ACTA MATHEMATICA VIETNAMICA 37
Volume 34, Number 1, 2009, pp. 37–47

MODELS, DATA AND IMAGES FOR PREDICTING THE EVOLUTION OF GEOPHYSICAL FLUIDS

FRANCOIS-XAVIER LE DIMET, INNOCENT SOUOPGUI, OLIVIER TITAUD AND ARTHUR VIDARD

Dedicated to Nguyen Van Hien on the occasion of his sixty-fifth birthday

Abstract. Predicting the evolution of the environment is an important and difficult task. To achieve this goal we need to consider all the available information: mathematical information under the form of models, observations, statistics, images from satellites. We will see how to use optimal control methods in order to link together these sources of information for retrieving at best the state of the environment. A particular attention will be paid to the assimilation of dynamic images in numerical models.

1. Introduction

The prediction of the atmosphere, the oceans or continental waters is an important social issue. Basically the forecast is obtained after a numerical integration of the mathematical models representing the geophysical fluid. Therefore a key problem is to retrieve the state of the fluid at an initial time corresponding to the starting date of the integration of the model.

The equations governing geophysical fluids are non linear, this fact has some important consequences on the behavior of the flows:

- Any situation is unique. The equations modeling a geophysical fluid do not have steady state or periodic solution. A consequence is that they have no general properties but only properties in the vicinity of some situation which have to be precised by data.
- There are interactions between the various scales of the flow and because the numerical models are necessarily finite, the flux of energy and matter between the resolved and unresolved scales has to be parameterized introducing some parameters which are not accessible to direct measurements.

Received April 3, 2008; in revised form September 18, 2008.
2000 Mathematics Subject Classification. 86A22.
Key words and phrases. Geophysical Fluids, Data Assimilation, Optimal Control, Prediction, Images.
This work, ADDISA project, is still under developments, it receives the financial support of Agence Nationale de la Recherche (ANR).

Briefly said: a model is not sufficient to retrieve the state of a geophysical flow at a given date. Some additional information must be added:

- data from in situ or remote observations,
- statistics,
- images,
- qualitative information.

Data Assimilation methods are the methods able to retrieve the state of a flow by mixing these various sources of information. In a first approximation there are two basic techniques :

- variational methods based on optimal control techniques,
- stochastic methods based on Kalman filter.

2. Variational data assimilation

2.1. **Principle of VDA.** Let us briefly describe the basic principle of Variational Data Assimilation (VDA). The state of the flow is described by a state variable X depending on time and space. It represents the variables of the model (velocity, temperature, elevation of the free surface, salinity, concentrations in biological or chemical species, ...). After discretisation, the evolution of the flow is governed by the differential system

$$\frac{dX}{dt} = F(X, U)$$

$$X(0) = V.$$

U contains some unknown parameters of the model: boundary condition, model error, parametrization of subgrid effect. U may depend on space and time. The initial condition V is unknown and depends on space. We assume that U and V being given, the model has a unique solution between 0 and T. We assume that we know an observation of the fields between 0 and T. To make things simple, we suppose its continuity in time. The discrepancy between the observation and the state variable is defined by a so-called cost-function in the form

$$J(U, V) = \frac{1}{2} \int_0^T \|CX - X_{obs}\|^2 \, dt + \|U - U_0\|^2 + \|V - V_0\|^2.$$

C is a mapping from the space of the state variable toward the space of observations where the comparison is carried out. The second and the third terms are regularization terms in Tikhonov's sense. They also permit to introduce some a priori information. It is important to point out that the norms are on three different spaces. They can take into account the statistical information by introducing the error covariance matrix. Here we will only consider identities in the three spaces. The problem of VDA can be considered as the determination of U^* and V^* minimizing $J(U, V)$. As a first approximation (humidity, salinity, concentration are non-negative), we have to solve an unconstrained optimization problem. From the numerical point of view, U^* and V^* will be estimated by a descent type algorithm i.e., as the limit of a sequence in the form

$$\begin{pmatrix} U_{k+1} \\ V_{k+1} \end{pmatrix} = \begin{pmatrix} U_k \\ V_k \end{pmatrix} + \lambda_k D_k,$$

where D_k is the direction of descent deduced from the gradient of J and λ_k is the stepsize realizing the minimum of J along the direction of descent. For computing the gradient we introduce a so-called adjoint variable P as the solution of the adjoint model

$$\begin{aligned} &\frac{dP}{dt} + \left[\frac{\partial F}{\partial X}\right]^T P = C^T\left(CX - X_{obs}\right) \\ &P(T) = 0. \end{aligned}$$

After a backward integration of the adjoint model, the gradient of J is given by

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

The derivation of the system can be found in [1]. The model plus the adjoint model with the optimality condition $\nabla J = 0$ is the Optimality System (O.S.). Let us point out that the optimality system contains all the available information and therefore sensitivity studies with respect to the observations must be carried out from the O.S. rather than from the model.

2.2. **Deriving the adjoint model.** The adjoint variable has the same dimension as the direct model leading to a large increase in the size of the code. Another inconvenient is that, in the case of nonlinear model, the solution of the direct system must be stored.

From the direct model toward the adjoint model, there are two basic operations:

- computing the jacobian of the model with respect to the state variable. This operation is not very difficult. The code can be derived statement by statement,
- transposing the jacobian. This is the most difficult operation because of multiple dependencies.

For preexisting codes the derivation of an adjoint model is an expensive operation in term of (wo)manpower. Fortunately these last years have known the development of software on automatic differentiation such as TAPENADE [2].

2.3. **Application in meteorology.** At the present time, most of the important meteorological centers have adopted VDA (European Center for Medium Range Weather Forecast, Japan Meteorological Agency, Météo France, . . .).

Models are based on the equation of fluid dynamics plus thermodynamics. After discretization in space (finite difference and spectral methods), the model is represented by a system of ODE's with around 500 millions of variables. This size will increase in the future.

Data are from several sources. The following figures display some sources of information (source ECMWF). They are heterogeneous in density, quality and

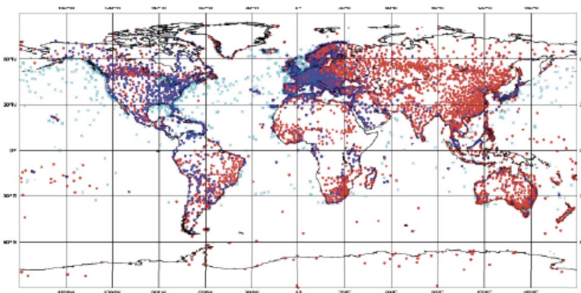

FIGURE 1. Synoptic data and ships

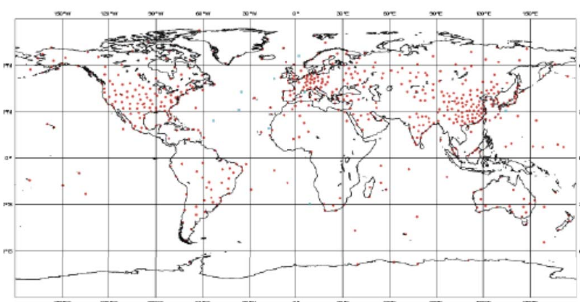

FIGURE 2. Wind, temperature, specific humidity

nature and therefore the covariance matrix associated to the error of observation has to be carefully estimated.

The observation of Earth by Satellites also provide many data, for instance there are polar orbiting satellites and geostationary satellites displayed in Fig. 3 and Fig. 4, respectively.

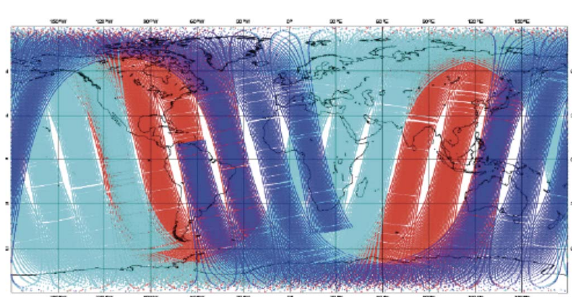

FIGURE 3. Polar orbiting satellites

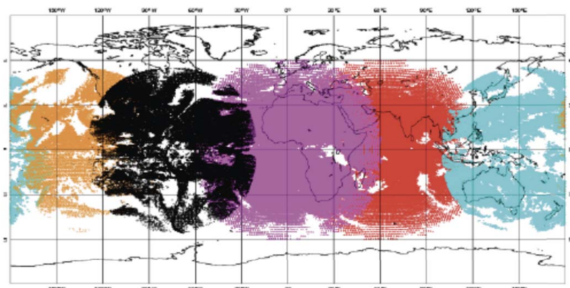

FIGURE 4. Geostationary satellites

These instruments do not observe directly the state variable of the model but radiances. Therefore, to be useful in the data assimilation, according to the general scheme defined above, an appropriated C operator transforming the observation into a state variable (e.g., radiances into a vertical profile of temperature) has to be provided. In this case it won't be a linear operator. Nevertheless the adjoint of this operator will be necessary.

Fig. 5 displays the increase in the total number of observations during the last decade including conventional and satellite observations. The total number has known an exponential increase. Nevertheless it remains smaller than the number of unknowns of the VDA problem. Consequently we have to deal with a ill-posed inverse problem and the role of the regularization term will be of importance.

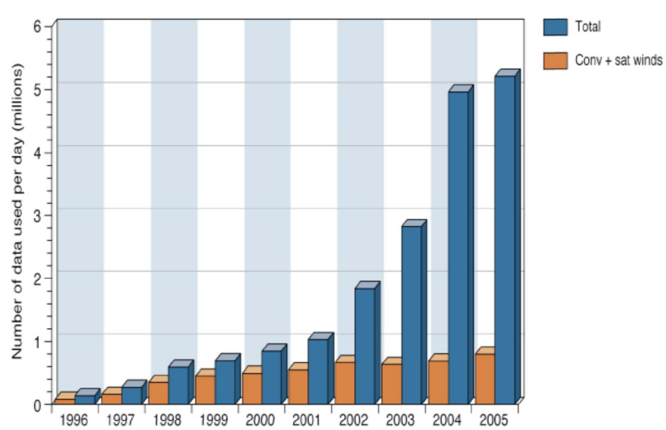

FIGURE 5. Total number of observations per days (source ECMWF)

3. Assimilation of images

3.1. **Dynamics of images.** The observation of the Earth by satellite is a source of information on the ocean and the atmosphere thanks to the measures of radiances. As it was said above, this data can be plugged in a data assimilation scheme. Moreover there is an important source of information captured by satellites: the dynamics of some meteorological or oceanic "objects": fronts, clouds, vortexes. Up to now this information has been used in a qualitative way rather than in a quantitative one. We present some preliminary works toward the assimilation of images into numerical models. Two preliminary questions arise:

- **What is an image?** Basically an image is composed of pixels to which is associated a scalar (grey level) field for black and white images (Fig. 6) and three scalars for colored images (Fig. 7), therefore to 3D vector field. From the mathematical point of view an image can be considered as an eulerian field.

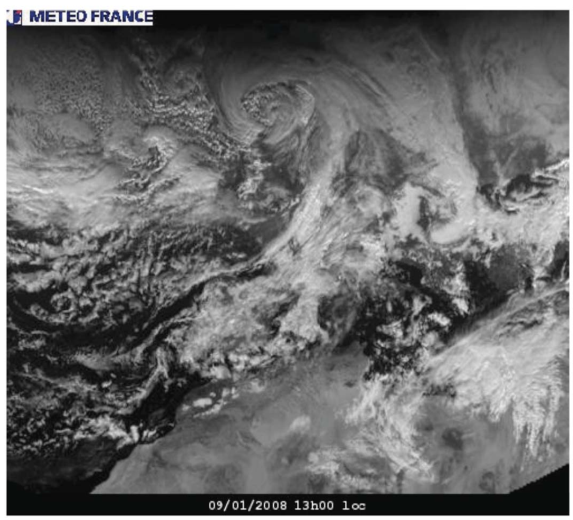

Figure 6. Europe on 09/01/08 at 13:00 on visible channel (source MétéoFrance)

- **What is seen?** Clearly on Fig. 6 and Fig. 7, the eye can distinguish shapes. If they are similar, they are not identical. Fig. 6 shows the nebulosity: this is a complex function of the state variables of the model: clouds are not directly in the solution of the model because they depend both on thermodynamics (temperature and humidity) and also on the microphysics of the cloud (water, ice, snow and the size of these particles).

On Figs. 7–9 we can see the evolution of a system on the Atlantic Ocean. These images provide an information on the dynamics of the system. Is it possible to quantify this information in order to combine it with numerical models ? There are two classes of methods.

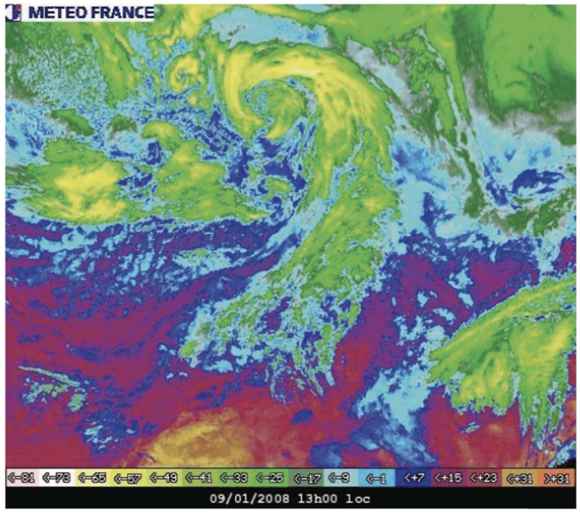

Figure 7. Europe on 09/01/08 at 13:00 on visible channel (source MétéoFrance)

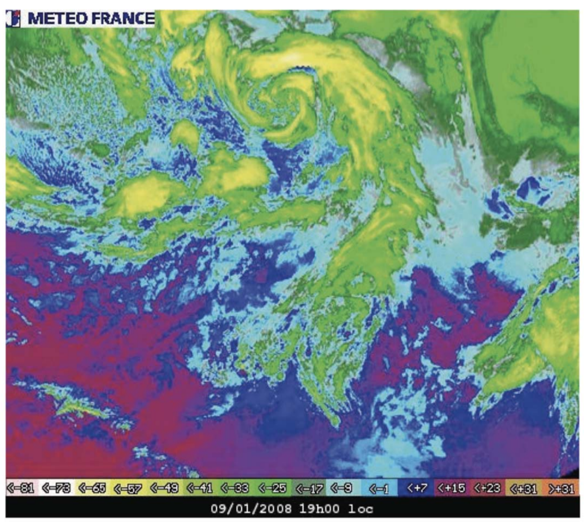

Figure 8. Europe on 09/01/08 at 19:00 on infrared channel (source MétéoFrance)

3.2. **Retrieving "pseudo-observations".** The principle of this approach [3] is to evaluate from the images some values of the velocity field and to use them in a regular scheme of VDA. This approach is a classical one in computer vision. It is based on the conservation of grey level values for individual pixels. Given a pixel of coordinates, I is the luminance of the pixel, this quantity is supposed to be conservative and this property is mathematically written as the total derivative of I is equal to 0:

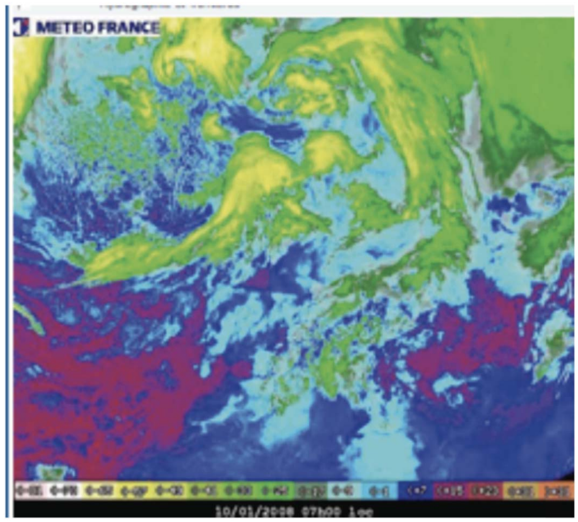

FIGURE 9. Europe on 10/01/08 at 07:00 on infrared channel (source MétéoFrance)

$$\frac{dI}{dt}\,(x(t),\,y(t),\,t) = 0,$$

leading to

$$\frac{\partial I}{\partial x}\,\frac{dx}{dt} + \frac{\partial I}{\partial y}\,\frac{dy}{dt} + \frac{\partial I}{\partial t} = \frac{\partial I}{\partial x}\,u + \frac{\partial I}{\partial y}\,v + \frac{\partial I}{\partial t} = 0,$$

where 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. Nevertheless it can be included in a variational formulation. We will look for the velocity field minimizing J with

$$\begin{aligned}
&J(u,v) = E_1 + E_2,\\
&E_1 = \textstyle\int_I \Big(\frac{\partial I}{\partial x}\,u + \frac{\partial I}{\partial y}\,v + \frac{\partial I}{\partial t}\Big)^2\,d\Omega,\\
&E_2 = \textstyle\int_I \|\nabla W\|^2\,d\Omega,\\
&W = (u,v).
\end{aligned}$$

E_2 can be considered as a regularization term to smooth the computed fields.

In E_1 we recognize the scalar product of the gradient of luminance with the velocity field. A consequence is that in locations where these vectors are orthogonal the equation doesn't bring any quantitative information on the velocity field.

This is not the only law of conservation which can be considered according to the nature of the image:

- with an image of the color of the ocean, an equation of conservation of chlorophyll (with source and sink) terms can be considered.
- with an image of Sea Surface Temperature (SST), the Boussinesq approximation can be used.

A selection of points on which these lagrangian equations are applied should be done. Must be discarded from the analysis structures the points for which the velocities and the gradient of luminance are almost orthogonal and also the quasi steady state structures. This is the case of filaments, which are elongated structures: they are detected by operators of mathematical morphology: peak operator detecting brighter areas of maximum width and valley operator detecting the darkest areas with the same characteristics.

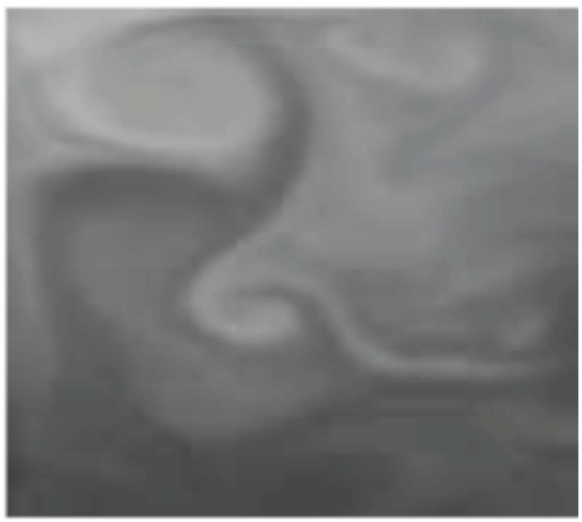

FIGURE 10. Image issued from the OPA model

3.3. **Direct assimilation of images.** Another way [4] to proceed is a direct assimilation of images using a variational formalism. We will avoid the stage of generating pseudo physical observations. The principle is to add in the discrepancy between the solution of the model and the observation a term directly linked to the images and their dynamics. It will take the form

$$J_2(X) = \int_0^T \int_\Omega \|DX - I\|_F^2 \, dt \, d\Omega.$$

In this expression F is the space of images, D is a mapping from the space of the state variable toward the space of images. Therefore the question is: how to define F? The imposed constraints are

- a structure of normed space should be considered in order to use simple rules for the differentiation of J_2 .

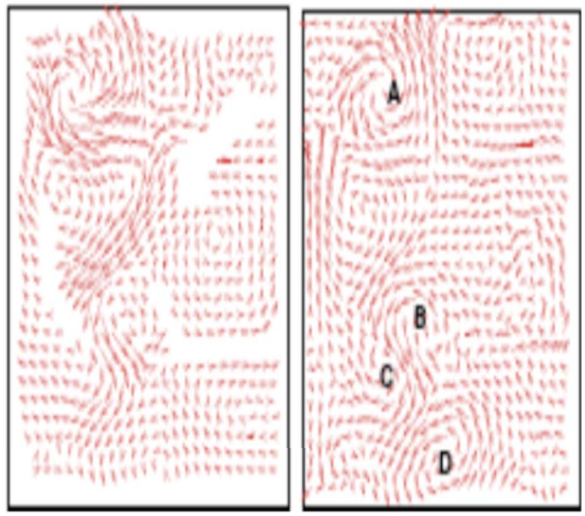

FIGURE 11. Estimated (left) and actual fields

- the dimensionality of F must remain reasonable to store the evolution of the images.

D is a mapping from the space of the state variable toward the space of images. It will be nonlinear and with a complex structure. With the usual variational approach, a term of the form

$$\left[\frac{\partial D}{\partial X}\right]^T (DX - I)$$

will be added in the right hand side of the adjoint model, where D is the process permitting to retrieve an image from the computed fields.

How to mathematically define images ? The first idea would be to consider an image characterized by edges and these edges will be approximated in a functional space. Wavelets is the first idea for such a discretization. It is to carry out a wavelet analysis of the contours detected in images.

4. CONCLUSION

Predicting the evolution of the environment requires to link together heterogeneous sources of information. There is a strong demand for the improvement of prediction and VDA is a powerful tool to achieve this goal. The insertion of images into numerical models is a major challenge with applications in many scientific and technological fields. Some years ago the development of complex prediction methods was unreachable for many developping countries because of their high computational cost, but we can expect that in a near future this will be no longer true and that reliable prediction on typhoons, floods will be carried out locally. A prerequisite is the formation of scientists in these fields.

References

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

[2] L. Hascoet, V. Pascual and R.-M. Greboroe, *The TAPENADE AD Tool*, TROPIC Project, INRIA Sophia Antipolis, AD Workshop, Cranfield, June 5-6, 2003.

[3] I. Herlin, F.-X. Le Dimet, E. Huot and J. P. Berroir, Coupling models and data: which possibilities for remotely-sensed images, in *e-Environment: Progress and Challenge*, Prastacos, Cortès, Diaz and Murillo (Eds), Research on Computing Sciences, Mexico, **11**, 2004.

[4] J. Ma, A. Antoniadis and F.-X. Le Dimet, Geometric wavelet-based snakes for multiscale detection and tracking of geophysical fluids, *IEEE Geosc. Rem. Sens.* **44** (12) (2006).

Université de Grenoble,
Laboratoire Jean-Kuntzmann and Inria,
38041 Grenoble, France

E-mail address: `ledimet@imag.fr`

E-mail address: `Innocent.Souopgui@imag.fr`

E-mail address: `Olivier.Titaud@hmg.inpg.fr`

E-mail address: `Arthur.Vidard@imag.fr`