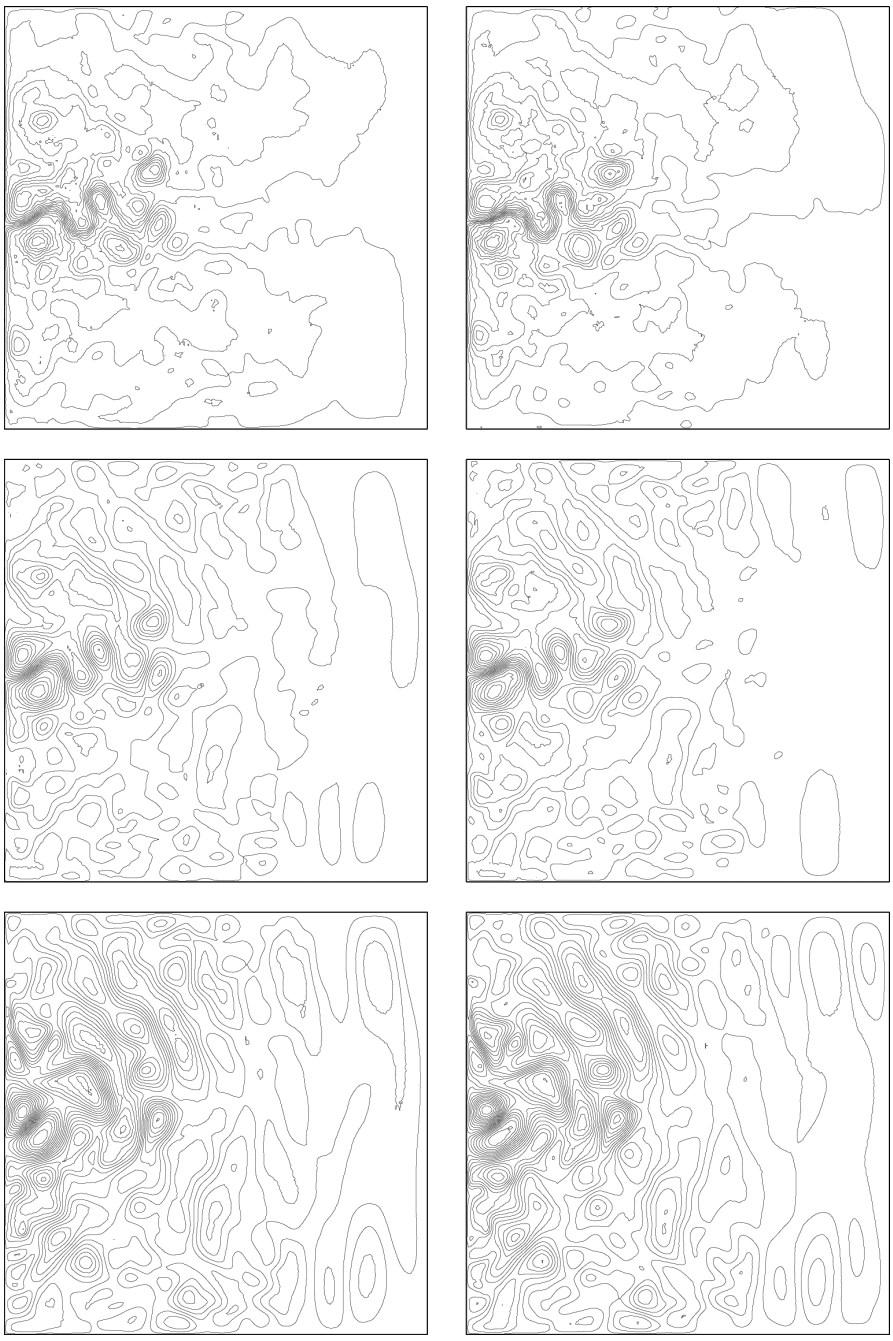


FIG. 3.2 4D-VAR data assimilation results identified initial condition (left) and corresponding solution at the end of the assimilation period (right) for the three (from top to bottom) layers of a quasi-geostrophic model.

- Introducing two second-order adjoint variables.
- Transposition to exhibit the linear dependence with respect to the directions.

If the model (Eq. (3.1)) has the form

$$\begin{cases} \dfrac{\mathrm{d}X}{\mathrm{d}t} = F(X) + B.U \\ X(0) = V \end{cases},$$

U and V being the control variables, and if we consider the cost function defined by Eq. (3.2) and the adjoint equation given by Eq. (3.5), from a backward integration of this adjoint model, the gradient is deduced

$$\nabla J = \begin{pmatrix} \nabla_U J \\ \nabla_V J \end{pmatrix} = \begin{pmatrix} -B^T P \\ -P\,(0) \end{pmatrix}.$$

To calculate the second-order derivative of J with respect to U and V, we have to derive the optimality system (i.e., the model plus the adjoint system). By analogy to the first-order case, we introduce two so-called second-order adjoint variables R and Q as the solution of the system:

$$\begin{cases} \dfrac{\mathrm{d}R}{\mathrm{d}t} = \left[\dfrac{\partial F}{\partial X}\right].R + B.\Lambda \\ \dfrac{\mathrm{d}Q}{\mathrm{d}t} + \left[\dfrac{\partial F}{\partial X}\right]^T.Q = -\left[\dfrac{\partial^2 F}{\partial X^2}.R\right]^T.P + C^T C R, \end{cases} \tag{4.1}$$

where Λ has the dimension of U.

If the Hessian of J is written

$$H(U, V) = \begin{pmatrix} J_{U,U} & J_{U,V} \\ J_{U,V} & J_{V,V} \end{pmatrix}$$

and if system (Eq. (4.1)) is integrated with the conditions:

$$Q(T) = 0$$
$$R(0) = \Theta$$

and $\Lambda = 0$, then

$$J_{V,V}.\Theta = -Q(0),$$
$$J_{V,U}.\Theta = -B^T.Q.$$

Now if the system is integrated with the conditions:

$$Q(T) = 0,$$
$$R(0) = 0,$$

then we obtain

$$J_{U,U}.\Lambda = -B^T.Q.$$

Therefore, without an explicit computation of the Hessian, it is possible to compute the product of the Hessian by any vector and consequently, using classical methods of linear algebra, to evaluate its eigenvalues and eigenvectors and also to carry out Newton-type methods. It is worth pointing out that the R variable is the solution of the linear tangent model (when $\Lambda = 0$), and therefore no extra code has to be written in this case. The left-hand side of the equation verified by Q is the adjoint model, and only the code associated to its right-hand side has to be written.

In the case of the shallow water equations with the initial condition V as unique control vector (no model error), the state variable is $X = (u, v, \phi)$, the adjoint variable is $P = (\tilde{u}, \tilde{v}, \tilde{\phi})$, which is solution of Eqs. (3.13)–(3.15). For the second order, the variable $R = (\bar{u}, \bar{v}, \bar{\phi})$ is the solution of the linear tangent model (Eqs. (3.10)–(3.12)), while the variable $Q = (\hat{u}, \hat{v}, \hat{\phi})$ is the solution of the equations:

$$\begin{aligned}
&\frac{\partial \hat{u}}{\partial t} + u\frac{\partial \hat{u}}{\partial x} + v\frac{\partial \hat{v}}{\partial y} + \hat{u}\frac{\partial v}{\partial y} - \hat{v}\frac{\partial v}{\partial y} - f\hat{v} + \phi\frac{\partial \hat{\phi}}{\partial x} \\
&\quad = \tilde{v}\frac{\partial \bar{v}}{\partial x} - \bar{u}\frac{\partial \tilde{u}}{\partial x} - \bar{v}\frac{\partial \bar{u}}{\partial y} + \tilde{u}\frac{\partial \bar{v}}{\partial y} - \bar{\phi}\frac{\partial \tilde{\phi}}{\partial x} - \bar{u}, \qquad (4.2)\\
&\frac{\partial \hat{v}}{\partial t} + \hat{u}\frac{\partial u}{\partial y} - u\frac{\partial \hat{v}}{\partial x} + \hat{v}\frac{\partial u}{\partial x} + v\frac{\partial \hat{v}}{\partial y} + f\hat{u} + \phi\frac{\partial \hat{\phi}}{\partial y} \\
&\quad = \tilde{u}\frac{\partial \bar{u}}{\partial x} - \bar{u}\frac{\partial \tilde{v}}{\partial x} - \tilde{v}\frac{\partial \bar{u}}{\partial y} + \bar{u}\frac{\partial \tilde{v}}{\partial y} - \bar{\phi}\frac{\partial \tilde{\phi}}{\partial y} - \bar{v}, \qquad (4.3)\\
&\frac{\partial \hat{\phi}}{\partial t} + \frac{\partial \hat{u}}{\partial x} + \frac{\partial \hat{v}}{\partial y} + u\frac{\partial \hat{\phi}}{\partial x} + v\frac{\partial \hat{\phi}}{\partial y} = -\bar{u}\frac{\partial \tilde{\phi}}{\partial x} - \bar{v}\frac{\partial \tilde{\phi}}{\partial x} - \gamma\bar{\phi}. \qquad (4.4)
\end{aligned}$$

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

The system obtained, i.e., 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 obtain the complete Hessian.

The determination of this product permits access to 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 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 method which requires computing the product Hessian-vector.

4.2. 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 as follows:

- X is the state vector of the model and K a vectorial parameter of the model $F(X, K) = 0$.
- $G(X, K)$ 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

$$\frac{\partial G}{\partial e_i} \simeq \frac{G(X(K + \alpha e_i), K + \alpha e_i) - G(X(K), K)}{\alpha}.$$

  The main inconvenience of this method is its computational cost: it requires solving the model as many times as the dimension of K. Furthermore, the determination of the parameter α may be tricky. If it is too large, the variation of G could be nonlinear, while for small values roundoff 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:

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

  Then the gradient is given by

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

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.

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, and they take place only as a forcing term 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 Le Dimet, Ngodock, Luong and Verron [1997] 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 as it has been seen in the previous subsection.

An important problem is the propagation of errors from models and observations toward the predicted fields. Second-order methods provide important tools for this purpose, especially for the estimation of the covariance of errors for the background term (prediction error) and the model error (see Le Dimet, Shutyaev and Gejadze [2006], Parmuzin, Le Dimet and Shutyaev [2006], Gejadze, Le Dimet and Shutyaev [2007]).

5. Nudging method

Nudging is a four-dimensional data assimilation (NDA) method that uses dynamical relaxation to adjust toward observations (observation nudging) or toward an analysis (analysis nudging).

Nudging is accomplished through the inclusion of a forcing term in the model dynamics, with a tunable coefficient that represents the relaxation time scale. Computationally inexpensive nudging is based on both heuristic and physical considerations.

The NDA method relaxes the model state toward the observations during the assimilation period by adding a nonphysical diffusive-type term to the model equations. The nudging terms are defined as the difference between the observation and the model solution multiplied by a nudging coefficient. The size of this coefficient is chosen by numerical experimentation so as to keep the nudging terms small in comparison to the dominating forcing terms in the governing equations, in order to avoid the rebounding effect that slows down the assimilation process, yet large enough to impact the simulation. NDA techniques have been used successfully on the global scale by Lyne, Swinbank and Birch [1982] and Krishnamurti, Jishan, Bedi, Ingles and Oosterhof [1991] and in a wide variety of research applications on mesoscale models (Hoke and Anthes [1976], Ramamurthy and Carr [1987], Ramamurthy and Carr [1988], Wang and Warner [1988], Stauffer and Seaman [1990], Verron, Molines and Blayo [1992] to cite but a few).

The NDA method is a flexible assimilation technique which is computationally much more economical than the VDA method. However, results from NDA experiments are quite sensitive to the adhoc specification of the nudging relaxation coefficient, and it is not at all clear how to choose a nudging coefficient so as to obtain an optimal solution (Lorenc [1986], Lorenc [1988]).

5.1. Optimal nudging specification

We assume that the model equations have been discretized in space by a finite difference, finite element, or spectral discretization method. The time continuous model satisfies dynamical equations of the form

$$\frac{\partial X}{\partial t} = F(X), \tag{5.1}$$

$$X(0) = V, \tag{5.2}$$

where X represents the discretized state variable of the model atmosphere, t is time, and V represents the initial condition for the model. Say, for instance, $X^o(t)$ is a given observation, then the objective of VDA is to find model initial conditions that minimize a cost function defined by

$$J(V) = \int_0^T \left\langle W(X - X^o), X - X^o \right\rangle \mathrm{d}t, \tag{5.3}$$

where W is a diagonal weighting matrix. Note that J is only a function of the initial state because X is uniquely defined by the model equations (Eqs. (5.1) and (5.2)). An implicit assumption made in VDA is that the model exactly represents the state of the atmosphere. However, this assumption is not true.

The NDA technique introduced by ANTHES [1974] consists in achieving a compromise between the model and the observations by considering the state of the atmosphere to be defined by

$$\frac{\partial X}{\partial t} = F(X) + G(X^o - X), \tag{5.4}$$

where G is a diagonal matrix.

Together with the initial conditions

$$X(0) = V, \tag{5.5}$$

the system (Eq. (5.1)) has a unique solution $X(V, G)$. The main difficulty in the NDA scheme resides in the estimation of the nudging coefficient G (STAUFFER and SEAMAN [1990]). If G is too large, the fictitious diffusion term will completely dominate the time tendency and will have an effect similar to replacing the model data by the observations at each time-step. Should a particular observation have a large error that prevents obtaining a dynamic balance, an exact fit to the observation is not required since it may lead to a false amplification of observational errors. On the other hand, if G is too small, the observation will have little effect on the solution. In general, G decreases with increasing observation error, increasing horizontal and vertical distance separation, and increasing time separation. In the experiment of ANTHES [1974], a nudging coefficient of 10^{-3} was used for all the fields for a hurricane model and was applied on all the domain of integration. In the experiment of KRISHNAMURTI,

Jishan, Bedi, Ingles and Oosterhof [1991], the relaxation coefficients for the estimated NDA experiment were kept invariant both in space and time, and their values were simply determined by numerical experience. The implicit dynamic constraints of the model then spread the updated information to the other variables (temperature and moisture) resulting eventually in a set of balanced conditions at the end of the nudging period.

In the work of Zou, Navon and Le Dimet [1992], a new parameter estimation approach was designed to obtain optimal nudging coefficients. They were optimal in the sense that the difference between the model solution and the observations will be small. For a comprehensive review of parameter estimation addressing issues of identifiability, see Navon [1998].

5.2. *Parameter estimation of optimal nudging coefficients*

The application of the variational approach to determine model parameters is conceptually similar to that of determining the initial conditions. Here, we present a brief illustration of the method. For the parameter estimation of the nudging coefficients, the cost function J can be defined as

$$J(G) = \int_0^T \left\langle W(X - X^o), X - X^o \right\rangle \mathrm{d}t + \left\langle K(G - \hat{G}), G - \hat{G} \right\rangle, \tag{5.6}$$

where $\hat{G}$ denotes the estimated nudging coefficients, and W and K are specified weighting matrices. The second term plays a double role. On one hand, it ensures that the new value of the nudging parameters is not too far away from the estimated quantity. On the other hand, it enhances the convexity of the cost function since this term contributes a positive term K to the Hessian matrix of J. An optimal NDA procedure can be defined by the optimal nudging coefficients G^* such that

$$J(G^*) \leq J(G), \quad \forall G. \tag{5.7}$$

The problem of extracting the dynamical state from observations is now identified as the mathematical problem of finding initial conditions or external forcing parameters that minimize the cost function. Due to the dynamical coupling of the state variables to the forcing parameters, the dynamics can be enforced through the use of a Lagrange function constructed by appending the model equations to the cost function as constraints in order to avoid the repeated application of the chain rule when differentiating the cost function. The Lagrange function is defined by

$$L(X, G, P) = J + \int_0^T \left\langle P, \frac{\partial X}{\partial t} - F(X) - G(X^o - X) \right\rangle \mathrm{d}t, \tag{5.8}$$

where P is a vector of Lagrange multipliers. The Lagrange multipliers are not specified but computed in determining the best fit. The gradient of the Lagrange function must be

zero at the minimum point. This results in the following first-order conditions:

$$\frac{\partial L}{\partial X} = 0 \sim \text{adjoint model forced by } \ 2\mathbf{W}(X - X^o), \tag{5.9}$$

$$\frac{\partial L}{\partial P} = 0 \sim \text{direct model (Eq. (5.4))}, \tag{5.10}$$

$$\frac{\partial L}{\partial G} = 0 \sim \int_0^T - < P, X^o - X) > \mathrm{d}t + 2K(G - \hat{G}) = 0. \tag{5.11}$$

The solution of equations (Eqs. (5.9)–(5.11)) is called a stationary point of L. Even if the dynamical evolution operator is nonlinear, the equation $\left(\dfrac{\partial L}{\partial X} = 0\right)$ will be the same as those derived by constructing the adjoint of the linear tangent operator; the linearization is automatic due to the Lagrange function L being linear in terms of the Lagrange multipliers P. An important relation between the gradient of the cost function (Eq. (5.7)) with respect to parameters G and the partial derivative of the Lagrange function with respect to the parameters is

$$\nabla_G J(G) = \frac{\partial L}{\partial G} | \text{at stationary point}, \tag{5.12}$$

i.e., the gradient of the cost function with respect to the parameters is equal to the left hand side of Eq. (5.11) which can be obtained in a procedure where the model state P is calculated by integrating the direct model forward and then integrating the adjoint model backward in time with the Lagrange multipliers as adjoint variables. Using this procedure, we can derive the following expressions of the adjoint equation and gradient formulation:

$$\begin{cases} \dfrac{\mathrm{d}P}{\mathrm{d}t} + \left[\dfrac{\partial F}{\partial X}\right]^T . P - G^T . P = 2W(X - X^o) \\ P\,(T) = 0, \end{cases} \tag{5.13}$$

and

$$\nabla_G J = -\int_0^T \left\langle X^o - X,\, P \right\rangle \mathrm{d}t + 2K(G - \hat{G}). \tag{5.14}$$

We see that the adjoint equation of a model with a nudging term added is the same as that without a nudging term except for the additional term $-G^T P$ added to the left hand side of the adjoint equation. Having obtained the value of cost function by integrating the model (Eq. (5.4)) forward and the value of the gradient $\nabla_G J$ by integrating the adjoint equation (Eq. (5.13)) backward in time, any large-scale unconstrained minimization method can be employed to minimize the cost function and to obtain an optimal parameter estimation.

If both the initial condition and the parameter are controlled, the gradient of the cost function for performing the minimization would be

$$\nabla J = (\nabla_V J, \nabla_G J)^T, \tag{5.15}$$

where

$$\nabla_V J = -P(0). \tag{5.16}$$

Zou, Navon and Le Dimet [1992] have shown that estimated NDA, optimal NDA, and KF differ from each other in the choice of the weight matrix often called the gain matrix:

$$G_n^* \equiv W_n^f H_n^T \left(H_n W_n^f H_n^T + R_n\right)^T. \tag{5.17}$$

The VDA, on the other hand, takes both the model forecasts and the observations as perfect. It attempts to obtain an optimal initial condition which minimizes the cost function

$$J^f = E\left(X_n^f - X_n^o\right)^T \left(X_n^f - X_n^o\right). \tag{5.18}$$

The theoretical framework of estimation and control theory provides the foundation of data assimilation techniques. The estimated NDA and the KF are closer to the estimation theory, the VDA to the optimal control aspect while optimal NDA is a combination of both (see also Lorenc [1986]).

See also work of Vidard, Piacentini and Le Dimet [2003] on optimal estimation of nudging coefficients.

5.3. Back-and-forth nudging

The backward nudging algorithm consists in solving the state equations of the model backwards in time, starting from the observation of the state of the system at the final instant. A nudging term, with the opposite sign compared with the standard nudging algorithm, is added to the state equations, and the final obtained state is in fact the initial state of the system (Auroux [2008]). The idea is to consider that we have a final condition V_T in Eqs. (5.1) and (5.2) instead of an initial condition V and then to apply nudging to this backward model with the opposite sign of the feedback term (in order to have a well-posed problem). We obtain

$$\begin{cases} \dfrac{\partial \tilde{X}}{\partial t} = F(\tilde{X}) - G'(X^o - \tilde{X}), \quad T > t > 0, \\ \tilde{X}(T) = V_T. \end{cases} \tag{5.19}$$

The back and forth nudging algorithm, introduced in Auroux and Blum [2005], consists in solving first the forward nudging equation and then the direct system backwards in time with a feedback term whose sign is opposite to the one introduced in the forward equation. The "initial" condition of this backward resolution is the final state obtained by the standard nudging method. After resolution of this backward equation, one obtains an estimate of the initial state of the system. We repeat these forward and backward resolutions (with the feedback terms) until convergence of the algorithm:

$$\begin{cases} \dfrac{\partial X_k}{\partial t} = F(X_k) + G(X^o - X_k), \\ X_k(0) = \tilde{X}_{k-1}(0), \end{cases} \tag{5.20}$$

$$\begin{cases} \dfrac{\partial \tilde{X}_k}{\partial t} = F(\tilde{X}_k) - G'(X^o - \tilde{X}_k), \\ \\ \tilde{X}_k(T) = X_k(T), \end{cases} \tag{5.21}$$

with $\tilde{X}_0(0) = V$. Then, $X_1(0) = V$, and a resolution of the direct model gives $X_1(T)$ and hence $\tilde{X}_1(T)$. A resolution of the backward model provides $\tilde{X}_1(0)$, which is equal to $X_2(0)$, the new initial condition of the system, and so on.

This algorithm can be compared with the 4D-VAR algorithm, which consists also in a sequence of forward and backward resolutions. In this algorithm, even for nonlinear problems, it is useless to linearize the system, and the backward system is not the adjoint equation but the direct system with an extra feedback term that stabilizes the resolution of this ill-posed backward resolution.

Auroux and Blum [2005] proved the convergence of the BFN algorithm on a linear model, provided that the feedback term is large enough. Auroux and Blum [2008] discussed the choice of the gain matrices G and G′ and tested the algorithm for Lorenz, Burgers and QG models. This algorithm is hence very promising to obtain a correct initial state, with a very easy implementation because it does not require neither the linearization of the equations to obtain the adjoint model nor any minimization process.

6. POD model reduction methods application to geosciences and 4D-VAR data assimilation

6.1. Introduction

Interest in reduced cost of implementation of 4D-VAR data assimilation in the geosciences motivated research efforts aimed toward reducing dimension of control space without significantly compromising quality of the final solution. POD, also known, when restricted to a finite dimensional case and truncated after a few terms, as equivalent to principal component analysis (PCA) and as empirical orthogonal function (EOF) in oceanography and meteorology, has emerged as a method of choice to be employed in flow control and optimization.

POD is a procedure for extracting a basis for a modal decomposition from an ensemble of signals. POD was introduced in the context of analysis of turbulent flow by Lumley [1967], Berkooz, Holmes and Lumley [1993].

In other disciplines, the same procedure goes by the names of Karhunen-Loeve decomposition or PCA.

The POD method was independently rediscovered several times: Kosambi [1943], Loeve [1945], and Karhunen [1946]. For introductory discussion for POD in fluid mechanics, see Sirovich [1987a,b,c] and Holmes [1990]. The mathematical theory behind it is the spectral theory of compact, self-adjoint operators.

The POD is equivalent to PCA methodology which originated with the work of Pearson [1901], a means of fitting planes by orthogonal least squares also put forward by Hotelling [1933].

If the POD spectrum decays fast enough, practically all the support of the invariant measure is contained in a compact set. Roughly speaking, all the likely realizations in the ensemble can be found in a relatively small set of bounded extent. "Regularity of solutions" is a mathematical property describing, essentially, the rate of decay of the tail of the wave number spectrum of instantaneous solutions of a partial differential equation. The method of snapshots introduced by SIROVICH [1987] is a numerical procedure for saving time in computation of empirical eigenfunctions. KIRBY and SIROVICH [1990] applied the POD procedure directly to the reconstruction of images of human faces. See also KIRBY [2001].

Snapshot bases consist of the flow solution for several flow solutions corresponding to different sets of parameter values evaluated at different time instants of the model evolution. This involves solving the fully discretized model and saving states at various time instants in the time interval under consideration.

POD approximation can be thought of as a Galerkin approximation in the spatial variable, with basis functions corresponding to the solution of the physical system at prespecified time instances. These are called the snapshots.

Due to possible linear dependence or almost linear dependence, the snapshots themselves are not appropriate as a basis. Rather singular value decomposition (SVD) is carried out and the leading generalized eigenfunctions are chosen as a basis, referred to as the POD basis.

6.2. *POD: the discrete case*

We consider the discrete Karhunen-Loeve expansion to find an optimal representation of the ensemble of snapshots. In general, each sample of snapshots $u_i(\vec{x})$ (defined on a set of m nodal points $\vec{x}$) can be expressed as a dimensional vector as follows:

$$\vec{u}_i = [u_{i1}, \ldots, u_{im}]^T, \tag{6.1}$$

where u_{ij} denotes the jth component of the vector $\vec{u}_i$. The mean vector is given by

$$\bar{u}_k = \frac{1}{n} \sum_{i=1}^{n} u_{ik}, \quad k = 1, \ldots, m. \tag{6.2}$$

We also can form a new ensemble by focusing on deviations from the mean value as follows:

$$v_{ik} = u_{ik} - \bar{u}_k, \quad k = 1, \ldots, m \tag{6.3}$$

Let the matrix A denotes the new ensemble

$$A = \begin{pmatrix} v_{11} & v_{21} & \cdots & vn1 \\ v_{12} & v_{22} & \cdots & v_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ v_{1m} & v_{2m} & \cdots & v_{mm} \end{pmatrix}_{m \times n},$$

where the discrete covariance matrix of the ensemble u may be written as

$$Cy_k = AA^T y_k = \lambda_k y_k. \tag{6.4}$$

Thus, to compute the POD mode, one must solve an eigenvalue problem.

For a discretization of an ocean problem, the dimension often exceeds 10^4 so that a direct solution of this eigenvalue problem is often not feasible. We can transform the $m \times m$ eigenvalue problem into an n eigenvalue problem. In the method of snapshots, one then solves the $n \times n$ eigenvalue problem

$$Dw_k = A^T A w_k = \lambda_k w_k, \quad w_k \in R^n, \tag{6.5}$$

where $1 \leq \lambda_k \leq n$ are the eigenvalues. The eigenvectors w_k may be chosen to be orthonormal, and the POD modes are given by $\phi_k = Aw_k/\sqrt{\lambda_k}$. In matrix form, with $\Phi = [\phi_1, \ldots, \phi_n]$, and $W = [w_1, \ldots, w_n]$, this becomes $\Phi = AW$.

The $n \times n$ eigenvalue problem (Eq. (6.4)) is more efficient than the $m \times m$ eigenvalue problem (Eq. (6.4)) when the number of snapshots n is much smaller than the number of states m.

6.3. POD 4D-VAR

In order to reduce the computational cost of 4D-VAR data assimilation, we consider minimization of the cost functional in a space whose dimension is much smaller than that of the original one. A way to drastically decrease the dimension of the control space without significantly compromising the quality of the final solution but sizably decreasing the cost in memory and CPU time of 4D-VAR motivates us to choose to project the control variable on a basis of characteristic vectors capturing most of the energy and the main directions of variability of the model, i.e., SVD, EOF, Lyapunov, or bred vectors. One would then attempt to control the vector of initial conditions in the reduced space model.

In the 1990s, most efforts of model reduction have been centered on KF and extended Kalman filter (EKF) data assimilation techniques, see Todling and Cohn [1994], Todling, Cohn and Sivakumaran [1998], Pham, Verron and Roubaud [1998], Cane, Kaplan, Miller, Tang, Hackert and Busalacchi [1996], Fukumori and Malanotte-Rizzoli [1995], Verlaan and Heemink [1997] and Hoteit and Pham [2003]. In particular, Cane, Kaplan, Miller, Tang, Hackert and Busalacchi [1996] employed a reduced-order method in which the state space is reduced through the projection onto a linear subspace spanned by a small set of basis functions, using an EOF analysis. This filter is referred to as the reduced rank EKF (see next section).

Some initial efforts aiming at the reduction of the dimension of the control variable, referred to as reduced-order strategy for 4D-VAR ocean data assimilation, were put forward initially by Blayo, Blum and Verron [1998] and Durbiano [2001] and more recently by Hoteit and Kohl [2006] and Robert, Durbiano, Blayo, Verron, Blum and Le Dimet [2005]. They used a low dimension space based on the first few EOFs, which can be computed from a sampling of the model trajectory. Hoteit and Kohl [2006] used the reduced-order model for part of the 4D-VAR assimilation then switched to the full model in a manner done earlier by Peterson [1989].

For a comprehensive description of POD theory and state of the art research, see GUNZBURGER [2003].

At the analysis time $[0, T_N]$, strong constraint 4D-VAR looks to minimize a cost function

$$J(U_0) = (U_0 - U_b)^T B^{-1}(U_0 - U_b) + (HU - y^0)^T O^{-1}(HU - y^0). \tag{6.6}$$

In POD 4D-VAR, we look to minimize the cost function

$$\begin{aligned} J(c_1(0), \ldots, c_M(0)) = {} & (U_0^{POD} - U_b) B^{-1}(U_0^{POD} - U_b) \\ & + (HU^{POD} - y^0) O^{-1}(HU^{POD} - y^0), \end{aligned} \tag{6.7}$$

where U_0^{POD} is the control vector, H is an observation operator, B is the background error covariance matrix, and O is the observation error covariance matrix. In Eq. (6.7),

$$U_0^{POD}(x) = U_0^{POD}(0, x) = \overline{U}(x) + \sum_{i=1}^{M} c_i(0)\Phi_i(x), \tag{6.8}$$

$$U^{POD}(x) = U^{POD}(t, x) = \overline{U}(x) + \sum_{i=1}^{M} c_i(t)\Phi_i(x). \tag{6.9}$$

In POD 4D-VAR, the control variables are $c_1(0), \cdots, c_M(0)$. As shown later, the dimension of the POD reduced space could be much smaller than that of the original space. In addition, the forward model is the reduced model which can be very efficiently solved. The adjoint model is used to calculate the gradient of the cost function (Eq. (6.7)) and that will greatly reduce both the computational cost and coding effort.

To establish POD model in POD 4D-VAR, we need first to obtain an ensemble of snapshots, which is taken from the background trajectory, or integrate original model with background initial conditions.

6.4. Adaptive POD 4D-VAR

Since the POD model is based on the solution of the original model for a specified initial condition, it might be a poor model when the new initial condition is significantly different from the one on which the POD model is based upon. Therefore, an adaptive POD 4D-VAR procedure is as follows:

(i) Establish POD model using background initial conditions and then perform optimization iterations to approximate the optimal solution of the cost function (Eq. (6.7)).
(ii) If after a number of iterations, the cost function cannot be reduced significantly as measured by a preset criterion, we generate a new set of snapshots by integrating the original model using the newest initial conditions.
(iii) Establish a new POD model using the new set of snapshots and continue optimization iteration.
(iv) Check if the optimality conditions are reached, if yes, stop; if no, go to step (ii).

6.5. Methods for 4D-VAR POD adaptivity

Working with a low-dimensional model during the computation of a control problem solution has to face the problem that these reduced models are possibly unreliable models if they are not correctly updated during the optimization process.

Consequently, some iterative technique is required, in which the construction of reduced-order models is coupled with the progress of the optimization process. Such an approach leads to the use of reduced-order models that adequately represent the flow dynamics as altered by the control. Crucial at this point is to decide whether or not the reduced-order model has to be adapted to a new flow configuration.

In adaptivity based on a trust-region method (Fahl and Sachs [2003]), the range of validity of a reduced-order model is automatically restricted, and the required update decision for the reduced-order models can be made by employing information that is obtained during the control problem solution.

Ravindran [2002] and Ravindran [2006] propose an adaptive procedure that successively updates the reduced-order model to be used in a sequential quadratic programming constrained optimization algorithm.

6.6. Goal-oriented model-based reduction

Bui-Thanh, Willcox, Ghattas and Van Bloemen Waanders [2007], Willcox, Ghattas, Van Bloemen Waanders and Bader [2005], and Daescu and Navon [2008] proposed an alternative method to determine the reduced-space basis. This method seeks to minimize an error similar in form to Eq. (6.7) to be presented below; however, it will improve upon the POD, first, by minimizing the error in the outputs (as opposed to states) and, second, by imposing additional constraints that $\hat{u}^k(t)$ should result from satisfying the reduced-order governing equations for each parameter instance k.

For a fixed basis size, the POD basis therefore minimizes the error between the original snapshots and their representation in the reduced space defined by

$$E = \sum_{k=1}^{S} \sum_{j=1}^{T} [u^k(t_j) - \tilde{u}^k(t_j)]^T [u^k(t_j) - \tilde{u}^k(t_j)], \tag{6.10}$$

where

$$\tilde{u}^k(t_j) = \Phi \Phi^T u^k(t_j), \tag{6.11}$$

Here, $u^k(t_j)$, $j = 1, \cdots, T$; $k = 1, \ldots, S$ is a snapshot, i.e., the solution of the governing equations at time t_j for parameter instance k. T time instants are considered for each parameter instance, yielding a total of ST snapshots.

The projection matrix $\Phi \in R^{N \times m}$ contains as columns the basis vectors ϕ_i, i.e.,

$$\Phi = [\phi_1, \phi_2, \ldots, \phi_m], \tag{6.12}$$

This error is equal to the sum of the singular values corresponding to those singular vectors (SVs) not included in the POD basis:

$$E = \sum_{i=m+1}^{ST} \sigma_i, \tag{6.13}$$

where σ_i is the ith singular value of U. The POD is an optimal basis in the sense that it minimizes the data reconstruction error given by Eq. (6.7). The goal-oriented, model-based optimization approach presented here provides a general framework for construction of reduced models, and is particularly applicable for optimal design, optimal control, and inverse problems. The optimization approach provides significant advantages over the usual POD by allowing the projection basis to be targeted to output functionals, by providing a framework to consider multiple parameter instances, and by incorporating the reduced-order governing equations as constraints in the basis derivation (see also MEYER and MATTHIES [2003]). Using this method, it is possible to obtain an a priori error estimate for a certain target functional of the solution. This error estimate can be used for adaptively resizing the number of basis vectors and the length of the time-step to satisfy a given error tolerance. It can also be used to form a very efficient low-dimensional basis especially tailored to the target functional of interest. This basis yielded a significantly better approximation of the functional when compared with conventionally chosen bases (see DAESCU and NAVON [2008]).

6.7. *State of the art of POD research*

ROBERT, DURBIANO, BLAYO, VERRON, BLUM and LE DIMET [2005] apply POD reduced-order modeling in a twin experiment setup for a primitive equation model of the equatorial Pacific Ocean model using an incremental formulation and using a background covariance matrix in the reduced space, obtaining a fast convergence of the minimization of the cost functional.

In a related work, ROBERT, BLAYO and VERRON [2006] applied reduced-order 4D-VAR as a preconditioner to incremental 4D-VAR data assimilation method. They used a low-dimensional space based on first few EOFs chosen from sampling of the model trajectory. See also work of LAWLESS, NICHOLS, BOESS and BUNSE-GERSTNER [2006] using a linear balanced truncation reduced-order modeling as the preconditioner of the inner iteration of an incremental 4D-VAR using a 1D shallow water equation model.

CAO, ZHU, LUO and NAVON [2006] proposed for the first time a 4D-VAR approach based on POD. Their proposed POD-based 4D-VAR methods are tested and demonstrated using a reduced gravity wave ocean model in Pacific domain in the context of identical twin data assimilation experiments.

LUO, CHEN, ZHU, WANG and NAVON [2007] present an error estimate of a new reduced-order optimizing finite difference system (FDS) model. Numerical examples are presented illustrating that the error between the POD approximate solution and the full FDS solution is consistent with previously obtained theoretical results. The preconditioning aspect of POD for efficient optimization is another topic of active research. DAESCU

and NAVON [2008] use a POD approach to model reduction to identify a reduced-order control space for a 2D global shallow water model. A reduced second-order adjoint model is developed and used to facilitate the implementation of a Hessian-free truncated-Newton (HFTN) minimization algorithm in the POD-based space. The HFTN algorithm benefited most from the order reduction since computational savings were achieved both in the outer and inner iterations of the method.

For use of centroidal Voronoi tessellations (CVT) combined with POD, see the work of BURKARDT, GUNZBURGER and LEE [2006]. Here, POD and CVT approaches to reduced-order modeling are provided, including descriptions of POD and CVT reduced-order bases, their construction from snapshot sets, and their application to the low-cost simulation of a Navier-Stokes system (see also GUNZBURGER, PETERSON and SHADID [2007]). Direct and inverse POD model reduction was applied to a 3D time-dependent finite element adaptive ocean model (Imperial College Ocean Model, ICOM) (FANG, PAIN, NAVON, PIGGOTT, GORMAN and GODDARD [2008]) (see Fig. 6.1).

A novel POD model has been developed for use with an advanced unstructured mesh finite element ocean model, the ICOM, which includes many recent developments in ocean modeling and numerical analysis. The advantages of the POD model developed over existing POD approaches are the ability to increase accuracy when representing geostrophic balance (the balance between the Coriolis terms and the pressure gradient). This is achieved through the use of two sets of geostrophic basis functions where each one is calculated by basis function for velocities u and v.

When adaptive meshes are employed in both the forward and adjoint models, the mesh resolution requirements for each model may be spatially and temporally different as the meshes are adapted according to the flow features of each model. This unavoidably brings to difficulties in the implementation of a POD-based reduced model for an inverse adaptive model. Such challenges include snapshots can be of different length at different time levels and the POD base of the forward model can differ from the POD base of

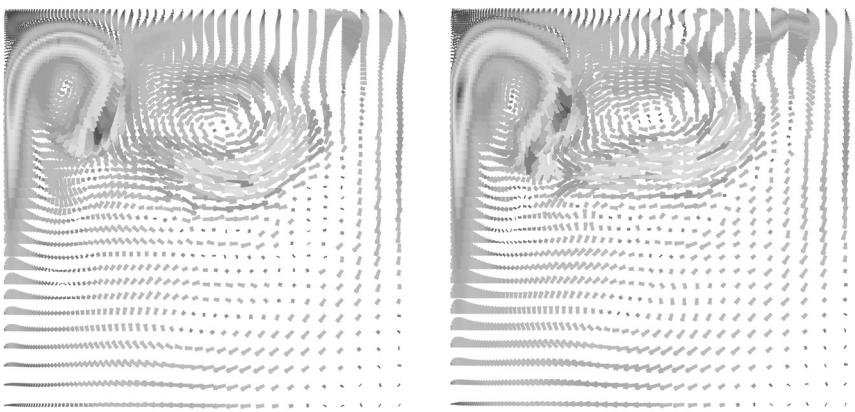


FIG. 6.1 Application of POD model reduction method to Imperial College Ocean Model (ICOM) adaptive finite element ocean Model. Full model and reduced-order model based on first 30 base functions.

the adjoint model. To overcome these difficulties, a standard reference fixed mesh is adopted for both the forward and adjoint reduced models. The solutions for both are interpolated from their own mesh onto the same reference fixed mesh at each time level. This allows the same number of base modes for both reduced forward and adjoint models. The referenced mesh can also be obtained by superimposing the resolution at each mesh level associated with a goal-based function.

7. Data assimilation with EnKF

7.1. Introduction

In recent years, two trends for operational data assimilation are prevalent for the data assimilation practitioners.

On one hand are the variational methods subdivided between computationally economical 3D-VAR methods which exclude the flow-dependent forecast errors (see PARRISH and DERBER [1992]) while few centers endowed with powerful computing resources adopted 4D-VAR, requiring availability and constant updating of an adjoint model. The latter requires a computationally demanding effort but is significantly more accurate than 3D-VAR (KALNAY, LI, MIYOSHI, YANG and BALLABRERA-POY [2007]) in pretest implementation comparisons. Moreover 4D-VAR allows the assimilation of asynoptic data at their correct observation time along with other advantages such as possibility of inclusion of model error term as a weak constraint 4D-VAR.

On the other hand, in view of the obvious shortcomings of usual KF and EKF, more efficient filter methods have emerged, obviating the prohibitive storage and computational time due to explicit treatment of the state error covariance matrix for KF and EKF such as the EnKF. These new filter algorithms are of special interest due to their simplicity of implementation since no adjoint operators are required, along with their potential for efficient use on parallel computers with large-scale geophysical models (NERGER, HILLER and SCHRATER [2005]). Research on EnKF started with work of EVENSEN [1994], EVENSEN and LEEUWEN [1996], BURGERS, VAN LEEUWEN and EVENSEN [1998], and HOUTEKAMER and MITCHELL [1998]. Their methods can be classified as perturbed observations (or stochastic) EnKF and are essentially a Monte-Carlo approximation of the KF which avoids evolving the covariance matrix of the pdf of the state vector x. A second type of EnKF is a class of square-root (or deterministic) filters (ANDERSON [2003], BISHOP, ETHERTON and MAJUMDAR [2001], WHITAKER and HAMILL [2002], see review of TIPPETT, ANDERSON, BISHOP, HAMILL and WHITAKER [2003]), which consist of a single analysis based on the ensemble mean, and where the analysis perturbations are obtained from the square root of the KF analysis error covariance.

Several variants of the EnKF have been proposed (ANDERSON [2003], BISHOP, ETHERTON and MAJUMDAR [2001], WHITAKER and HAMILL [2002]) which can be interpreted as ensemble square-root KFs. For an improved treatment of nonlinear error evolution in EKF, the singular evolutive interpolated Kalman filter (PHAM, VERRON and ROUBAUD [1998]) was introduced as a variant of the singular evolutive extended Kalman (SEEK) filter. It combines the low-rank approximation with an ensemble representation of the covariance matrix. This idea has also been followed in the concept

of error subspace statistical estimation (Lermusiaux and Robinson [1999]). Another approach is based on a low-rank approximation of the state covariance matrix of the EKF to reduce the computational costs. Using finite difference approximations for the tangent linear model, these algorithms display better abilities to treat nonlinearity as compared with the EKF. Examples of low-rank filters are the reduced rank square-root algorithm (Verlaan and Heemink [1995]) and the similar SEEK filter (Pham, Verron and Roubaud [1998]).

We will first present in a short section the basic linear KF (Kalman [1960]) followed by the EKF and then devote our attention to a brief survey of various flavors of EnKF and its state-of-the-art implementation. See Ghil and Manalotte-Rizzoli [1991] for equivalence between 4D-VAR with strong constraint and the linear KF and Li and Navon [2001].

Finally, some open issues of advantages of EnKF versus 4D-VAR as have emerged from recent work of Kalnay, Li, Miyoshi, Yang and Ballabrera-Poy [2007] will be briefly addressed.

7.2. The KF

The KF is an efficient recursive filter that estimates the state of a dynamic system from a series of incomplete and noisy measurements. It was developed by Kalman [1960]. The KF has been derived in a number of books on control theory, e.g., Gelb, Kasper, Nash, Price and Sutherland [1974] and Jazwinski [1970] to mention but a few. See also early work of Du Plessis [1967]. In oceanography, the KF has been used by Budgell [1986] to describe nonlinear and linear shallow water wave propagation in branched channels, using one-dimensional (1D) cross-sectionally integrated equations. Miller [1986] used a 1D linear barotropic QG model to investigate the properties of the KF. He provided a derivation of the KF equations. In meteorology, Ghil [1980] and Ghil et al. [1981] promoted first the use of KF along with Cohn, Ghil and Isaacson [1981] and Cohn [1997]. See also Miller, Ghil and Gauthiez [1994]. Ghil [1989] discussed the KF as a data assimilation method in oceanography and used it with a simple linear barotropic model. The KF for use in meteorology has recently been addressed in work of Cohn and Parrish [1991] who discussed the propagation of error covariances in a 2D linear model.

7.3. The KF formulation

The KF is a recursive filter that estimates the state of a dynamic system from a series of incomplete and noisy measurements.

Consider a linear observation process described by

$$y_k^0 = H_k x_k^t + e_k, \tag{7.1}$$

where k is a multiple of the number of time-steps between two consecutive observations in time. y_k^0 is the vector of observations while the vector e_k is an additive noise representing the error in observations due for instance to instrumental error. Random noise e_k is

assumed white in time with mean 0 and covariance R_k, i.e.,

$$E(e_k e_k^T) = R_k \mathrm{d}_k \mathrm{d}_k^T, \tag{7.2}$$

All the time, we consider a discrete in time stochastic dynamic system

$$x_k^t = M_{k-1} x_{k-1}^t + \eta_{k-1}, \tag{7.3}$$

where M_k represents model dynamics while η_k is model error white in time with mean zero and covariance Q_k,

$$E(h_k h_k^T) = Q_k \mathrm{d}_k \mathrm{d}_k^T, \tag{7.4}$$

One can show that the linear KF (Gelb, Kasper, Nash, Price and Sutherland [1974], Jazwinski [1970], Todling [1999]) consists of following stages:

- Advance in time:

$$\begin{cases} x_k^f = M_{k-1} x_{k-1}^a \\ P_k^f = M_{k-1} P_{k-1}^a M_{k-1}^T + Q_{k-1}, \end{cases} \tag{7.5}$$

  where the forecast and analysis error covariance matrices at time k are given by

$$\begin{cases} P_k^f = E\{(x_k^t - x_k^f)(x_k^t - x_k^f)^T\} \\ P_k^a = E\{(x_k^t - x_k^a)(x_k^t - x_k^a)^T\}. \end{cases} \tag{7.6}$$

  Q_{k-1} is the model error covariance matrix at time $t = t_{k-1}$, and M_{k-1} is the model dynamics. x_{k-1}^a and x_{k-1}^f are the analysis and the forecast at time $t = t_{k-1}$.
- Compute the Kalman gain:

$$K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1}. \tag{7.7}$$

  The matrix K_k is the optimal weighting matrix known as the Kalman gain matrix.
- Update the state:

$$x_k^a = x_k^f + K_k(y_k^0 - H_k x_k^f), \tag{7.8}$$

  where y_k^0 is the observation at time $t = t_k$, H_k is the observation matrix at time $t = t_k$.
- Update error covariance matrix:

$$P_k^a = (I - K_k H_k) P_k^f. \tag{7.9}$$

7.4. Computational cost of KF

The KF assuming the dynamical model has n unknowns in the state vector then error covariance matrix has n^2 unknowns. The evolution of the error covariance in time requires the cost of $2n$ model integrations. Thus, KF in usual form can only be used for rather low-dimensional dynamical models.

The basic KF is limited to a linear assumption. However, most non-trivial systems are nonlinear. The nonlinearity can be associated either with the process model or with the observation model or with both. In EKF, the state transition and observation models need not be linear functions of the state but may instead be functions.

$$\begin{cases} x_k = f(x_{k-1}, u_k, w_k) \\ z_k = h(x_k, v_k). \end{cases} \tag{7.10}$$

The function f can be used to compute the predicted state from the previous estimate, and similarly, the function h can be used to compute the predicted measurement from the predicted state. However, f and h cannot be applied to the covariance directly. Instead, a matrix of partial derivatives (the Jacobian or Tangent Linear Model) is computed. At each time-step, the Jacobian is evaluated with current predicted states. These matrices can be used in the KF equations. This process essentially linearizes the nonlinear function around the current estimate. This results in the following EKF equations:

- Predict:

$$\begin{cases} x_{k|k-1} = f(\hat{x}_{k-1|k-1}, u_k, 0) \\ P^a_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k, \end{cases} \tag{7.11}$$

- Update:

$$\begin{cases} \tilde{y}_k = z_k - h(\hat{x}_{k|k-1}, 0) \\ S_k = H_k P_{k|k-1} H_k^T + R_k \\ K_k = P_{k|k-1} H_k^T S_K^{-1} \\ x_{k|k-1} = f(\hat{x}_{k-1|k-1}, u_k, 0) \\ P_{k|k} = (I - K_k H_k) P_{k|k-1}, \end{cases} \tag{7.12}$$

  where the state transition and observation matrices are defined to be the following Jacobians

$$\begin{cases} F_k = \dfrac{\partial f}{\partial x}|_{\hat{x}_{k-1|k-1}, u_k} \\ H_k = \dfrac{\partial h}{\partial x}|_{\hat{x}_{k|k-1}}, \end{cases} \tag{7.13}$$

For use in meteorology, see GHIL and MANALOTTE-RIZZOLI [1991], GAUTHIER, COURTIER and MOLL [1993], and BOUTTIER [1994].

7.5. Shortcomings of the EKF

Unlike its linear counterpart, the EKF is not an optimal estimator. In addition, if the initial estimate of the state is wrong, or if the process is modeled incorrectly, the filter may quickly diverge, owing to its linearization. We have really effectuated a closure by discarding moments of third and higher order giving us an approximate equation for the error variance. Usefulness of EKF will depend on properties of the model dynamics. See discussion of MILLER, GHIL and GAUTHIEZ [1994]. EVENSEN [1992] provided the first application of EKF on a nonlinear ocean circulation model. Another problem with the EKF is that the estimated covariance matrix tends to underestimate the true covariance matrix and therefore risks becoming inconsistent in the statistical sense without the addition of "stabilizing noise."

7.6. EnKF

Here, we follow algorithmic presentation of MANDEL [2006]. The EnKF is a Monte Carlo approximation of the KF avoiding evolving the covariance matrix of the pdf of the state vector x. Instead, the probability distribution is represented by a sample

$$X = [x_1, x_2, \ldots, x_N] = [x_i]. \tag{7.14}$$

X is an $n \times N$ matrix whose columns are the ensemble members, and it is called the prior ensemble. Ideally, ensemble members would form a sample from the prior distribution. However, the ensemble members are not in general independent except in the initial ensemble since every EnKF step ties them together. They are deemed to be approximately independent, and all calculations proceed as if they actually were independent. Replicate the data $\mathbf{d}$ into a $m \times N$ matrix

$$D = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N] = [\mathbf{d}_i] \tag{7.15}$$

so that each column consists of the data vector $\mathbf{d}$ plus a random vector from the n-dimensional normal distribution $\mathcal{N}(0, R)$.

Because randomness is introduced in ENKF at every assimilation cycle, the algorithm updates every ensemble member to a different set of observations perturbed by a random noise.

For details, see work of HOUTEKAMER and MITCHELL [1998, 2001], HAMILL and SNYDER [2000, 2002], and more recently HOUTEKAMER, MITCHELL, PELLERIN, BUEHNER, CHARRON, SPACEK and HANSEN [2005]. If, in addition, the columns of X are a sample from the prior probability distribution, then the columns of

$$\hat{X} = X + K(D - HX) \tag{7.16}$$

form a sample from the posterior probability distribution. The EnKF is now obtained simply by replacing the state covariance Q in Kalman gain matrix

$$K = QH^T(HQH^T + R)^{-1} \tag{7.17}$$

by the sample covariance C computed from the ensemble members (called the ensemble covariance).

7.7. Implementation

7.7.1. Basic formulation

Suppose the ensemble matrix X and the data matrix D are as above. The ensemble mean and the covariance are

$$E(X) = \frac{1}{N}\sum_{k=1}^{N} x_k, \quad C = \frac{AA^T}{N-1}, \tag{7.18}$$

where

$$A = X - E(X) = X - \frac{1}{N}(Xe_{N\times 1})e_{1\times N}, \tag{7.19}$$

and e denotes the matrix of all ones of the indicated size. The posterior ensemble X^p is then given by

$$\hat{X} \simeq X^p = X + CH^T(HCH^T + R)^{-1}(D - HX), \tag{7.20}$$

where the perturbed data matrix D is as above. Since C can be written as

$$C = (X - E(X))(X - E(X))^T, \tag{7.21}$$

one can see that the posterior ensemble consists of linear combinations of members of the prior ensemble. Note that since R is a covariance matrix, it is always positive semidefinite and usually positive definite, so the inverse above exists and the formula can be implemented by the Cholesky decomposition. In EVENSEN [2004], R is replaced by the sample covariance $\dfrac{DD^T}{N-1}$ and the inverse is replaced by a pseudoinverse, computed using the SVD. Since these formulas are matrix operations with dominant Level 3 operations, they are suitable for efficient implementation using software packages such as LAPACK (on serial and shaby it, it is much better (several times cheaper and also more accurate) to compute the Cholesky decomposition of the matrix anred memory computers). Instead of computing the inverse of a matrix and multiplying d treat the multiplication by the inverse as solution of a linear system with many simultaneous right-hand sides. For complex Numerical Weather Prediction (NWP) models, deriving explicitly the background error covariance estimate from original method will be prohibitive

$$\hat{P}^b = \frac{\tilde{X}^b(\tilde{X}^b)^T}{m-1}, \tag{7.22}$$

where $\tilde{X}^b = (\tilde{x}_1^b K, \tilde{x}_m^b)$ and $\tilde{x}_i^b = x_i^b - \bar{x}^b$. Where the ensemble mean is defined by

$$\bar{x}^b = \frac{1}{m}\sum_{i=1}^{m} x_{x_i}^b.$$

Different parallel implementations of parallelized EnKF were proposed by Keppenne and Rienecker [2002] and other method by Reichle, McLaughlin and Entekhabi [2002], Reichle, Walker, Koster and Houser [2003]. Ensemble data assimilation algorithms with the assumptions of linear error dynamics and Gaussian error statistics will converge as the number of members of the ensemble increases to the state and covariance estimate of those of the EKF (Burgers, Van Leeuwen and Evensen [1998]).

7.8. Reduced rank KFs

Due to the expense involved in EKF, many simplified approaches have been proposed attempting to capture only a subset of the flow-dependent error covariances. The covariances evolve with the model dynamics within a specified reduced dimension subspace defined by a fixed set of basis functions (Todling and Cohn [1994]). Possible choices for the basis functions include EOFs (Cane, Kaplan, Miller, Tang, Hackert and Busalacchi [1996]), SVs (Cohn and Todling [1996], Fisher [1998]), or a balanced truncation of the Hankel operator (see Farrell and Ioannou [2001]). The EOF basis may not be optimal in the sense of providing the best subsequent forecast, which is often the goal of assimilating data. SVs or partially evolved SVs turn out to be more effective. SVs represent the directions that will evolve to optimally account for the error at a future time.

The SEEK filter for data assimilation in oceanography is a variant of the EKF with a low-rank error covariance matrix. It is quite similar in some aspects to reduced rank KF introduced by Cohn and Todling [1996], but differs in some aspects. It is derived from the EKF by approximating the state error covariance matrix by a matrix of reduced rank and evolving this matrix in decomposed form.

7.9. Algorithm of SEEK

For initialization, choose the initial estimate for the model state and an approximate state covariance matrix of low Rank in the decomposed form LUL^T. For forecast, evolve the guessed state with the full nonlinear model and the column vectors L_i with the tangent-linear model. For analysis, compute the updated state covariance matrix by an equation for the matrix U which relates the model state error to the observation error in the spirit of the Riccati equation. With this updated covariance matrix, the state update is given by the analysis step of the EKF. To avoid successive alignment of the vectors L_i, occasionally perform a reorthogonalization of these vectors (see Pham, Verron and Roubaud [1998], Carme, Pham and Verron [2001], Hoteit, Pham and Blum [2002, 2001], Hoteit and Pham [2003]).

7.10. Deterministic update ensemble filters

There is also a family of nonstochastic filters (see for instance Tippett, Anderson, Bishop, Hamill and Whitaker [2003]). These filters do not use perturbed observations, which, it is argued, can be a source of sampling error when ensembles are small. Instead,

they deterministically transform the ensemble of background fields into an ensemble of analyses using

$$P^a(t) = (I - K(t)H(t))P^f(t), \tag{7.23}$$

without adding random noise. However, this equation is valid only when the gain K is optimal, which depends, in turn, on Q and R being accurately known. A more general, but more complicated, equation for P^a, which reduces to this equation when the gain is optimal, is given by COHN [1982, Eq. (2.10b)], DALEY [1991, Eq. (13.3.19)], and GHIL and MANALOTTE-RIZZOLI [1991, Eq. (4.13b)]. The performance of stochastic and deterministic filters has been compared in a hierarchy of perfect-model scenarios by LAWSON, and HANSEN, [2004].

7.11. Ensemble square-root filter

The serial ensemble square-root filter (EnSRF) (WHITAKER and HAMILL [2002]) algorithm has been used for the assimilation at the scale of thunderstorms by SNYDER and ZHANG [2003], ZHANG, SNYDER and SUN [2004], and DOWELL, ZHANG, WICKER, SNYDER and CROOK [2004]. WHITAKER, COMPO, WEI and HAMILL [2004] used the algorithm for the global data assimilation of surface pressure observations. Similar to EnKF, the EnSRF conducts a set of parallel data assimilation cycles. In the EnSRF, one updates the equations for the ensemble mean (denoted by an overbar) and the deviation of the ith member from the mean separately:

$$\begin{cases} \bar{x}^a = \bar{x}^b + \hat{K}(y - H\bar{x}^b) \\ \bar{x}_i^a = \bar{x}_i^b - \tilde{K}H\bar{x}_i^b. \end{cases} \tag{7.24}$$

Here, $\hat{K}$ is the traditional Kalman gain, and $\tilde{K}$ is the reduced gain used to update deviations from the ensemble mean. In the EnSRF, the mean and departures from the mean are updated independently according to Eq. (7.24). If observations are processed one at a time, the EnSRF requires about the same computation as the traditional EnKF with perturbed observations, but for moderately sized ensembles and processes that are generally linear and Gaussian, the EnSRF produces analyses with significantly less error WHITAKER and HAMILL [2002]. Conversely, LAWSON, and HANSEN, [2004] suggest that if multimodality is typical and ensemble size is large, the EnKF will perform better. For details, see review of HAMILL [2006] in book of PALMER and HAGEDORN [2006].

7.12. Local ensemble Kalman filtering

Local ensemble Kalman filtering proposed by OTT, HUNT, SZUNYOGH, ZIMIN, KOSTELICH, CORAZZA, KALNAY, PATIL and YORKE [2004] and SZUNYOGH KOSTELICH and GYARMATI et al. [2005] treats local patches surrounding every grid point independently, thus avoiding correlations between distant points. This is preferable in view of the low-rank assumption for the error covariance matrix. These local patches are analyzed independently and are then combined to yield the global analysis.

7.13. Covariance localization

One aspect of ensemble assimilation methods is the requirement of accuracy for covariance matrices. Erroneous representation of error statistics affects the analysis-error covariance, which is propagated forward in time.

The covariance estimate from the ensemble is multiplied point by point with a correlation function that is 1.0 at the observation location and zero beyond some prespecified distance (correlation length).

Two approaches are used : one consists in a cut-off radius so that observations are not assimilated beyond a certain distance from the grid point (see HOUTEKAMER and MITCHELL [1998], EVENSEN [2003]). This may introduce spurious discontinuities.

The second approach is to use a correlation function that decreases monotonically with increasing distance. This results in the Kalman gain

$$K = P^b H^T (H P^b H^T + R)^{-1}, \tag{7.25}$$

being replaced by a modified gain

$$K = (\rho_s \circ P^b) H^T (H(\rho_s \circ P^b) H^T + R)^{-1}, \tag{7.26}$$

where the operation $\rho_s \circ$ denotes a Schur product (an element-by-element multiplication) of a correlation matrix S with local support with the covariance model generated by the ensemble. The Schur product of matrices A and B is a matrix C of the same dimension, where $c_{ij} = a_{ij} b_{ij}$. When covariance localization is applied to smaller ensembles, it may result in more accurate analyses than would be obtained from larger ensembles without localization HOUTEKAMER and MITCHELL [2001]. Localization increases the effective rank of the background error covariances HAMILL, WHITAKER and SNYDER [2001].

Generally, the larger the ensemble, the broader the optimum correlation length scale of the localization function (HOUTEKAMER and MITCHELL [2001], HAMILL, WHITAKER and SNYDER [2001]). See WHITAKER, COMPO, WEI and HAMILL [2004] and HOUTEKAMER, MITCHELL, PELLERIN, BUEHNER, CHARRON, SPACEK and HANSEN [2005] for examples performing ensemble assimilations that also include a vertical covariance localization.

8. Assimilation of images

The observation of the Earth by geostationary or polar-orbiting satellites clearly displays the evolution of some characteristics features such as fronts, the color, or the temperature of the ocean. Figure 8.1 represents the sea surface temperature (SST) of the Black Sea observed by the satellite Moderate Resolution Imaging Spectroradiometer (MODIS), and some geometric features are identified and their temporal evolution has an important informative content. At the present time, this information is used more in a qualitative fashion rather than in a quantitative one. The question arising is: is it possible to couple this information with a numerical model, i.e., how to assimilate images?

Two basic approaches can be considered:

- The first one (HERLIN, LE DIMET, HUOT and BERROIR [2004]) consists, in a first step, to extract from the images some "pseudo" measurements (e.g., surface velocity in

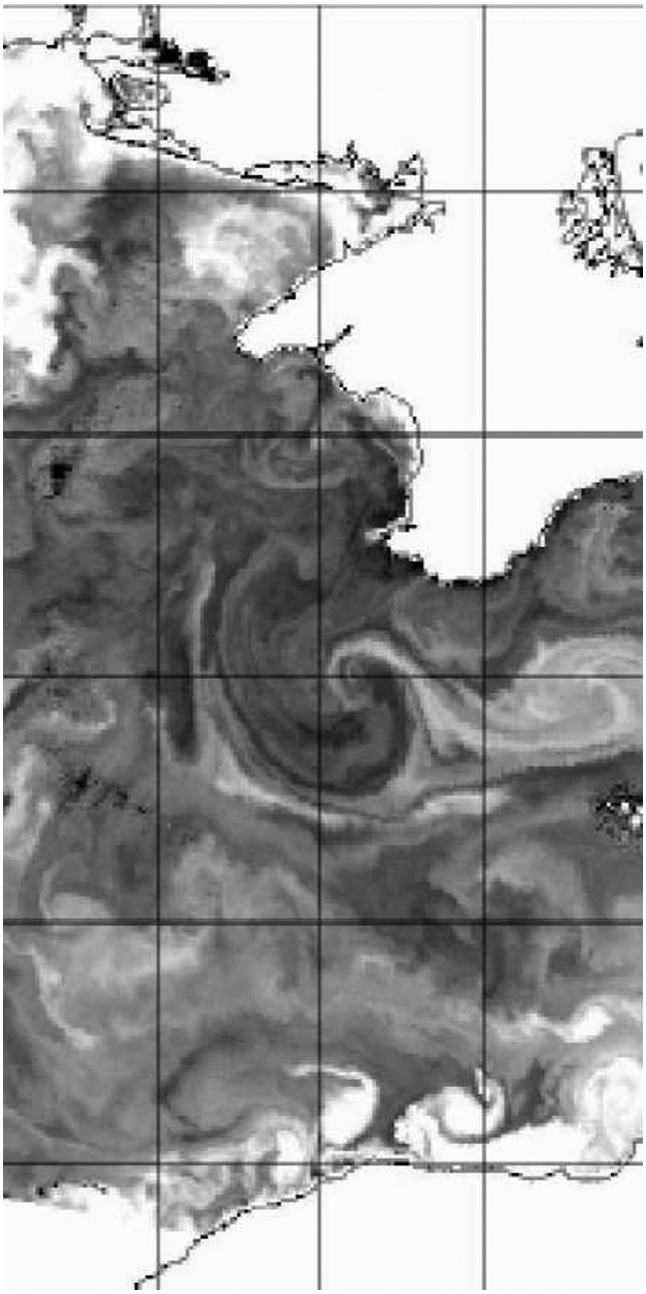


FIG. 8.1 Sea surface temperature in the Black Sea from MODIS.

oceanography, wind components in meteorology), then these measurements will be used in a classical scheme of VDA. For instance, in meteorology some identified clouds can be used as Lagrangian tracers, and assimilated as such. This information makes sense only if the altitudes of the clouds are known, it can be done by an evaluation of the temperature of the cloud and a comparison with the vertical profile of temperature.

- The second approach (MA, ANTONIADIS and LE DIMET [2006]) consists to consider images as objects and insert them directly in the variational analysis.

8.1. Retrieving velocities from images

This approach is a classical one in computer vision, and it is based on the conservation of grey level for individual pixels. Let us consider a pixel of coordinates $(x(t), y(t))$, if I is the luminance of the pixel, this quantity is conservative and its total derivative is equal to 0:

$$\frac{\mathrm{d}I}{\mathrm{d}t}(x(t), y(t), t) = 0.$$

By developing this expression we get

$$\frac{\partial I}{\partial x}.\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial I}{\partial y}.\frac{\partial y}{\partial t} + \frac{\partial I}{\partial t} = \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0,$$

u and v are the components of the velocity of the flow and are unknown. This equation is not sufficient to retrieve the velocity field, but it can be included in a variational formulation: we will seek for the velocity field minimizing the functional J defined by

$$J(u, v) = E_1 + E_2,$$

with

$$E_1 = \int_\Omega \left(\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t}\right)^2 d\Omega,$$

and

$$E_2 = \int_\Omega \|\nabla W\|^2 \, d\Omega,$$

with

$$W = (u, v).$$

E_2 can be considered as a regularization term to smooth the retrieved fields. E_1 is the scalar product of the gradient of luminance with the velocity field, and if these vectors are orthogonal, then the equation does not contain any quantitative information on the

velocity field. Some other conservation laws can be considered according to the nature of the image:

- With an image displaying the color of the ocean, an equation of chlorophyll conservation must be used, and it will have to include sink and source terms, and therefore an equation modeling biological processes must be added to the physical model.
- With an image of the SST, the Boussinesq approximation can be considered.

A selection of points on which the minimization of J is carried out to determine (u, v) has to be done. The structures where the velocity and the gradient of luminance are almost orthogonal must be discarded from the analysis, this being the case of filaments which are elongated structures. These structures are detected by applying operators of mathematical morphology: peak operator detecting brighter areas of maximum width and valley operator detecting darker areas with the same characteristics.

After detection of these structures, a second selection is carried out according to two criteria; the first one consists in removing the filaments having a large elongation – this is done by evaluating the condition number of their inertia matrix, the second selection will discard the quasi-steady state structures. At the pixel level, another selection is done: only points with a significant displacement are considered. Once the pixels have been selected, then an optimization procedure of the function J is performed and a field of velocity is obtained. The result will be considered as pseudo-observation and then included in a VDA scheme. The process is illustrated in Fig. 8.2. An inconvenience of this method is its large number of degrees of freedom: in the choice of the laws of conservation and also in the choice of threshold parameters for the selection of pixels.

8.2. Direct assimilation of images

In the former method, it is necessary to solve several problems of optimization, each iterative algorithm requires the choice of at least one stopping criterion, the accumulation of these quantities being detrimental to the control of the global algorithm. A way to alleviate this difficulty is to consider some characteristic features of the images (e.g., fronts) as objects and they will be assimilated as such in addition to the usual state variables.

In the cost function defining the VDA, an additional term will be added in the form:

$$J_2(X) = \int_0^T \int_\Omega \|DX - I\|^2 \mathrm{d}t\mathrm{d}\Omega.$$

I is the image and D an operator from the space of the state variable of the model toward the space of images. Therefore, the comparison between the images retrieved from the model and the observed images is carried out in the space of images. The questions are

- How to choose the space of images?
- What metric must be used in this space for obtaining an efficient and pertinent comparison?

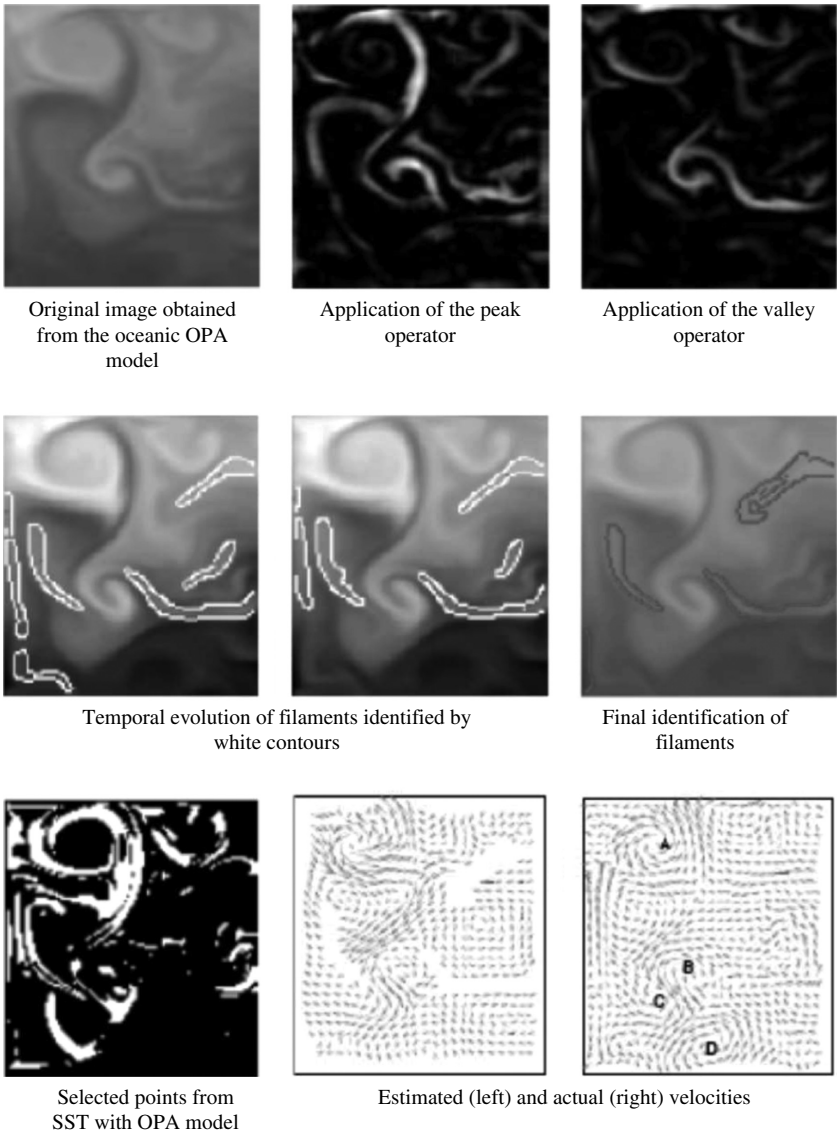
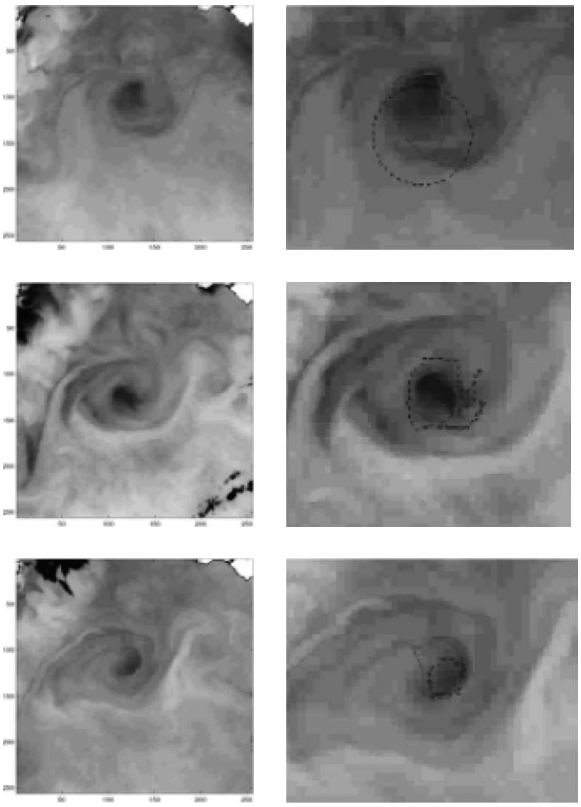


FIG. 8.2 Retrieving velocities from images.

In MA, ANTONIADIS and LE DIMET [2006], the choice has been done to use curvelets (CANDES, DEMANET, DONOHO and YING [2006]). The features of interest in the images are defined by contours: snakes. A snake is a virtual object which can be deformed elastically (thus possessing an internal energy) and which is immersed in a potential field. The main difficulties for applying a snake model to a temporal sequence of images consist in the determination of an initial contour and the design of external forces.



Evolution of a vortex in SST

Tracking in triplet successive frames using the curvelets based method. The initial and final snakes are displayed

FIG. 8.3 Direct assimilation of images.

The advantage of curvelet-based multiscale methods for snake modeling and tracking is its ability for simultaneously detecting edges and suppressing noise. An example of application of this method is given in Fig. 8.3.

Assimilating images is a generic problem with potential developments not only for geophysical fluids but also in biology and medicine. With respect to the classical data assimilation, due to the multiscale approach, it has the potential to focus on local features such as storms or hurricanes in meteorology. We can expect that further developments will be achieved in the following years.

9. Conclusion

Presently, data assimilation is a very active domain of research with extensions toward several directions.

- The domain of application of these methods has been extended to some other domains in geophysics especially in hydrology for the water cycle surface and underground water. Atmospheric chemistry is an important domain of potential applications.
- From the computational point of view, there is a demand for efficient and fast methods saving both storage and computing time.
- From the theoretical point of view, these methods are not always clearly justified, especially in the nonlinear case. Many problems remain open such as the optimal location of sensors.

Data assimilation has become an essential tool for modeling and prediction of the evolution of geophysical fluids. In many other domains for which data and models are the main sources of information, these methods could be developed in the near future.

10. Acknowledgments

The authors would like to thank Didier Auroux, William Castaings, Mohamed Jardak, and Diana Cheng for their help in the preparation of the manuscript. This paper was partly (assimilation of images) funded by Agence Nationale de la Recherche, France (ANR) in the framework of the ADDISA project. Prof. I. M. Navon would like to acknowledge the support of NSF grants ATM-0201808 and CCF-0635162.

References

Anderson, J.L. (2003). A local least squares framework for ensemble filtering. *Mon. Wea. Rev.* **131**, 634–642.

Anderson, J.L., Anderson, S.L. (1999). A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127**, 2741–2758.

Anderson, B.D.O., Moore, J.B. (1979). *Optimal Filtering* (Prentice-Hall, Englewood Cliffs, NJ).

Anthes, R.A. (1974). Data assimilation and initialization of hurricane prediction models. *J. Atmos. Sci.* **31**, 702–719.

Auroux, D. (2003). Étude de différentes méthodes d'assimilation de données pour l'environnement, Thesis, Université de Nice Sophia-Antipolis.

Auroux, D., Blum, J. (2005). Back and forth nudging algorithm for data assimilation problems. *C. R. Acad. Sci. Paris*. Sér. I **340** (12), 873–878.

Auroux, D., Blum, J. (2008). A nudging-based data assimilation method: the Back and Fourth Nudging (BFN) algorithm. *Nonlinear Processes in Geophysics*, **15**, 305–319.

Bengtsson, T., Snyder, C., Nychka, D. (2003). Toward a nonlinear ensemble filter for high dimensional systems. *J. Geophys. Res. Atmos.* **108** (D24).

Berkooz, G., Holmes, P., Lumley, J.L. (1993). The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.* **25**, 539–575.

Bishop, C.H., Etherton, B.J., Majumdar, S.J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Mon. Wea. Rev.* **129** (3), 420–436.

Blayo, E., Blum, J., Verron, J. (1998). Assimilation variationnelle de données en océanographie et réduction de la dimension de l'espace de contrôle. In: *Équations aux dérivées partielles et applications* (Articles dédiés 'a Jacques-Louis Lions, Gauthier-Villars), pp. 199–219.

Blum, J., Luong, B., Verron, J. (1998). Variational assimilation of altimeter data into a non-linear ocean model: temporal strategies, Contrôle et e.d.p. (Luminy, juin 1997). *ESAIM Proc.* **4**, 21–57.

Bouttier, F. (1994). A dynamical estimation of forecast error covariances in an assimilation system. *Mon. Wea. Rev.* **122**, 2376–2390.

Budgell, W.P. (1986). Nonlinear data assimilation for shallow water equations in branched channels. *J. Geophys. Res.* **91** (C9), 10633–10644.

Bui-Thanh, T., Willcox, K., Ghattas, O., Van Bloemen Waanders, B. (2007). Goal-oriented model-constrained optimization for reduction of large-scale systems. *J. Comput. Phys.* **224** (2), 880–896.

Burgers, G., Van Leeuwen, P.J., Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.* **126**, 1719–1724.

Burkardt, J., Gunzburger, M., Lee, H.C. (2006). POD and CVT-based reduced-order modeling of Navier-Stokes flows. *Comput. Methods Appl. Mech. Eng.* **196** (1–3), 337–355.

Candes, E., Demanet, D., Donoho, L., Ying, L. (2006). Fast discrete curvelets transforms. *Multiscale Model. Simul.* **5** (3), 861–899.

Cane, M.A., Kaplan, A., Miller, R.N., Tang, B., Hackert, E.C., Busalacchi, A.J. (1996). Mapping tropical Pacific sea level: data assimilation via a reduced state space Kalman filter. *J. Geophys. Res.* **101**, 22599–22617.

Cao, Y., Zhu, J., Luo, Z., Navon, IM. (2006). Reduced-order modeling of the upper tropical Pacific ocean model using proper orthogonal decomposition. *Comput. Math. Appl.* **52** (8–9), 1373–1386.

Cao, Y., Zhu, J., Navon, I.M., Luo, Z. (2007). A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *Int. J. Numer. Methods Fluids* **53** (10), 1571–1583.

10-Ch09-N51893 [13:43 2008/9/13] Temam & Tribbia: Computational Methods for the Atmosphere and the Oceans Page: 427 377–434

Carme, S., Pham, D.T., Verron, J. (2001). Improving the singular evolutive extended Kalman filter for strongly nonlinear models for use in ocean data assimilation. *Inverse Probl.* **17** (5), 1535–1559.

Chen, Y., Snyder, C. (2007). Assimilating vortex position with an ensemble Kalman filter. *Mon. Wea. Rev.* **135**, 1828–1845.

Cohn, S.E., Ghil, M., Isaacson, E. (1981). Applications of estimation theory to numerical weather prediction. *Bull. Am. Meteorol. Soc.* **62** (6), 899.

Cohn S.E. (1982). ~~Mehtods~~ of sequential estimation for determining initial data in numerical weather prediction. Thesis (Ph.D.)–New York University (Source: Dissertation Abstracts International, Vol. 43-07, Sec. B, p. 2227).

Cohn, S.E. (1997). An introduction to estimation theory. *J. Meteorol. Soc. Jpn.* **75** (1B), 257–288.

Cohn, S.E., Todling, R. (1996). Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteorol. Soc. Jpn.* **74** (1), 63–75.

Cohn, S.E., Parrish, D.E. (1991). The behavior of forecast error covariances for a kalman filter in two dimensions. *Mon. Wea. Rev.* **119** (8), 1757–1785.

Courtier, Ph., Talagrand, O. (1987). Variational assimilation of meteorological observations with the adjoint equations part 2. Numerical results. *Q. J. Roy. Meteorol. Soc.* **113**, 1329–1347.

Courtier, Ph., Thepaut, J-N., Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. Roy. Meteorol. Soc.* **120**, 1367–1388.

Daescu, D.N., Navon, I.M. (2008). Efficiency of a Pod-based reduced second order adjoint model in 4D-Var data assimilation. *Int. J. Numer. Methods Fluids.* **53**, 985–1004.

Daescu, D.N., Navon, I.M. (2008). A dual-weighted approach to order reduction in 4D-Var data assimilation. *Monthly Weather Review*. **136** (3), 1026–1041.

Daley, R. (1991). *Atmospheric Data Analysis*, Cambridge University Press, 471 ~~Pages~~.

Dowell, D.C., Zhang, F., Wicker, L.J., Snyder, C., Crook, N.A. (2004). Wind and thermodynamic retrievals in the 17 may 1981 Arcadia, Oklahoma supercell: ensemble Kalman filter experiments. *Mon. Wea. Rev.* **132**, 1982–2005.

Du Plessis, R.M. (1967). *Poor Man's Explanation of Kalman Filters or How I Stopped Worrying and Learned to Love Matrix Inversion*. (Rockwell International Technical Note, Anaheim, CA), pp. 24.

Durbiano, S. (2001). Vecteurs caractéristiques de modéles océaniques pour la réduction d'ordre en assimilation de données, Ph.D. Thesis, Université Joseph Fourier, Laboratoire de Modélisation et calcul. Grenoble, France, pp. 214.

Evensen, G. (1992). Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res. Ocean.* **97** (C11), 17905–17924.

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

Evensen, G. (2007). *Data Assimilation: The Ensemble Kalman Filter* (Springer, Berlin, Germany).

Evensen, G., Van Leeuwen, P.J. (1996). Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.* **124** (1), 85–96.

Evensen, G. (2004). Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dyn.* **54**, 539–560.

Evensen, G. (2003). The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53**, 343–367.

Fahl, M., Sachs, E. (2003). Reduced-order modeling approaches to pdeconstrained optimization based on proper orthogonal decomposition. In: Biegler, L.T., Ghattas, O., Heinkenschloss, M., Van Bloemen Waanders, B. (eds.), *Large Scale PDE-Constrained Optimization* (Springer Verlag), pp. 268–281, 350.

Fang, F., Pain, C., Navon, I.M., Piggott, M.D., Gorman, G.J., Goddard, A.J.H. (2008). Reduced-order modeling of an adaptive mesh ocean model. In Early view with *International Journal for Numerical Methods in fluids*.

Farrell, B.F., Ioannou, P.J. (2001). Estimation using a reduced-order Kalman filter. *J. Atmos. Sci.* **58** (23), 3666–3680.

Fisher, M. (1998). *Development of a Simplified Kalman Filter. ECMWF Research Department Technical Memorandum 260.* (European Centre for Medium-Range Weather Forecasts), pp. 16. Available from Library, ECMWF, Shinfield Park, Reading, Berkshire, RG29AX England.

Fukumori, I., Malanotte-Rizzoli, P. (1995). An approximate Kalman filter for ocean data assimilation - an example with an idealized gulf-stream model. *J. Geophys. Res. Oceans* **100** (C4), 6777–6793.

Gauthier, P., Courtier, P., Moll, P. (1993). Assimilation of simulated lidar data with a Kalman filter. *Mon. Wea. Rev.* **121**, 1803–1820.

Gejadze, I., Le Dimet, F.-X., Shutyaev, V. (2007). On error covariances in variational data assimilation. *Russ. J. Numer. Anal. Math. Model.* **22** (2), 163.

Gelb, A., Kasper, J.F., Nash, R.A., Price, C.F., Sutherland, A.A. (1974). *Applied Optimal Estimation* (The M.I.T. Press), pp. 374.

Ghil, M., Manalotte-Rizzoli, P. (1991). Data assimilation in meteorology and oceanography. *Adv. Geophys.* **23**, 141–265.

Ghil, M. (1989). Meteorological data assimilation for oceanographers. Part I: description and theoretical framework. *Dyn. Atmos. Oceans* **13**, 171–218.

Ghil, M., Cohn, S., Tavantzis, J., Bube, K., and Isaacson, E. 1981: Applications of estimation theory to numerical weather predication, Dynamic Meteorology: Data Assimilation Methods, L. Bengtsson, M. Ghil and E. Källén (eds.), Springer Verlag, pp. 139–224.

Gilbert, J.C., Lemarechal, C. (1989). Some numerical experiments with variable storage quasi-Newton algorithms. *Math. Prog.* **B25**, 407–435.

Golub, G.H., Van Loan, C.F. (1989). *Matrix Computations*, second ed. (Johns Hopkins Univ. Press).

Gunzburger, MD., Peterson, JS., Shadid, JN. (2007). Reduced-order modeling of time-dependent PDEs with multiple parameters in the boundary data. *Comput. Methods Appl. Mech. Eng.* **196** (4–6), 1030–1047.

Gunzburger, M.D. (2003). *Perspectives in Flow Control and Optimization* (Society for Industrial and Applied Mathematics, Philadelphia, PA), pp. 261.

Hager, W.W. (1989). Updating the inverse of a matrix. *SIAM Rev.* **31**, 221–239.

Hamill, T.M., Snyder, C. (2000). A hybrid ensemble Kalman filter/3dvariational analysis scheme. *Mon. Wea. Rev.* **128**, 2905–2919.

Hamill, T.M., Snyder, C. (2002). Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.* **130**, 1552–1572.

Hamill, T.M., Whitaker, J.S. (2005). Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Mon. Wea. Rev.* **133**, 3132–3147.

Hamill, T.M., Whitaker, J.S., Snyder, C. (2001). Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129**, 2776–2790.

Herlin, I., Le Dimet, F.-X., Huot, E., Berroir, J.-P. (2004). Coupling models and data: which possibilities for remotely-sensed images. In: Prastacos, P., Cortes, U., Diaz, J.L., Murillo, J. (eds.), *e-Environement: Progress and Challenge*, vol. 11 (Research on Computing Sciences, Mexico).

Hoke, J.E., Anthes, R.A. (1976). The initialization of numerical models by a dynamical initialization technique. *Mon. Wea. Rev.* **104**, 1551–1556.

Holland, W.R. (1978). The role of mesoscale eddies in the general circulation of the ocean-numerical experiments using a wind-driven quasigeostrophic model. *J. Phys. Oceanogr.* **8**, 363–392.

Holmes, P., Lumley, J.L., Berkooz, G. (1996). *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge University Press, Cambridge, UK).

Holmes, P.J. (1990). Nonlinear dynamics, chaos and mechanics. *Appl. Mech. Rev.* **43** (5–2), S23–S39.

Hoteit, I., Pham, D.-T. (2003). Evolution of the reduced state space and data assimilation schemes based on the Kalman filter. *J. Meteorol. Soc. Jpn.* **81** (1), 21–39.

Hoteit, I., Kohl, A. (2006). Efficiency of reduced-order, time-dependent adjoint data assimilation approaches. *J. Oceanogr.* **62** (4), 539–550.

Hoteit, I., Pham, D.T., Blum, J. (2002). A simplified reduced-order Kalman filtering and application to altimetric data assimilation in Tropical Pacific. *J. Mar. Syst.* **36** (1–2), 101–127.

Hoteit, I., Pham, D.T., Blum, J. (2001). A semi-evolutive partially local filter for data assimilation. *Mar. Pollut. Bull.* **43** (7–12), 164–174.

Houtekamer, P.L., Mitchell, H.L. (1998). Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811.

Houtekamer, P.L. Mitchell, H.L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon., Wea. Rev.* **129** (1), 123–137.

Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., Hansen, B. (2005). Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. *Mon. Wea. Rev.* **133** (3), 604–620.

Houtekamer, P.L., Mitchell, H.L. (2005). Ensemble Kalman filtering. *Q. J. Roy. Meteorol. Soc.* **131** (613), 3269–3289, part C.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417–441, 498–520.

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory* (Academic Press), pp. 376.

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

Kalman, R.E. (1961). New methods and results in linear filtering and prediction theory. *ASME J. Basic Eng. Ser. D.* **83**.

Kalnay, E. (2003). *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge University Press, Cambridge, UK).

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., Ballabrera-Poy, J. (2007). 4D-Var or ensemble Kalman filter? *Tellus A* **59** (5), 758–773.

Keppenne, C.L., Rienecker, M.M. (2002). Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.* **130**, 2951–2965.

Karhunen, K. (1946). Zur spektraltheorie stochastischer prozesse. *Ann. Acad. Sci. Fennicae* 37.

Kirby, M., Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Anal. Mach. Intell.* **12** (1), 103–108.

Kirby, M. (2001). *Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns* (John Wiley).

Kosambi, D.D. (1943). *J. Indian Math. Soc.* **7**, 76–88.

Krishnamurti, T.N., Jishan, X., Bedi, H.S., Ingles, K., Oosterhof, D. (1991). Physical initialization for numerical weather prediction over the tropics. *Tellus* **43A**, 53–81.

Lawless, A.S., Nichols, N.K., Boess, C., Bunse-Gerstner, A. (2006). *Using Model Reduction Methods with in Incremental 4D-Var, Numerical Analysis Report 10/06* (Department of Mathematics, The University of Reading).

Lawson, W.G., Hansen, J.A. (2004). Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Wea. Rev.*, **132**, 1966–1981.

Le Dimet, F.-X. (1980). *Une Étude Générale D'analyse Objective Variationnelle Des Champs Mét éorologiques* (Rapport Scientifique LAMP 28, Université de Clermond II, BP 45, 63170 Aubi'ere, France).

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

Le Dimet, F.-X., Ngodock, H.-E., Luong, B., Verron, J. (1997). Sensitivity analysis in variational data assimilation. *J. Met. Soc. Jpn.* **75** (1B), 245–255.

Le Dimet, F.-X., Navon, I.M., Daescu, D.N. (2002). Second order information in data assimilation. *Mon. Wea. Rev.* **130** (3), 629–648.

Le Dimet, F.-X., Shutyaev, V., Gejadze, I. (2006). On optimal solution error in variational data assimilation: theoretical aspects. *Russ. J. Numer. Anal. Math. Model.* **21** (2), 139.

Lions, J.L. (1971). *Optimal Control of Systems Governed by Partial Differential Equations* (Springer Verlag).

Loeve, M. (1945). Fonctions aléatoires de second ordre. *Compte Rend. Acad. Sci. Paris* **220**.

Lermusiaux, P.F.J., Robinson, A.R. (1999). Data assimilation via error subspace statistical estimation. 1: theory and schemes. *Mon. Wea. Rev.* **127**, 1385–1407.

Li, Z., Navon, I.M. (2001). Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Q. J. Roy. Meteorol. Soc.* **127**, 661–683.

Lorenc, A.C. (1986). Analysis methods for numerical weather prediction. *Q. J. Roy. Meteorol. Soc.* **112**, 1177–1194.

Lorenc, A.C. (1988). A practical approximation to optimal four-dimensional objective analysis. *Mon. Wea. Rev.* **116**, 730–745.

Lumley, J.L. (1967). The structure of inhomogeneous turbulence. In: Yaglom, A.M., Tatarski, V.I. (eds.), *Atmospheric Turbulence and Wave Propagation* (Nauka, Moscow), pp. 166–178.

Luo, Z., Chen, J., Zhu, J., Wang, R., Navon, I.M. (2007). An optimizing reduced-order FDS for the tropical Pacific Ocean reduced gravity model. *Int. J. Numer. Methods Fluids* **55** (2), 143–161.

Luong, B., Blum, J., Verron, J. (1998). A variational method for the resolution of a data assimilation problem in oceanography. *Inverse Probl.* **14**, 979–997.

Lyne, W.H. (1979). *Data Assimilation by Repeated Correction of Model Fields-Theoretical Considerations* (Met O 11 Tech. Note No. 130, Meteorological Office).

Lyne, W.H., Swinbank, R., Birch, N.T. (1982). A data assimilation experiment and the global circulation during the FGGE special observing periods. *Q. J. R. Met. Soc.* **108**, 575–594.

Ma, J., Antoniadis, A., Le Dimet, F.-X. (2006) Curvelet-based snake for multiscale detection and tracking of geophysical fluids. *IEEE Trans. Geosci. Remote Sens.* **14** (12), 3626–3638.

Mandel, J. (2007). *A Brief Tutorial on the Ensemble Kalman Filter* (CCM Report 242, University of Colorado at Denver and Health Sciences Center).

Mandel, J. (2006). *Efficient Implementation of the Ensemble Kalman Filter* (CCM Report 231, University of Colorado at Denver and Health Sciences Center).

Mandel, J., Beezley, J.D. (2007). *Predictor-Corrector and Morphing Ensemble Filters for the Assimilation of Sparse Data into High-Dimensional Nonlinear Systems* CCM Report 239, 11th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS), CD-ROM, Paper 4.12, 87th American Meteorological Society Annual Meeting, San Antonio, TX, January 2007 presentation, November 2006.

Maybeck, P.S. (1979). *Stochastic Models, Estimation, and Control, vol. 1* (Academic Press).

Meyer, M., Matthies, HG. (2003). Efficient model reduction in non-linear dynamics using the Karhunen-Loeve expansion and dual-weighted-residual methods. *Comput. Mech.* **31**, 179–191.

Miller, R.N., Ghil, M., Gauthiez, F. (1994). Advanced data assimilation in strongly nonlinear dynamical-systems. *J Atmos. Sci.* **51** (8), 1037–1056.

Miller, R.N. (1986). Toward the application of the Kalman filter to regional open ocean modeling. *J. Phys. Oceanogr.* **16** (1), 72–86.

Mohammadi, B., Pironneau, O. (2001). *Applied Shape Optimization for Fluids* (Clarendon Press, Oxford, UK).

Navon, I.M., Zou, X., Derber, J., Sela, J. (1992). Variational data assimilation with an adiabatic version of the NMC spectral model. *Mon. Wea. Rev.* **120** (7), 1433–1446.

Navon, I.M. (1998). Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dyn. Atmos. Ocean.* **27** (1–4), 55–79, Special Issue in honor of Richard Pfeffer.

Nerger, L., Hiller, W., Schrater, J. (2005). A comparison of error subspace Kalman filters. *Tellus A* **57** (5), 715–735.

Nodet, M. (2006). Variational assimilation of Lagrangian data in oceanography. *Inverse Probl.* **22**, 245–263.

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, E., Kalnay, E., Patil, D., Yorke, J.A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A* **56**, 415–428.

Palmer, T., Hagedorn, R. (2006). *Predictability of Weather and Climate* (Cambridge University Press), pp. 718.

Parmuzin, E., Le Dimet, F.-X., Shutyaev, V. (2006). On error analysis in variational data assimilation problem for a nonlinear convection-diffusion model. *Russ. J. Numer. Anal. Math. Model.* **21** (2), 169.

Parrish, D., Derber, J. (1992). The national meteorological centers spectral statistical interpolation analysis system. *Mon. Wea. Rev.* **120**, 1747–1763.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **2**, 559–572.

Peterson, J.S. (1989). The reduced basis method for incompressible viscous flow calculations. *SIAM J. Sci. Stat. Comput.* **10** (4), 777–786.

Pham, DT., Verron, J., Roubaud, MC. (1998). A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.* **16** (3–4), 323–340.

Puel, J.P. (2002). A non-standard approach to a data assimilation problem. *C. R. Acad. Sci.* Paris, Ser.1 **335**, 161–166.

Rabier, F. (2005). Overview of global data assimilation developments in numerical weather-prediction centres. *Q. J. Roy. Meteorol. Soc.* **131** (613), 3215–3233, part C.

Ramamurthy, M.K., Carr, F.H. (1987). Four-dimensional data assimilation in the monsoon region. Part I: experiments with wind data. *Mon. Wea. Rev.* **115**, 1678–1706.

Ramamurthy, M.K., Carr, F.H. (1988). Four-dimensional data assimilation in the monsoon region. Part II: role of temperature and moisture data. *Mon. Wea. Rev.* **116**, 1896–1913.

Ravindran, S.S. (2002). Adaptive reduced-order controllers for a thermal flow system using proper orthogonal decomposition. *SIAM J. Sci. Comput.* **23** (6), 1924–1942.

Ravindran, S.S. (2006). Reduced-order controllers for control of flow past an airfoil. *Int. J. Numer. Methods Fluids* **50** (5), 531–554.

Ravindran, S.S. (2000). A reduced-order approach for optimal control of fluids using proper orthogonal decomposition. *Int. J Numer. Methods Fluids* **34** (5), 425–448.

Reichle, R.H., McLaughlin, D.B., Entekhabi, D. (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.* **130**, 103–114.

Reichle, R.H., Walker, J.P., Koster, R.D., Houser, P.R. (2003). Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometeorol.* **3**, 728–740.

Robert, C., Durbiano, S., Blayo, E., Verron, J., Blum, J., Le Dimet, F.X. (2005). A reduced-order strategy for 4D-Var data assimilation. *J. Mar. Syst.* **57** (1–2), 70–82.

Robert, C., Blayo, E., Verron, J. (2006). Reduced-order 4D-Var: a preconditioner for the incremental 4D-Var data assimilation method. *Geophys. Res. Let.* **33**, L18609, doi: 10.1029/2006GL026555.

Robert, C., Blayo, E., Verron, J. (2006). Comparison of reduced-order, sequential and variational data assimilation methods in the tropical Pacific Ocean. *Ocean Dyn.* **56**, 624–633, doi: 10.1007/s10236-006-0079-9.

Rostaing-Schmidt, N., Hassold, E. (1994). Basic function representation of programs for automatic differentiation in the odyssée system. In: Le Dimet, F.-X. (ed.), *High Performance Computing in the Geosciences* (Kluwer Academic Publishers B.V. NATO ASI Series), pp. 207–222.

Sachs, E.W. (2003). Managing POD models in PDE-constrained optimization. Optimization in simulation-based models IMA, ~~m~~Minneapolis. Available at 2003; http://www.ima.umn.edu/talks/workshops/1-9-16.2003/sachs/sachs.pdf.

Sasaki, Y.K. (1958). An objective analysis based on the variational method. *J. Met. Soc. Jpn. II* **36**, 77–88.

Sirovich, L.P. (1987). Turbulence and the dynamics of coherent structures, Part I: coherent structures. *Q. Appl. Math.* **45**, 561–571.

Sirovich, L.P. (1987). Turbulence and the dynamics of coherent structures, Part II: symmetries and transformations. *Q. Appl. Math.* **45**, 573–582.

Sirovich, L.P. (1987). Turbulence and the dynamics of coherent structures, Part III: dynamics and scaling. *Q. Appl. Math.* **45**, 583–590.

Snyder, C., Zhang, F. (2003). Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.* **131**, 1663–1677.

Stauffer, D.R., Seaman, N.L. (1990). Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: experiments with synoptic-scale data. *Mon. Wea. Rev.* 118, 1250–1277.

Szunyogh, I., Kostelich, E.J., Gyarmati, G. et al. (2005). Assessing a local ensemble Kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus Series a-dynamic meteorology and oceanography.* **57** (4), ~~p.~~ 528–545.

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

Thepaut, J.N., Courtier, P. (1987). Four-dimensional data assimilation using the adjoint of a multi-level primitive equation model. *Q. J. Roy. Meteorol. Soc.* **117**, 1225–1254.

Thomas M. Hamill. (2006). Ensemble-based atmospheric data assimilation. In: Tim Palmer and Renate Hagedorn (eds.), Predictability of Weather and Climate, Chapter 6, Cambridge University Press, pp. 124–156.

Tikhonov, A., Arsenin, V. (1977). *Solutions of Ill-posed Problems* (Winston, Washington, DC).

Tippett, M.K., Anderson, J.L., Bishop, C.H., Hamill, T.M., Whitaker, J.S. (2003). Ensemble square root filters. *Mon. Wea. Rev.* **131**, 1485–1490.

Todling, R., Cohn, S.E. (1994). Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.* **122**, 2530–2557.

Todling, R., Cohn, S.E., Sivakumaran, N.S. (1998). Suboptimal schemes for retrospective data assimilation based on the fixed-lag Kalman smoother. *Mon. Wea. Rev.* **126** (8), 2274–2286.

Todling, R. (1999). *Estimation Theory and Foundations of Atmospheric Data Assimilation* (DAO Office Note 1999-01).

Van Leeuwen, P. (2003). A variance-minimizing filter for large-scale applications. *Mon. Wea. Rev.* **131**, 2071–2084.

Verlaan, M., Heemink, A.W. (1995). Reduced rank square root filters for large scale data assimilation problems. *Proc. Second Int. Symp. On Assimilation of Observations in Meteorology and Oceanography* (WMO TD/-No.651, Tokyo, Japan), pp. 247–252.

Verlaan, M., Heemink, A.W. (1997). Tidal flow forecasting using reduced rank square root filters. *Stoch. Hydrol. Hydraul.* **11** (5), 349–368.

Verron, J., Molines, J.M., Blayo, E. (1992). Assimilation of Geosat data into a quasigeostrophic model of the North Atlantic between 20 and 50 N: preliminary results. *Oceanol. Acta* **15** (5), 575–583.

Vidard, P.A. (2001). Vers une prise en compte des erreurs modéle en assimilation de données 4D variationnelle. Application á un modéle réaliste d'océan. Thesis (Université Joseph-Fourier, France).

Vidard, P.A., Piacentini, A., Le Dimet, F.-X. (2003). Determination of optimal nudging coefficients. *Tellus A* **55** (1), 1–15.

Wang, Z., Navon, I.M., Le Dimet, F.-X., Zou, X. (1992). The second order adjoint analysis: theory and application. *Met. Atmos. Phys.* **50**, 3–20.

Wang, W., Warner, T.T. (1988). Use of four-dimensional data assimilation by Newtonian relaxation and latent heat forcing to improve a mesoscale-model precipitation forecast: a case study. *Mon. Wea. Rev.* **116**, 2593–2613.

Whitaker, J.S., Hamill, T.M. (2002). Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.* **130**, 1913–1924.

Whitaker, J.S., Compo, G.P., Wei, X., Hamill, T.M. (2004). Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.* **132**, 1190–1200.

Whitaker, J.S. (2003). Ensemble square root filters. *Mon. Wea. Rev.* **131**, 1485–1490.

Willcox, K., Ghattas, O., Van Bloemen Waanders, B., Bader, B. (2005). An optimization framework for goal-oriented, model-based reduction of large-scale systems. Paper presented at 44th IEEE Conference on Decision and Control and European Control Conference, Seville, 2005.

Zhang, F., Snyder, C., Sun, J. (2004). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.* **132**, 1238–1253.

Zou, X., Navon, I.M., Le Dimet, F.-X. (1992). An optimal nudging data assimilation scheme using parameter estimation. *Q. J. Roy. Met. Soc.* **118**, 1163–1186.

Zou, X., Navon, I.M., Berger, M., Phua, K.H., Schlick,T., Le Dimet, F.-X. (1993). Numerical experience with limited Memory Quasi-Newton and Truncated Newton methods. *SIAM J. Optim.* **3** (3), 582–608.