Adjoint formalism and the configuration retrieval of an evolutionary system represented by Burgers' Equations

P. S. D. Silva* I. M. Navon L. Landau
COPPE/UFRJ CSIT/FSU COPPE/UFRJ

August 6, 2002

Abstract

4D variational data assimilation is proposed as an efficient procedure to determine a past configuration of a dynamical system of which only a few (and, under certain hypothesis, even partial) observations and its dynamics are known. The retrieved configuration is in good accordance with the test configuration.

The leading mathematical concept, adjoint equations, and the role in the design of algorithms of the representation of the differential of a functional defined in a vectorial space with inner product are discussed in detail.

Keywords: Optimal control, adjoint equations, variational methods, data assimilation.

AMS 2000: 49J20, 35A99, 35A15.

1 Introduction

Data assimilation consists in combining a set of space-time distributed observations with the numerical model of the dynamics of a given system in order to produce a configuration of the system that reconciles model outputs with observations. In the variational approach, it is defined a linear functional evaluating the discrepancy between available observations of the system and its numerical model produced configurations. Thus, an initial configuration of the system is sought that minimizes the functional. Being a proper way to change from the control over state variables to that over input variables (initial or/and boundary data), the method of adjoint equations gives a computationally efficient strategy to this task.

Due to its importance, a rigorous and hopefully clear discussion of the method of adjoint equations is presented in section 2. In the exposition, we emphasize the geometrical aspects involved in changing from state variables to input variables, as well as a particular

*paulosd@iis.com.br

presentation of the differential of a given functional, that uses the gradient of the functional, when it is defined in a vectorial space with an inner product. The aforementioned change of variables is a very appropriate procedure to reduce the computational effort to the resolution of the variational data assimilation problem.

Knowing the differential of the functional that measures the error between the observations of the system and its model generated configurations, it is possible, as in the two cases presented here, to perform the sensitivity analysis of the model to its inputs. Although a more general sensitivity analysis requires also a second order adjoint analysis ([1] and [2]), as in the interesting situation of assessing the effect of changes in observation sets ([3] and [4]), which can lead to a rational solution to the sensor location problem.

In this work, we examine the capacity of the variational data assimilation method to retrieve a past configuration of a given system by means of a numerical model of its dynamics and a set of system configurations. The studied system consists of a scalar field being transported in a two-dimensional domain by a known medium. Two cases are analyzed: in one of them, each observation produces the entire information about the scalar field, at the time the observation is obtained; in the other case, at each instant of time, only partial information on the scalar field is available, but in such a way that, if the model were integrated forward starting each time from the instant of time at which each observation is obtaining, there would be an instant of time such that all model outputs together generate a complete knowledge of the system, that is, the system, in this second case, is observable.

The procedure outlined here would correspond to the first step to remedy an oil spill on water.

2 Adjoint formalism

The method of adjoint equations has been successfully used in numerical solution of control of systems problems. Although there is a considerable amount of work dealing with adjoint formalism (a very good one is [5]), we present in this section a rigorous approach to the method as well as a brief discussion of its mathematical meaning with applications in view.

Let E and V be vectorial spaces with inner products $<,>_E$ and $<,>_V$, respectively, being

$$\text{E}= \{X : [t_0, t_1] \to V\}.$$

The linear version of the variational data assimilation problem is formally given by the following problem (P):

Find the solution Y ∈ E of

$$(P) \quad \begin{cases} \dfrac{\partial Y}{\partial t} = \dfrac{\partial F}{\partial X} Y \\ \text{that minimizes the functional} \\ \text{J:E} \rightarrow \Re \\ \quad \text{Y} \mapsto \int_{t_0}^{t_1} T(Y(t))\, dt \end{cases}$$

being F: E $\rightarrow$ E a differentiable operator,

$$\begin{aligned} &\text{T:E} \rightarrow \Re \\ &X \mapsto \frac{1}{2} \ \langle W(t)(X(t) - \tilde{X}(t)), X(t) - \tilde{X}(t) \rangle_E \end{aligned}$$

and W $\in$ E and $\tilde{\text{X}}$ $\in$ E given.

Define the operators

$$\begin{aligned} N : \text{E} &\rightarrow \text{E} \qquad \text{and} \qquad & N^* : \text{E} &\rightarrow \text{E} \\ Y &\mapsto \frac{\partial Y}{\partial t} - \frac{\partial F}{\partial X} Y & Y &\mapsto - \frac{\partial Y}{\partial t} - (\frac{\partial F}{\partial X})^* Y. \end{aligned}$$

Let X $\in$ E and Y $\in$ E be such that

$$N(X) = 0 \qquad \text{and} \qquad N^*(Y) = \nabla_X T.$$

By taking the inner product between $N(X)$ and Y and $N^*(Y)$ and X and subtracting the latter from the former, we obtain:

$$\langle N(X), Y \rangle_E - \langle N^*(Y), X \rangle_E = -\langle \nabla_X T, X \rangle_E$$

$$\therefore \langle \frac{\partial X}{\partial t} - \frac{\partial F}{\partial X} X, Y \rangle_E - \langle - \frac{\partial Y}{\partial t} - (\frac{\partial F}{\partial X})^* Y, X \rangle_E = -\langle \nabla_X T, X \rangle_E$$

$$\therefore \langle \frac{\partial X}{\partial t}, \text{Y} \rangle_E - \langle \frac{\partial F}{\partial X} X, Y \rangle_E + \langle (\frac{\partial F}{\partial X})^* Y, X \rangle_E = -\langle \nabla_X T, X \rangle_E$$

$$\therefore \langle \frac{\partial X}{\partial t}, Y \rangle_E + \langle \frac{\partial Y}{\partial t}, X \rangle_E + \langle \frac{\partial F}{\partial X} X, Y \rangle_E + \langle Y, \frac{\partial F}{\partial X} X \rangle_E = -\langle \nabla_X T, X \rangle_E$$

$$\therefore \frac{\partial}{\partial t} \langle X, Y \rangle_E = -\langle \nabla_X T, X \rangle_E$$

By integrating the last expression between t_0 and t_1, we have

$$\langle X_1, Y_1 \rangle_V - \langle X_0, Y_0 \rangle_V = - \int_{t_0}^{t_1} \langle \nabla_X T, X \rangle_E \, dt = -\langle \nabla_X J, X \rangle_E.$$

Choosing Y such that $Y(t_1) = Y_1 = 0$, the last expression reduces to:

$$\langle Y_0, X_0 \rangle_V = \langle \nabla_X J, X \rangle_E,$$

for all $X \in E$.
The main point here is the fact that in a vectorial space E endowed with an inner product it is possible to represent the differential of a linear functional f defined in E in the following form [6]:

$$f'(x) \cdot v = \langle \nabla f(x), v \rangle_E, \text{ for all } v \in E,$$

what has been used in the expression of the differential of the functional J above.
Therefore, as well as X_0 is the projection of X onto the vectorial space V, Y_0 is the projection of $\nabla_X J$ over V, which can be denoted by $\nabla_{X_0} J$, whence

$$Y_0 = \nabla_{X_0} J.$$

The notation $\nabla_{X_0} J$, abusive if understood as the gradient of J in relation to the variable X_0, since this functional is defined in the vectorial space E, should be correctly understood as the projection of the gradient of J over a vectorial sub-space of E isomorphic to V.
So, the control of the functional J relative to the data in E (initial conditions) is obtaining from the adjoint problem to problem (P): find the solution of

$$-\frac{\partial Y}{\partial t} - (\frac{\partial F}{\partial X})^* Y = \nabla_X T \ .$$

that verifies the condition $Y(t_1) = 0$.
There is undoubtedly an educational gain in thinking of $\nabla_{X_0} J$ as the restriction of $\nabla_X J$ to the vectorial space V: if the operator $\frac{\partial F}{\partial X}$ verifies certain regularity conditions, the problem of finding the solution of the equation:

$$\frac{\partial Y}{\partial t} = \frac{\partial F}{\partial X} Y,$$

being $Y(t_1) = Y_1$ is well posed in the Hadamard sense. Therefore, given $Y_1 \in V$, there exists an unique solution $Y \in E$ of the above equation such that $Y(t_1) = Y_1$. Then the evaluation of the functional J could be reduced to the evaluation of the functional restricted to the vectorial space V, which is isomorphic to a vectorial subspace of E, a computationally advantageous procedure.

The geometrical appeal of this strategy should be clear: restricting the evaluation of J to the vectorial space V instead of considering its evaluation in the vectorial space E is equivalent to the reduction of an integration defined on the set D in the vectorial space E to an integration on a set in the vectorial space V isomorphic to the vectorial space that contains the boundary of D.

Also it should be noted that, while classical partial differential equations are the mathematical expression of physical phenomena, adjoint equations, as introduced above, seem to be merely a mathematical artifice without any physical meaning. However consider the following problem defined in the interval $[t_0, t_1]$:

$$\frac{\partial X}{\partial t} - A(X) = 0, X(t_0) = X_0 \qquad (1)$$

$A : E \rightarrow E$ being a linear operator, whose adjoint problem is given by :

$$-\frac{\partial Y}{\partial t} - A^*(Y) = 0, Y(t_0) = Y_0 \qquad (2)$$

Taking again the inner product between Y and expression (1) and between X and expression (2) and subtracting them, we have:

$$\langle Y, \frac{\partial X}{\partial t} - A(X)\rangle_E - \langle X, -\frac{\partial Y}{\partial t} - A^*(Y)\rangle_E = 0$$

$$\therefore \frac{\partial}{\partial t}\langle X, Y\rangle_E = 0.$$

Integrating the last expression between t_i and t_j , $t_0 < t_i < t_j < t_1$, we obtain:

$$\langle X_{t_i}, Y_{t_i}\rangle_E = \langle X_{t_j}, Y_{t_j}\rangle_E \ ,$$

or

$$X_{t_i}^T Y_{t_i} = X_{t_j}^T Y_{t_j} \ .$$

Choosing $t_j = t_1$ and for the initial condition of (2) $Y(t_1) = X(t_1) = X_1$, we have:

$$X_{t_i}^T Y_{t_i} = X_{t_1}^2 \ . \qquad (3)$$

If the system is conservative, then:

$$X_0^2 = X_1^2.$$

then, with $t_i = t_0$, it results from (3) that:

$$X_0^T Y_0 = X_0^2$$

$$\therefore Y_0 = Y(t_0) = X(t_0).$$

Therefore, if the adjoint variable Y, solution of the adjoint equation (2), is the configuration of the system (1) at $t = t_1$, given by $Y(t_1) = X_1$, then it retrieves the system configuration at $t = t_0$.

3 Numerical simulations

As we know, there are two classes of algorithms able to perform the data assimilation: sequential and variational algorithms. In the first case, the structure of the algorithm is basically a prediction-correction one: based on the dynamics of the system and on observations made until the instant of time t_i it is produced an estimation of the state of the system at the time t_{i+1} , which is corrected using the observation of the system at that time. In the variational algorithms, one minimizes a functional that evaluates the discrepancy between the states of the system produced by the numerical model and the set of system observations.

Therefore, in the sequential algorithms, the flux of information about the system is always in increasing time direction, that is, the information about the system is used only in the correction of future states of the system produced by the numerical model, while, in the variational algorithms, the flux of information and the consequent corrections in the model predictions are both backward and forward. This is also the case when using the Kalman smoother [7].

Then the variational data assimilation process can be used in the reconstruction of a state of a system being studied, when its dynamics and a set of observations are available. In particular, this would be useful in the case of an accident in a medium whose dynamic is known, since we have a set of observations of the region after the accident.

In this study, we have considered the two-dimensional transport of a scalar amount released in a given medium, having a discrete set of observations of the generated spot, in both cases, complete and partial (at each instant of time) observations.

The spatial domain, with 6,000 m of length and 4,400 m of width, was discretized by means of 21×21 grid points, being $\Delta x = 300$ m and $\Delta y = 220$ m. The transport of the scalar $C(x, y, t)$ was obtained from the integration of the two-dimensional advection-diffusion equation, also referred to as Burgers' equation (see for instance [8]),

$$\frac{\partial C}{\partial t} + u_0 \frac{\partial C}{\partial x} - D\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2}\right) = 0 \qquad (1)$$

starting from the initial configuration as showed in figure 1 during a time interval equals to $100 \times \Delta t$, being $\Delta t = 300$ s .

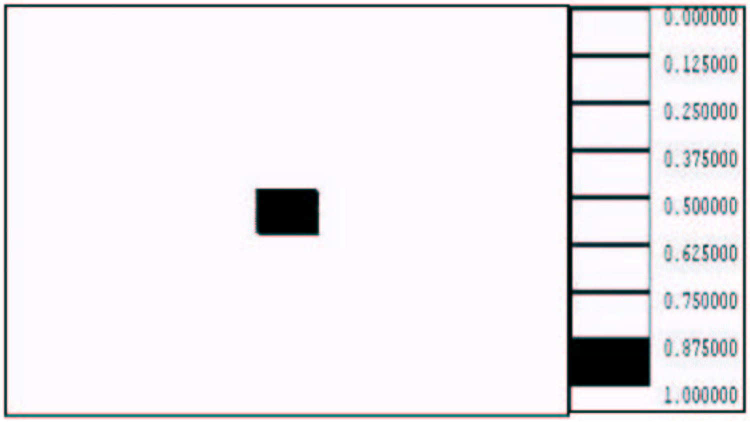

Figure 1: initial configuration

3.1 Complete observations case

The observations of the system consist of its configurations at time t $= 10 \times i \times \Delta t$, i $= 1, 9$, generated by the numerical model (see figures 2 and 3).
Equation (1) was solved using central in space and forward in time finite differences with the following boundary conditions:

left boundary C(1,J,K) = C(2,J,K)	right boundary C(21,J,K) = C(20,J,K)
top boundary C(I,21,K) = C(I,20,K)	bottom boundary C(I,1,K) = C(I,2,K)

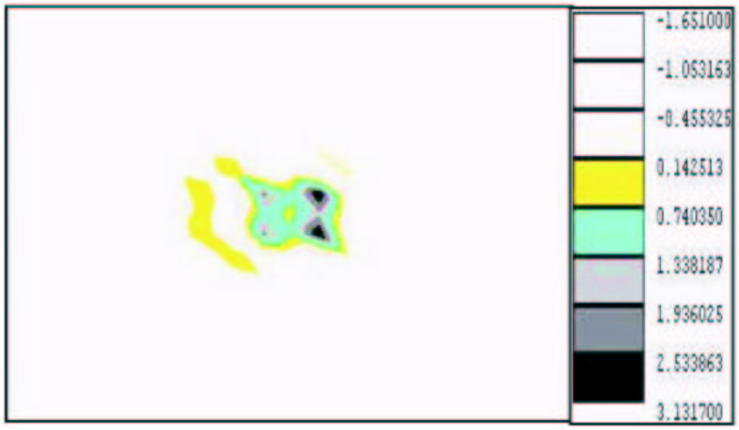

Figure 2: configuration at $t = 50 \times \Delta t$

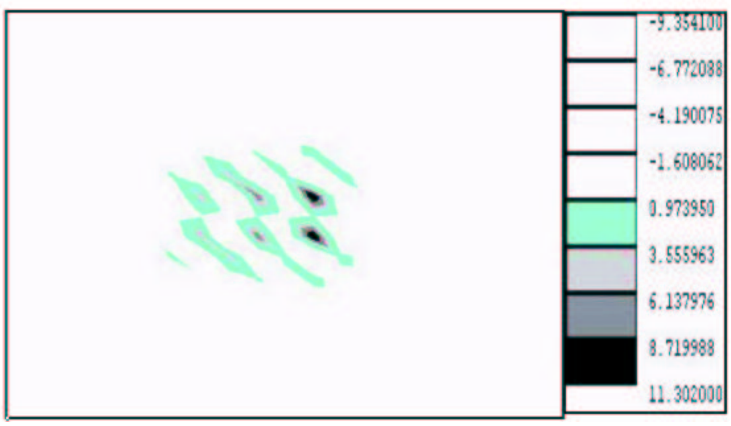

Figure 3: configuration at $t = 90 \times \Delta t$

The initial configuration of the system was randomly perturbed (figure 4) in order to produce the first guess for the initial configuration. Starting from this perturbed initial condition, one integration of the model gives the states of the system which will be compare with the available observations by means of a discretized version of the functional J.

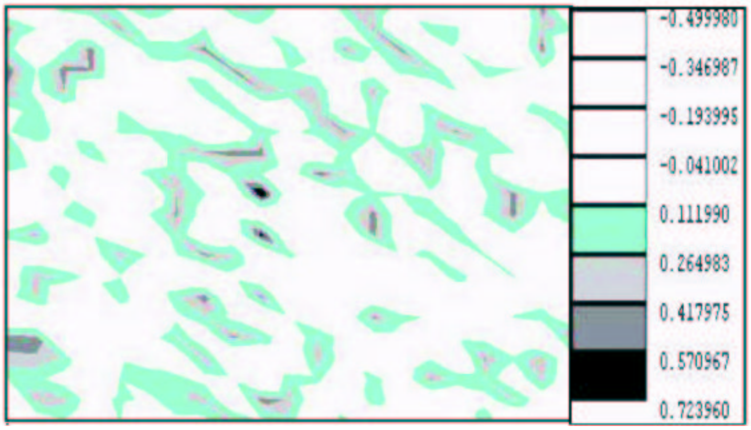

Figure 4: perturbed initial configuration

The minimization of the functional J with respect to the initial configuration of the system produces the retrieved initial configuration of the system. In this process it was used a routine based on the Quasi-Newton Limited Memory algorithm by Broyden, Fletcher, Goldfarb and Shanno, L-BFGS [9]. Figure 5 shows the retrieved initial configuration from the variational data assimilation process.

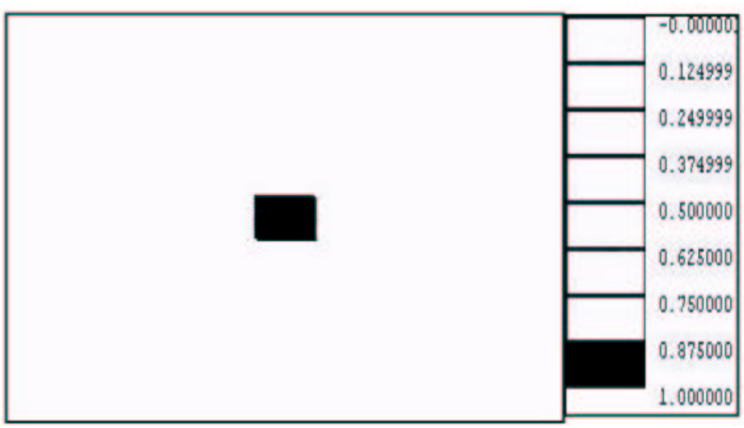

Figure 5: retrieved initial configuration

Apparently, see figure 6, the values of the perturbed configuration were adjusted according to the observations prior the shape adjustment.

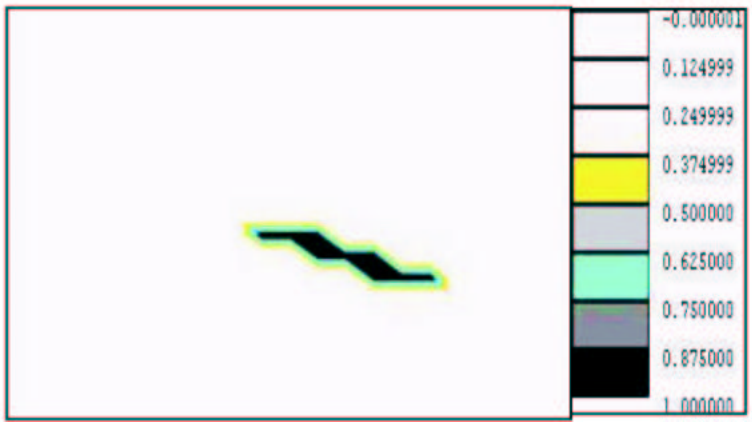

Figure 6: retrieved configuration after 7 minimization steps

3.2 Partial observations

With the same initial (perturbed) and boundary conditions as in the complete observation case, we used a set of matrices, M(k), in order to produce the partial observations of the system at each instant of time k. The matrices were defined as $M(I,J,10\times k) = 1$, if $2\times k \leq I \leq 2\times(k+1)$, and $M(I,J,10\times k) = 0$, in any other case. Thus the partial observations were obtained from the expressions $COBS(I,J,10\times k) = C(I,J,10\times k) \times M(I,J,10\times k)$, $I = 2, 20$, $J = 2, 20$, $k = 1, 9$, being $C(I,J,10\times k)$ the observations of the system at $t = 10\times k$, as in the case 3.1.

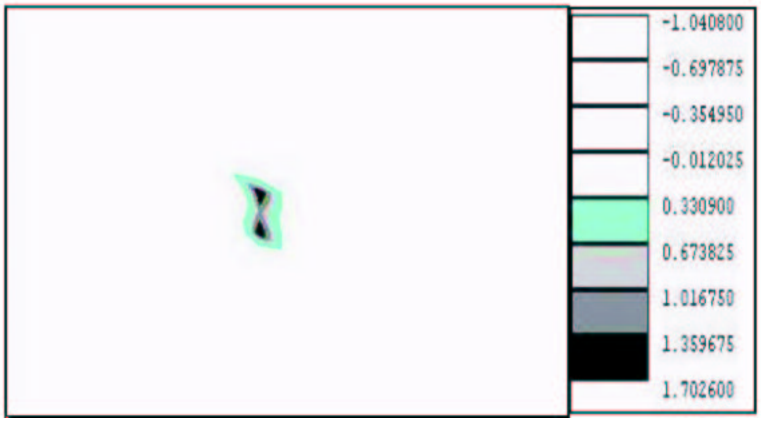

Figure 7: partial observation at t = 40 × Δt

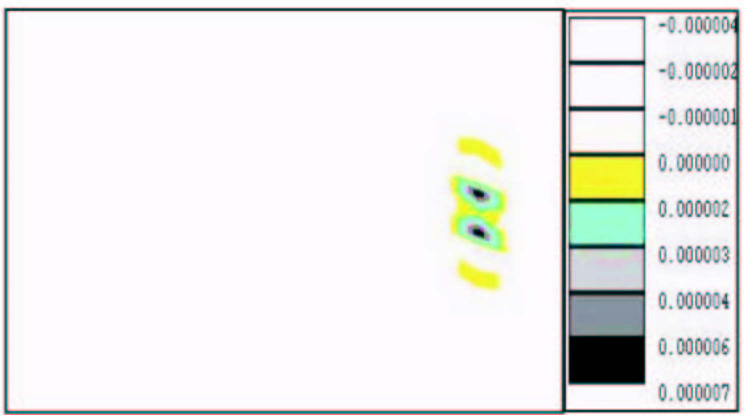

Figure 8: partial observation at t = 90 × Δt

Some observations are displayed in figures 8 and 9. They consist in taking only a narrow strip of data of the spatial domain at some instant of time during the observation period. It should be noted that, as defined, the partial observations constitute a complete set of observations in the sense that all points of the spatial domain have been observed at least at an instant of time during the observation period. In these circumstances, at least from a theoretical point of view, the recovering of the initial configuration of the system is feasible, considering only the observability of the system, but not its sensitivity to the initial conditions.

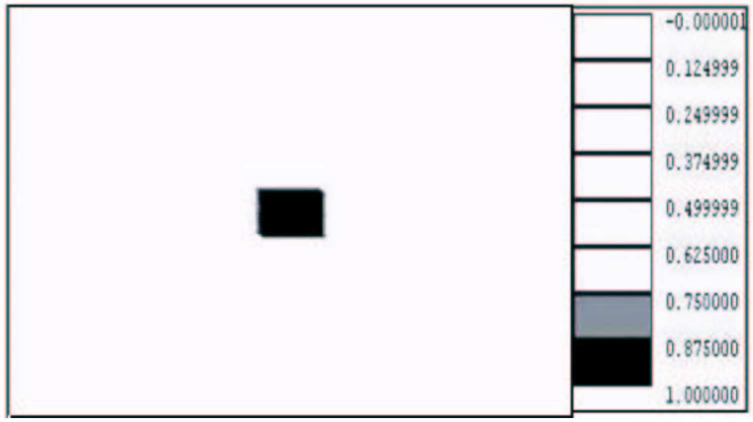

Figure 9: retrieved initial configuration using partial observations

Even in a simple experiment, as in the present case, with homogeneous data, the numerical simulations with partial observations reveal aspects of the problem that could be hardly anticipated. One of them could be formulated as follow: some sets of observations contain more information than others, even if the number of grid points in distinct sets is the same, that is, the influence of a set of observations on the solution of the problem is variable, and this influence would not be properly evaluated if considered only as a function of the number of grid points in a given spatial discretization of the domain, and this means that the information extracted from observations varies not only quantitatively, but also, and perhaps mainly, qualitatively.

In this numerical experiment, we observed the dependency of the retrieved initial configuration of the system on the quality of the observations. This was performed by disturbing a set of observations, for example, multiplying the variable C(I,J,K) by the factor 0.9, being $K \in \{1, 2, ..., 9\}$ (figure 10). By using the same perturbed initial configuration as in the previous section, the initial configuration of the system was recovered with a computational effort equivalent to that in the simulation of the complete observation case, but, this time, exhibiting a discrepancy with reference to the true initial configuration comparable with the error in observations introduced by the factor 0.9.

The impact of a set of incomplete observations in the assimilation process was analyzed in [10].

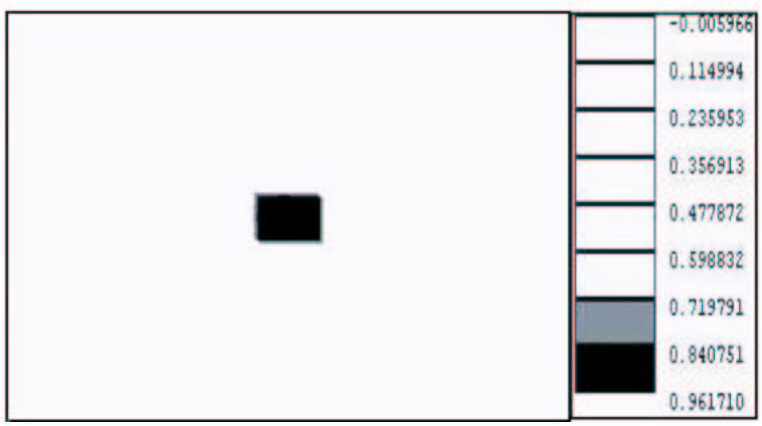

Figure 10: retrieved initial configuration using a modified observation set

4 Conclusions

The numerical experiments have shown, despite their simplicity, that the variational data assimilation process can be successfully used in the determination of a configuration of a transported scalar, at least when the system is completely observable, that is, when the value of the scalar is known in each point of the domain even if at different instants of time in the observation interval.
It should be emphasized that the procedure can only produce the system configuration at the initial instant of the observation interval, which does not necessarily coincide with the actual initial instant of the process. The practical consequence of this fact is that, although one can determine the evolution of the system, once its configurations during a certain observation interval is available, the exact determination of the initial configuration of the system as well as the determination of the initial time in the evolution of the system remain without answer. Nevertheless, there are results showing that, with an adequate modification in the expression of the functional J, it is possible to obtain, by using the adjoint formalism, the probability density function of the initial configuration and of the initial time.
Some experiments have shown the necessity of evaluating the impact of different observations of the system in study on the retrieved initial configuration based on this information and also on the system dynamics, what have been done in the domain of Adaptive Locations of the Observations.

References

[1] Z. Wang, I. M. Navon et al, The second order adjoint analysis: theory and applications, Meteorol. Atmos. Phys., 50, 3-20, 1992.

[2] F. X. le Dimet, H. E. Ngodock et al, Sensitivity analysis in variational data assimilation, Journal of the Meteorological Society of Japan, v.75, n.1B, 245-255, 1997.

[3] D. Daescu and G. Carmichael, An adjoint sensitivity method for the adaptive location of the observations in air quality modeling, pre-print, 2002.

[4] J. Pudykiewicz, Applications of adjoint tracer transport equations for evaluating source parameters, Atmos. Environment, v.32, n.17, 3039-3050, 1998.

[5] O. Talagrand and P. Courtier, Formalisation de la méthode du modèle adjoint, l'Etablissement d'Etudes et de Recherches Météorologiques, Note de travail, n.117, 1985.

[6] E. L. Lima, Curso de Análise, v.2, Projeto Euclides/IMPA, 1981.

[7] Z. Li and I. M. Navon, Optimality of 4D-Var and its relationship with the Kalman filter and Kalman smoother, Quarterly Journal of the Royal Meteorological Society, v.127, n.572, B, 661-684, 2001.

[8] C. A. J. Fletcher, Burgers equation: A model for all reasons, 1982.

[9] D. Liu and J. Nocedal, On the limited memory BFGS method for large scale minimization, Mathematical Programming, 45, 503-528, 1989.

[10] X. Zou, I. M. Navon and F. X. le Dimet, Incomplete observations and control of gravity waves in variational data assimilation, Tellus, 44A, 273-296, 1992.

[11] J. Noye, editor, Partial differential equations, North-Holland, 139-225, 1982.