Variational Data Assimilation, Optimal Parameter Estimation and Sensitivity Analysis for Environmental Problems*

I. M. Navon

Department of Mathematics and
Supercomputer Computations Research Institute,
Florida State University, Tallahassee, Florida, U.S.A.

1. INTRODUCTION

Optimal control theory of partial differential equations [1], [2] has emerged as a new way to attack problem of 4-D atmospheric/oceanic data assimilation problems. These variational techniques attempt to achieve a best fit between data (observations) and forecast model subject to some '*a priori*' criteria. This review paper presents new trends for utilizing model adjoint equations for variational data assimilation, parameter fitting and sensitivity analysis in general, as applicable to many areas of meteorological, oceanographic and environmental research.

2. VARIATIONAL DATA ASSIMILATION FORMALISM

The objective of variational 4-D data assimilation is to find the solution of a numerical forecast model which best fits a series of observational fields distributed over some space and time interval.

The distance between the solution and observations is given by

$$J(\mathbf{x}(t_0)) = \frac{1}{2}\sum_{r=0}^{R}\left(\mathbf{C}\mathbf{x}(t_r) - \mathbf{x}^{obs}(t_r)\right)^T \mathbf{W}(t_r)\left(\mathbf{C}\mathbf{x}(t_r) - \mathbf{x}^{obs}(t_r)\right) \tag{1}$$

where t_r represents the time when an observation occurs in the assimilation window $[t_0, t_a]$, R is the total number of time levels in the assimilation window when observations are available, $\mathbf{x}(t_r)$ is the N-component vector in the space R_N containing values of forecast model variables at time t_r, $\mathbf{x}^{obs}(t_r)$ is the M-component $(M \leq N)$ vector in space R_M containing values of observations at time t_r, $\mathbf{C}$ is a projection operator from space R_N to space R_M (i.e., a matrix $N \times M$) and $\mathbf{W}(t_r)$ is an $M \times M$ weighting matrix. J is defined as an $\mathcal{L}_2$-norm of $\mathbf{C}\mathbf{x} - \mathbf{x}^{obs}$. One can also choose an energy norm in finite dimensional space.

The values of $\mathbf{x}(t_r)$, $r = 0, 1, \ldots, R$ are obtained by integrating numerical model of form

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(t, \mathbf{x}) \tag{2}$$

*To appear in the Proceedings of the International Conference on Computational Engineering Science, Hawaii, July 30–August 3, 1995.

from the initial state $\mathbf{x}(t_0)$. In time discretized form, (2) will assume the following general form

$$\mathbf{x}(t_0 + \Delta t) = \mathbf{F}_1\left(\mathbf{x}(t_0)\right) + \mathbf{L}_1\mathbf{x}(t_0) \tag{3}$$

$$\mathbf{x}(t_r + \Delta t) = \mathbf{F}\left(\mathbf{x}(t_r)\right) + \mathbf{L}(t_r)\mathbf{x}(t_r - \Delta t) + \mathbf{m}(t_r)\mathbf{x}(t_r), \qquad \text{for} \quad r = 1, 2, \ldots, \tag{4}$$

where $\mathbf{F}_1(\mathbf{x})$ and $\mathbf{F}(\mathbf{x})$ are nonlinear operators and $\mathbf{L}_1$, $\mathbf{L}$ and $\mathbf{m}$ are linear operators.

To minimize the cost functional J, we need to use large-scale efficient minimization algorithms which require the gradient of cost function with respect to control variables (such as initial and/or boundary conditions).

To calculate the gradient of the cost function with respect to the initial condition, we will define a quantity J':

$$\begin{aligned} J'(\mathbf{x}(t_0)) &\equiv J\left(\mathbf{x}(t_0) + \mathbf{x}'(t_0)\right) - J\left(\mathbf{x}(t_0)\right) \\ &= \sum_{r=0}^{R} \left(\mathbf{W}(t_r)\left(\mathbf{C}\mathbf{x}(t_r) - \mathbf{x}^{obs}(t_r)\right)\right)^T \mathbf{C}\mathbf{x}'(t_r) + \sum_{r=0}^{R} O(\mathbf{x}'^2(t_r)). \end{aligned} \tag{5}$$

i.e., the change in the cost function resulting from a small perturbation $\mathbf{x}'(t_0)$ about the initial conditions $\mathbf{x}(t_0)$. In limit as $\|\mathbf{x}'\| \to 0$, J' is the directional derivative in the $\mathbf{x}'(t_0)$ direction and is given by

$$J'\left(\mathbf{x}(t_0)\right) = \left(\nabla J\left(\mathbf{x}(t_0)\right)\right)^T \mathbf{x}'(t_0). \tag{6}$$

Equating (5) and (6) results in

$$\left(\nabla J\left(\mathbf{x}(t_0)\right)\right)^T \mathbf{x}'(t_0) = \sum_{r=0}^{R} \left(\mathbf{W}(t_r)\left(\mathbf{C}\mathbf{x}(t_r) - \mathbf{x}^{obs}(t_r)\right)\right)^T \mathbf{x}'(t_r). \tag{7}$$

It is clear that if $\mathbf{x}'(t_r)$ can be expressed as a function of $\mathbf{x}'(t_0)$, then the gradient of the cost function with respect to the initial conditions can be found.

Since $\mathbf{x}'(t_r)$ is the perturbation at time t_r in the forecast resulting from the initial perturbation $\mathbf{x}'(t_0)$, it can be obtained by integrating the *tangent linear model*, which in turn can be obtained by linearizing the nonlinear model (3)-(4):

$$\mathbf{x}'(t_0 + \Delta t) = \frac{\partial \mathbf{F}_1\left(\mathbf{x}(t_0)\right)\mathbf{x}'(t_0)}{\partial \mathbf{x}} + \mathbf{L}_1\mathbf{x}'(t_0) \tag{8}$$

$$\mathbf{x}'(t_r + \Delta t) = \frac{\partial \mathbf{F}\left(\mathbf{x}(t_r)\right)\mathbf{x}'(t_r)}{\partial \mathbf{x}} + \mathbf{L}\mathbf{x}'(t_r - \Delta t) + \mathbf{m}(t_r)\mathbf{x}'(t_r), \qquad \text{for } r = 1, 2, \ldots, \tag{9}$$

which may then be rewritten symbolically as

$$\mathbf{x}'(t_r) = \mathbf{P}_r\mathbf{x}'(t_0) \tag{10}$$

where $\mathbf{P}_r$ represents the result of applying all the operator matrices in the linear model to obtain $\mathbf{x}'(t_r)$ from $\mathbf{x}'(t_0)$.

Using the tangent linear model (10), (7) becomes

$$\left(\nabla J\left(\mathbf{x}(t_0)\right)\right)^T \mathbf{x}'(t_0) = \sum_{r=0}^{R} \left(\mathbf{W}(t_r)\left(\mathbf{C}\mathbf{x}(t_r) - \mathbf{x}^{obs}(t_r)\right)\right)^T \mathbf{C}\mathbf{P}_r\mathbf{x}'(t_0). \tag{11}$$

This implies

$$\nabla J\left(\mathbf{x}(t_0)\right) = \sum_{r=0}^{R} \mathbf{P}_r^T\mathbf{C}^T\mathbf{W}(t_r)\left(\mathbf{C}\mathbf{x}(t_r) - \mathbf{x}^{obs}(t_r)\right), \tag{12}$$

where $\mathbf{P}_r^T$, $r = 1, 2, \ldots, R$ are the corresponding adjoint operators of the linear operators $\mathbf{P}_r$, $r = 1, 2, \ldots, R$ in the tangent linear model. Therefore, the gradient of the cost function may be obtained by a single integration of the adjoint model from final time t_a to initial time t_0 of the assimilation window with zero initial conditions for the adjoint variables at time t_a while the weighted differences

$$\mathbf{C}^T \mathbf{W}(t_r) \left(\mathbf{C}\mathbf{x}(t_r) - \mathbf{x}^{obs}(t_r) \right), \quad r = R, R-1, \ldots, 0 \tag{13}$$

are inserted on the right-hand-side of the following adjoint model

$$\hat{\mathbf{x}}(t_0) = \mathbf{P}^T \hat{\mathbf{x}}(t_a), \tag{14}$$

whenever an observational time $t_r (r = R, R-1, \ldots, 0)$ is reached, where $\hat{\mathbf{x}}$ represents the adjoint variables.

3. PARAMETER ESTIMATION

Parameter identification refers to the determination from observed data of unknown parameters in the system model such that the predicted response of the model is in some sense close to the process observations. Some of the parameters represent physical properties which cannot be measured readily, while others belong to parameterization schemes.

A sizable amount of work on adjoint parameter estimation was carried out in groundwater hydrology and petroleum reservoirs along with advances in the mathematical community.

In meteorology, Courtier (1986, 1987) [3], [4] estimated orography of a shallow-water equations model. Zou et al. (1992) [5] estimated modeling coefficients in operational weather prediction spectral model.

In oceanography, early work was carried out by [6] for bottom drag coefficients in a tidal channel.

Typically [5] use a cost function defined by

$$J(\mathbf{X},\ \mathbf{P}) = \int_{t_0}^{t_R} < W(\mathbf{X} - \mathbf{X}^o),\ \mathbf{X} - \mathbf{X}^o > \ dt + \int_{t_0}^{t_R} < K(\mathbf{P} - \hat{\mathbf{P}}),\ \mathbf{P} - \hat{\mathbf{P}} > \ dt \tag{15}$$

where $\hat{\mathbf{P}}$ are the estimated parameters, the K are specified weighting matrices different for the weighting matrix W. The scalar product $<, >$ is usually in $\mathcal{L}_2$, $\mathbf{X}$ is model forecast in R_N and $\mathbf{X}^o \equiv \mathbf{X}^{obs}$.

The adjoint model equation and the gradient of the cost function are

$$\frac{\partial Q}{\partial t} + \left[\frac{\partial F}{\partial \mathbf{x}} \right]^T Q - P^T Q = W(\mathbf{X} - \mathbf{X}^o) \tag{16}$$

(Here we took $C \equiv I$, otherwise we need to use the Moore-Penrose unique generalized inverse of C.) We see that an additional term $-P^T Q$ was added to the left-hand side of the adjoint model equation and the gradient of the cost function with respect to P is

$$\nabla_p J = - \int_{t_0}^{t_R} < (\mathbf{X} - \mathbf{X}^o,\ Q\ dt > \ + 2K(\mathbf{P} - \hat{\mathbf{P}}) \tag{17}$$

Q here are the adjoint variables.

A priori one may expect some forecasts to manifest insensitivity with respect to some parameters, so we intend to choose parameters by a "adjoint sensitivity analysis".

Another strategy is to optimally estimate classes of carefully chosen parameters related to a given physical environmental process at a time, e.g., precipitation, pollutant concentration and proceed to another class (say radiative transfer) only after in depth experimentation with the first set of parameters yielded satisfactory results. This approach called "history matching" was used to fine true parameters in a computationally feasible way [7].

Some parameters are bounded within a range of acceptable values, so one has to solve constrained optimization problem. For such solutions see work of [8] and references therein.

In depth mathematical analysis of the parameter estimation problem includes work of [9], [10], [11] and [12] and references therein.

Issues of identifiability and uniqueness of the solution as well as whether the solution of optimal parameter estimation depends continuously on the observations (stability) will be addressed in some detail in my talk (see [13], [9]).

4. SENSITIVITY ANALYSIS

Sensitivity is a measure of the effect of changes in a given input parameter on a selected response (any forecast aspect). The general definition of sensitivity of a response to variations in system parameters is the *Gâteau* differential.

The G differential $VR(\mathbf{X}^0, \boldsymbol{\alpha}^{\mathbf{o}}, \mathbf{h}_{\boldsymbol{x}}, \mathbf{h}_\alpha)$ of a specific response $R(\mathbf{X}, \boldsymbol{\alpha})$

$$R(\mathbf{x}, \boldsymbol{\alpha}) = \int_{\boldsymbol{t_o}}^{\boldsymbol{t_a}} \boldsymbol{r}(\boldsymbol{t}; \mathbf{x}, \boldsymbol{\alpha}) \boldsymbol{dt} \tag{18}$$

at the nominal values $(\mathbf{X}^0, \boldsymbol{\alpha}^{\mathbf{o}})$, where $\boldsymbol{\alpha}$ is a model parameter vector, for increments $(\mathbf{h}_x, \mathbf{h}_\alpha)$ around $(\mathbf{X}^0, \boldsymbol{\alpha}^{\mathbf{o}})$ is given by

$$VR(\mathbf{x}^0, \boldsymbol{\alpha}^{\mathbf{o}}; \mathbf{h}_{\boldsymbol{x}}, \mathbf{h}_\alpha) = \int_{\boldsymbol{t_o}}^{\boldsymbol{t_R}} \boldsymbol{r'}_{\mathbf{x}} \cdot \mathbf{h}_{\boldsymbol{x}} \boldsymbol{dt} + \int_{\boldsymbol{t_o}}^{\boldsymbol{t_R}} \boldsymbol{r'}_{\boldsymbol{\alpha}} \cdot \mathbf{h}_\alpha \boldsymbol{dt} \tag{19}$$

$$\boldsymbol{r'}_{\mathbf{x}} = \left[\left(\frac{\partial r}{\partial x_1}, \ldots, \frac{\partial r}{\partial x_P} \right) \right]_{(\mathbf{x}^0 \boldsymbol{\alpha}^{\mathbf{o}})} \tag{20}$$

$$\boldsymbol{r'}_{\boldsymbol{\alpha}} = \left[\left(\frac{\partial r}{\partial \alpha_1}, \ldots, \frac{\partial r}{\partial \alpha_N} \right) \right]_{(\mathbf{x}^0, \boldsymbol{\alpha}^{\mathbf{o}})} \tag{21}$$

N is the dimension of the vector of model parameters and P is the dimension of the model variable.

If a variation occurs solely in the n-th parameter the corresponding variation $h^n_{\boldsymbol{\alpha}}$ of the parameter vector is

$$h^n_{\boldsymbol{\alpha}} = (0, \ \ldots, \ h^n_\alpha, \ \ldots, \quad 0)^T \tag{22}$$

and the corresponding sensitivity is VR^n.

The relative sensitivity S_n is the dimensionless quantity

$$S_n = \frac{VR^n}{R} \left(\frac{\mathbf{h}^n_\alpha}{\boldsymbol{\alpha}^{\mathbf{o}}_{\boldsymbol{n}}} \right)^{-1} \tag{23}$$

The *relative* sensitivity clearly demonstrates the measure of the importance of the input parameter. The higher the relative sensitivity, the more important the input parameter in question. Thus, one of the crucial aspects of sensitivity analysis is to identify the most important input parameters whose changes impact the most the chosen response. The magnitudes of relative sensitivities can serve as a guide to ranking importance of model parameters for use in choosing candidates for optimal parameter estimation.

For models that involve a large number of parameters and comparatively few responses, sensitivity analysis can be performed very efficiently by using deterministic methods based on adjoint functions. It can be shown [14] that the changes in the response function can be expressed in terms of adjoint dynamics $q(t)$, which is an adjoint variable corresponding to the model variable $x(t)$. The use of the adjoint model eliminates the need to calculate, by forward integration, $\delta x(t)$ $(= x(t) - x^0(t))$, a quantity whose dynamics is governed by the so-called linear tangent equations, where $x(t)$ and $x^0(t)$ are the perturbed and the actual model trajectories in phase space. These forward calculations happen to be explicitly dependent on the changes in the initial conditions $\delta x(t_0)$ and the model parameters changes and must be repeated every time these are altered; the formulation using the adjoint solution to the linear tangent dynamics does not suffer from this shortcoming and is therefore extremely economical when dealing with large models possessing several parameters.

Based on work of [15], [16], [14] extended sensitivity analysis to general operator type responses such as time and space dependent functions of the model state variables and parameters. This since the most interesting and revealing meteorological cases involve sensitivity with respect to operator responses that depend on both time and space. We intend to use those methods to carry out extensive sensitivity studies using the NMC model with physics and its adjoint.

Sensitivities amongst other may quantify the extent that uncertainties in parameters contribute to uncertainties in model results. Furthermore the adjoint sensitivity analysis may also provide a quantitative measure of the importance of data or a region in phase space in contributing to an adequately chosen response function. One limitation of such sensitivity study is the restriction of each result to one of the forecast aspects. Therefore, one should carefully select different responses.

Sensitivity studies with the full physics adjoint model will be carried out with the intention of using a stratification based on various meteorological seasons corresponding to different meteorological situations - for instance, seasonally dependent sensitivities (see [17]). In our ranking of model parameters to be used in the parameter estimation, we shall use various response functions and compare resulting parameter rankings. This implies stratification of relative sensitivities both by response and by season - thus implying that the subset of model parameters chosen is of real impact on the model response. It is well known that the adjoint sensitivity approach indicates both geographical areas and meteorological parameters to which a given model function is most sensitive [18].

REFERENCES

[1] J. L. Lions, Optimal Control of Systems Governed by Partial Differential Equations, translated by S. K. Mitter, Springer-Verlag, Berlin-Heidelberg (1971), p.396.
[2] G. I. Marcuck, Numerical Solution of the Problems of the Dynamics of the Atmosphere and Ocean (in Russian, 1974), Gidrometeoizdat, 303pp.
[3] P. Courtier, Le modele adjoint, outil pour des experiences de sensibilite. Note de travail no 166. EERM (1986), Paris, France, 36pp. (available from Meteo-France).
[4] P. Courtier, Application du contrôle optimal á la prévision numerique en Météorologie. These de doctorat de l'universite Paris 6 (Reading 1987).
[5] X. Zou, I.M. Navon, and F. X. LeDimet, An Optimal Nudging Data Assimilation Scheme Using Parameter Estimation, Q. J. of Roy. Met. Soc., Vol. 118 (1992), pp.1163–

1186.
[6] V. G. Panchang and J. J. O'Brien, On the Determination of Hydraulic Model Parameters Using the Strong Constraint Formulation, Modeling Marine Systems, Vol. I (1988), editor: A. M. Davies, CRC Press, Inc., pp.5–18.
[7] R.E. Ewing and T. Lin, Parameter Identification Problems in Single Phase and Two-Phase Flow, in *Control and Estimation of Distributed Parameter Systems* (1989), Kappel, Kunish, and Schappacher, eds., ISNM 91, Birkhauser Verlag, Boston, pp.87–108, 434pp.
[8] K. Kunish and E.W. Sachs, Reduced SQP Methods for Parameter Estimation Problems, SIAM J. Numer Anal., Vol. 29, No. 6 (1992), pp.1793–1820.
[9] G. Chavent, Identification of Distributed Parameter Systems: About the Output Least Squares Methods, Its Implementation and Identifiability, Proc. 5th IFAC Symp. on Identification and System Parameter Estimation (Darmstadt) (1979), Pergamon Press, Oxford, pp.85–97.
[10] S. Omatu and John H. Seinfeld, Distributed Parameter Systems: Theory and Applications, Oxford Mathematical Monographs (1989), Oxford University Press, 430pp.
[11] H. T. Banks and K. Kunish, Estimation Techniques for Distributed Parameter Systems, Birkhauser (1989), Boston, 315pp.
[12] H. T. Banks, Computational Issues in Parameter Estimation and Feedback Control Problems for Partial Differential Equation Systems, Physica D, Vol. 60 (1992), pp.226–238.
[13] C. Kravaris and J.H. Seinfeld, Identification of Parameters in Distributed Parameters Systems by Regularization, SIAM J. on Control and Optimization, Vol. 23 (1985), pp.217–241.
[14] X. Zou, I.M. Navon, A. Barcilon, Jeff Whittaker, and Dan Cacuci, An Adjoint Sensitivity Study of Blocking in a Two-Layer Isentropic Model, Monthly Weather Review, Vol. 121, No. 10 (1993), pp.2833–2857.
[15] D. G. Cacuci, Sensitivity Theory for Nonlinear Systems: II. Extensions to Additional Classes of Responses, J. Math. Phys., Vol. 22 (1981), pp. 2803-2812.
[16] D. G. Cacuci, The Forward and Adjoint Methods of Sensitivity Analysis, pp.71-144, Chapter 3 of the book, *Uncertainty Analysis*, by Yigal Ronen (1988), CRC Press, Inc., 282pp.
[17] R. M. Errico, T. Vukićević, and K. Raeder, Comparison of Initial and Lateral Boundary Condition Sensitivity for a Limited Area Model, Tellus, Vol. 45A (1993), pp.539–557.
[18] F. Rabier, P. Courtier, and O. Talagrand, An Application of Adjoint Models to Sensitivity, Beitr. Phys. Atmosph., Vol .65 No. 3 (1992), pp.177–192.