ELSEVIER Dynamics of Atmospheres and Oceans 27 (1997) 55–79

dynamics of atmospheres and oceans

Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography [1]

I.M. Navon

Department of Mathematics and Supercomputer Computations Research Institute, Florida State University, Tallahassee, FL 32306-4052, USA

Received 3 April 1996; revised 28 January 1997; accepted 29 January 1997

Abstract

The present paper has two aims. One is to survey briefly the state of the art of parameter estimation in meteorology and oceanography in view of applications of 4-D variational data assimilation techniques to inverse parameter estimation problems, which bear promise of serious positive impact on improving model prediction. The other aim is to present crucial aspects of identifiability and stability essential for validating results of optimal parameter estimation and which have not been addressed so far in either the meteorological or the oceanographic literature.

As noted by Yeh (1986, *Water Resour. Res.* 22, 95–108) in the context of ground water flow parameter estimation the inverse or parameter estimation problem is often ill-posed and beset by instability and nonuniqueness, particularly if one seeks parameters distributed in space-time domain. This approach will allow one to assess and rigorously validate results of parameter estimation, i.e. do they indeed represent a real identification of physical model parameters or just compensate model errors? A brief survey of other approaches for solving the problem of optimal parameter estimation in meteorology and oceanography is finally presented. © 1997 Elsevier Science B.V.

1. Introduction

While a sizable amount of research on adjoint parameter estimation was carried out in the last 20 years in fields such as groundwater hydrology and petroleum reservoirs for instance by Carrera and Neuman (1986a,b,c), Yeh (1986), Cushman-Roisin (1986),

[1] Dedicated to Professor Richard Pfeffer.

PII S0377-0265(97)00032-8

Seinfeld and Kravaris (1982), Sun and Yeh (1990a,b, 1992), matched by a major effort of the mathematical community started early by seminar work of Richard Bellman and collaborators (Bellman et al., 1965a,b) Chavent and Lemonnier (1974) and Chavent et al. (1975), adjoint parameter estimation work in meteorology and oceanography is more recent and consists of fewer contributions lacking the in-depth approach for validation of the uniqueness of results taken in above mentioned research fields.

In this paper directed at the segment of the meteorological and oceanographic community which focuses on data assimilation by 4-D VAR or estimation theory, we aim to achieve two main goals:

The first is to present a brief review of the state of the art of parameter estimation in the meteorological and oceanographic community.

The second goal is to illuminate without recourse to the use of heavy mathematics the issue of ill-posedness of the problem of parameter estimation along with a description of the problems of identifiability and uniqueness which may preclude the possibility of a successful parameter estimation procedure. The rigorous mathematical background is provided in Appendix A.

The paper plan is as follows. Section 2 as mentioned above provides a brief survey of state of the art parameter estimation in meteorology and oceanography along with the typical set-up for adjoint parameter estimation. Section 2.3 describes issues of identifiability ill-posedness and regularization in a qualitative way and illustrates the issue with examples. An effort is made to point out the importance of these concepts for the adjoint parameter estimation procedure to be relevant and for the results to be uniquely validated. Section 3 describes an alternative approach to adjoint parameter estimation focusing on the maximum likelihood (ML) method. The use of total variation as a regularization method for parameters with discontinuities is presented in Section 4. Parameter estimation via the extended Kalman filter (EKF) is presented in Section 5. The issue of the regularization procedure is also discussed. Sensitivity analysis as an efficient tool in parameter estimation in meteorology and oceanography is discussed in Section 6. Summary and conclusions are then presented in Section 7. Mathematical results related to identifiability are presented in Appendix A for the sake of completion.

2. State of the art of adjoint parameter estimation in meteorology and oceanography

2.1. Parameter estimation in meteorology

The research effort on adjoint parameter estimation in meteorology can be dated back to the work of Courtier (1986, 1987) on estimating orography using a shallow-water equations model.

An early detailed survey addressing also issues of adjoint parameter identification was provided by Le Dimet and Navon (1988). Early monographs treating inverse problems in geophysics are those of Menke (1984) and Tarantola (1987).

Zou et al. (1992) estimated the magnitude of the nudging coefficient in the NMC adiabatic version of the spectral MRF model, while Wang (1993) and Wang et al. (1995)

estimated the same coefficient using the FSU (Florida State University) adiabatic spectral model. Stauffer and Bao (1993) performed a parameter estimation of nudging coefficients in a 1-D linearized shallow water equations model.

Wergen (1992) used also a 1-D linearized shallow-water equations model to recover both initial state and a set of forcing parameters from the observations. Wergen found out that even with noisy observation the parameters were recovered to an acceptable degree of accuracy.

Louis and Živković (1994) carried out physical parameters estimation in a simplified single column model, and their effort presents a more comprehensive approach to parameter estimation, making it amenable to generalization to problems of parameter estimation involving 3-D numerical weather prediction models.

The research methodology used in adjoint parameter estimation in meteorology can be viewed to be an extension of the 4-D VAR approach for controlling initial or initial and boundary conditions.

Some authors determine which are the crucial physical package parameters to be optimally identified based either on experience or using a relative adjoint sensitivity analysis. Such an analysis enables one to rank a subset of chosen parameters according to their relative sensitivities to adequately chosen model responses.

The usual procedure for assessing the impact of an optimized parameter requires testing impact on the model for a sufficiently long period, thus ensuring that no degradation of the forecast is caused by the optimally estimated parameter.

Since some parameters are known to vary between given bounds, the minimization of the cost functional (to be described below) will by necessity be of the constrained minimization type. Several efficient constrained minimization procedures are available (for instance see Nash and Sofer, 1996, for details). For methods of unconstrained minimization see Navon and Legler (1987, Navon et al. (1992a,b) and Zou et al. (1993b) along with an excellent review by Nocedal (1991).

An essential monograph explaining analysis and application of assimilation methods is the one by Daley (1991).

If an optimally estimated parameter attains unphysical values, one can deduce that either an overfitting of the data took place, or that this parameter is not identifiable with the data available. We will address this issue in detail in the section related to identifiability.

Stratification of groups of parameters to be optimally identified may proceed by either seasonal stratification or following a given physical process at a time. Due to the nonlinear feedbacks that exist between classes of physical parameters, one should proceed with caution when increasing the dimensionality of the problem, i.e. by adding a new class of physical parameters to be optimally identified.

2.2. *Parameter estimation in oceanography*

In oceanography optimal control parameter estimation issues have been addressed in early work by Bennett and McIntosh (1982) and Prevost and Salmon (1986) using Sasaki's weak constraint formalism. Early adjoint parameter estimation work is also the one carried out by Panchang and O'Brien (1989) for bottom drag coefficient identification in a tidal channel, while phase speeds were estimated in the review article by Smedstad and O'Brien (1991) using the adjoint method in a reduced gravity model for the tropical Pacific Ocean. Yu and O'Brien (1991) carried out a wind stress coefficient estimation along with the estimation of the oceanic eddy viscosity profile. Heemink (1987) used estimation theory for shallow water flow identification.

Early reviews of data assimilation methods in oceanographic applications are the lectures by Miller (1987) and the special issue of *Dynamics of Atmospheres and Oceans* edited by Haidvogel and Robinson (1989).

The book of Bennett (1992) provides a mathematically advanced and thorough synthesis of oceanographic assimilation methodologies.

An important comprehensive review serving as a milestone in adjoint methodology in meteorology and oceanography and summarizing state-of-the-art for beginning of the 1990s, including issues of parameter estimation, is the excellent review by Ghil and Malanotte-Rizzoli (1991).

Typical recent advances related to assimilating data into complex, time-dependent ocean general circulation models are those of Fukumori et al. (1993) and Tziperman et al. (1992).

Das and Lardner (1990, 1992) estimated bottom friction and water depth in a 2-D tidal model and related to the issue of the number of observations required for a given mesh resolution in order to achieve satisfactory parameter identification.

Richardson and Panchang (1992) are the first in oceanography to have noted the difficulties associated with parameter estimation (eddy viscosity) with adjoint methods (and inverse modeling techniques in general) due to their being beset by instabilities and nonuniqueness when identifying parameters distributed in the space time domain, especially when the data is noisy. To overcome these difficulties they proposed to insert an additional criterion in the cost functional formulation namely that the parameter profiles should be smoothly varying, i.e. a compromise between data-model misfit minimization and the solution smoothness. A similar approach was introduced in the context of ground water flow parameter estimation in the doctoral thesis of Cushman-Roisin (1986). Additional work on eddy viscosity profile parameter estimation in a 3-D tidal model was carried out by Lardner and his collaborators (see Lardner and Song, 1995, and Lardner (1993), Lardner et al. (1993), Lardner and Das (1994)).

Additional recent parameter estimation applications in oceanography include work of Tziperman and Thacker (1989), estimating friction and wind forcing in ocean circulation models, Marotzke (1992), Marotzke and Wunsch (1993), Chertok and Lardner (1996), Gunson (1995), Yu and O'Brien (1992) and ten Brummelhuis et al. (1993) which used parameter estimation for tidal models.

A recent book presenting a comprehensive picture of the state-of-the-art of data assimilation in oceanography in the mid 90s and addressing also in several chapters the topic of parameter estimation is the excellent monograph edited by Malanotte-Rizzoli (1996). In particular the contributions of Malanotte-Rizzoli and Tziperman (1996) and references therein and that of Hogg (1996) in the above monograph, represent modern lucid contributions to state-of-the-art in parameter estimation in oceanography.

The book by Wunsch (1996) is also a very well-written, lucid and useful contribution including Gauss–Markov estimation, sequential estimators and adjoint/Pontryagin principle devoted to finite dimensional problems and methods along with a very useful subsection on parameter estimation geared towards the general topics of parameter estimation.

I would finally like to caution the reader that I only addressed in this subsection literature on the adjoint parameter estimation aspect in oceanography and thus covered but a narrow aspect of the vast literature on oceanographic adjoint applications. The monograph edited by Malanotte-Rizzoli (1996) is an authoritative source for references to these applications.

2.3. Identifiability

The uniqueness problem in parameter identification is ultimately related to the issue of parameter identifiability.

The notion of identifiability addresses the question of whether it is at all possible to obtain unique solutions of the inverse problem for unknown parameters of interest in a meteorological/oceanic model from data collected in the spatial and temporal domains.

Simply put (see Kitamura and Nakarigi, 1977) the parameter identification problem can be formulated as the one-to-one property of mapping from the space of system outputs to the space of parameters. The uniqueness of such a mapping is extremely difficult to establish. A working definition of identifiability is as follows (Kitamura and Nakarigi, 1977): "We shall call an unknown parameter 'identifiable' if it can be determined uniquely in all points of its domain by using the input–output relation of the system and the input–output data".

An early review work in the field of identification of parameters was provided by Goodson and Polis (1979).

Chavent (1979, 1983, 1991) presented a definition of identifiability using the output least-squares error criterion as used in 4-D VAR. If such a criterion is used for solving the inverse problem of parameter identification the parameter is said to be output least-square identifiable if and only if a unique solution of the optimization problem exists and the solution depends continuously on the observations.

We will now provide two simple illustrative examples related to identifiability (see Banks and Kunish, 1989).

The first is related to ill-posedness of parameter estimation problems—where there is a lack of a continuous inverse of the parameter-to-observation mapping.

2.3.1. Example 1 (Banks and Kunish, 1989)

Let Q be the set of parameters which guarantee that a chosen model equation has solution $u(q)$, let C be observation operator mapping from solution space of model to observation space Z. Using these mappings we have the parameter-to-output mapping Φ: $Q \to Z$ with $\Phi(q) = Cu(q)$.

Let us assume our model is the 1-D equation

$$-(qu_x)_x = f \text{ on } (0,1) \quad (1)$$

$$u(0) = u(1) = 0$$

Here f is assumed to be known and q is the unknown parameter.

Integrating Eq. (1) formally we obtain for $x_p \in (0, 1)$

$$q(x) = \frac{u_x(x_p)}{u_x(x)} q(x_p) - \frac{1}{u_x(x)} \int_{x_p}^{x} f(s) \mathrm{d}s \quad \text{for} \quad x \in [0,1] \tag{2}$$

If $u_x > 0$ (or $u_x < 0$) on [0, 1] then q is uniquely determined by Eq. (2) provided $q(x_p)$ is given for some $x_p \in [0, 1]$. If u_x has precisely one root we may take this point as x_p and define from Eq. (1)

$$q(x_p) = \frac{f(x_p)}{u_{xx}(x_p)}$$

provided $u_{xx}(x_p) \neq 0$.

Then q is determined uniquely without specification of q at any point in its domain.

Also q can be bounded in terms of $u_x(q)$ provided that u_x can be bounded away from zero. However, we cannot bound the inverse of

$$\Phi(q) = u(q)$$

as a mapping from $C(0, 1)$ to $C(0, 1)$ even in a neighborhood of some u^* satisfying $u_x^* > 0$.

To see this let

$$u_k(x) = x + \frac{1}{(2k+1)\pi} \sin 2k\pi x \quad \text{on} \quad [0,1], \quad k = 1,2,\ldots$$

and assume $f = 0$ and $q(0) = 1$, are known, then

$$(u_k)_x = 1 - 2k/(2k+1)$$

$$u_k(0) > 0$$

$$u_k(1) = 1 \quad \text{for all} \quad k$$

and $u_k \to u^*$ with $u^*(x) = x$ in $C(0, 1)$.

On the other hand by Eq. (2) we have for the corresponding parameters

$$q_k(x) = \frac{(u_k)_x(0)}{(u_k)_x(x)} = \frac{1 + \frac{2k\pi}{(2k+1)\pi}}{1 + \frac{2k\pi}{(2k+1)\pi} \cos 2k\pi x} \tag{3}$$

which is a divergent series in $C(0, 1)$. This example illustrates the lack of a continuous inverse of the parameter-to-observation mapping of parameter estimation problems.

2.3.2. Example 2 (Banks and Kunish, 1989)

We consider the case of a parabolic model equation and show that the unknown parameters are identifiable. Then a spectral approximation to the model equation is performed for which the parameters are again identifiable.

Nevertheless it is not possible to use a spectral approximation in an output least squares (OLS) formulation to successfully estimate the unknown parameters.

The model equation is:

$$u_t = q_1 u_{xx} + q_2 u \text{ for } t>0,\ 0<x<1 \tag{4}$$
$$u(0,x) = \Phi(x),\ 0 \le x \le 1$$
$$u(t,0) = u(t,1) = 0,\ t>0$$

where Φ is the piecewise linear interpolant satisfying

$$\Phi(0) = \Phi(1) = 0,\ \Phi\left(\frac{1}{2}\right) = 1$$

The solution of Eq. (4) is given by

$$u(t,x,q_1,q_2) = \sum_{i=1}^{\infty} \tilde{u}_i \exp\left(\left(q_2 - q_1(i\pi)^2\right)t\right)\sin i\pi x \tag{5}$$

where $\tilde{u}_i$ are the Fourier coefficients of the sine series for Φ. Observe that

$$\tilde{u}_{2i} = 0,\ \text{for } i = 1,2,\ldots$$

Let us assume that the observations are of the form

$$u(t^*,\cdot) \in L^2 \text{ for such } t^* \in (0,\infty)$$

and let

$$q^* = (q_1^*,q_2^*),\ q_1^* > 0$$

be the 'true' parameters with corresponding observation

$$Z = u(t^*,\ldots,q^*)$$

Then if $u(t^*,\ \cdot,\ q^*) = u(t^*,\ \cdot,\ q)$, $q = (q_1,\ q_2)$ it follows from Eq. (5) that $q = q^*$.

Now let us take an approximation to Eq. (5) of the form

$$u^N(t,x,q_1,q_2) = \sum_{i=1}^{N} u_i \exp\left(q_2 - q_1(i\pi)^2 t\right)\sin i\pi x$$

Again $u^N(t^*,\ \ldots,\ q^*) = u^N(t^*,\ \ldots,\ q)$ implies $q = q^*$, provided $N \ge 3$.

Thus q is identifiable at q^* for the original as well as for the approximating model equation. Practically this is useless however, since for only moderately large t^* and i, the contribution of the term $\exp((q_2 - q_1(i\pi)^2 t^*)$ to $u^N(t,\ x,\ q_1,\ q_2)$ is negligible.

Since $\tilde{u}^2 = 0$ we essentially estimate only $(q_2 - q_1\pi^2)t^*$ in an OLS formulation. Thus either q_1 or q_2 separately can be estimated successfully, but a simultaneous estimation of q_1 and q_2 fails.

2.4. Mathematical background for parameter estimation

A typical cost functional in adjoint parameter estimation takes the form (Zou et al., 1992)

$$J(\boldsymbol{X},\boldsymbol{P}) = \int_{t_0}^{t_r} \langle W(\boldsymbol{X} - \boldsymbol{X}^{\text{obs}}),(\boldsymbol{X} - \boldsymbol{X}^{\text{obs}})\rangle \mathrm{d}t + \int_{t_0}^{t_r} K\langle \boldsymbol{P} - \hat{\boldsymbol{P}},\boldsymbol{P} - \hat{\boldsymbol{P}}\rangle \mathrm{d}t \tag{6}$$

where vector $\boldsymbol{P}$, represents model parameters, $\hat{\boldsymbol{P}}$ is the vector of estimated parameters, K are specified weighting matrices, $\boldsymbol{X}^{\text{obs}}$ is the observation vector, $\boldsymbol{X}$ are the model output variables, W is a weighting matrix and, for the more realistic case, there is an interpolation operator H from the model space to the observation space.

The adjoint model equation for

$$\frac{\partial \boldsymbol{X}}{\partial t} = \boldsymbol{F}(\boldsymbol{X}) + K(\boldsymbol{P} - \hat{\boldsymbol{P}}) \tag{7}$$

The adjoint model equation is obtained from a formulation of an augmented Lagrangian, where

$$\nabla_P J = 2K(\boldsymbol{P} - \hat{\boldsymbol{P}}) - \int_{t_0}^{t_r} \langle \boldsymbol{P},(\boldsymbol{X} - \boldsymbol{X}^{\text{obs}}) \rangle \mathrm{d}t \tag{8}$$

and the adjoint model equation is

$$\frac{\partial \boldsymbol{Q}}{\partial t} + \left[\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{X}} \right]^T \boldsymbol{Q} - \boldsymbol{P}^T \boldsymbol{Q} = W(\boldsymbol{X} - \boldsymbol{X}^{\text{obs}}) \tag{9}$$

where $\boldsymbol{Q}$ is a vector of Lagrangian multipliers identified with the adjoint variables, $\boldsymbol{X}$ is the discretized state variable, $\boldsymbol{X}^{\text{obs}}$ is the observation vector, and we see that an additional term, namely:

$$-\boldsymbol{P}^T \boldsymbol{Q}$$

was added to the left hand side of the last equation.

We can assess sensitivity of forecast to model parameters in a simplistic way (i.e. without taking into account presence of data) by considering

$$\delta J = \langle \nabla_P J, \delta \boldsymbol{P} \rangle$$

where $\delta \boldsymbol{P}$ is a small change in parameters vector resulting in a change δJ in forecast errors.

In a general set up the cost function is

$$J(\boldsymbol{X},\boldsymbol{P}) = \tfrac{1}{2}(\boldsymbol{X} - \boldsymbol{X}_{\text{b}})^T B^{-1}(\boldsymbol{X} - \boldsymbol{X}_{\text{b}}) + \tfrac{1}{2}(H\boldsymbol{X} - \boldsymbol{X}^{\text{o}})^T O^{-1}(H\boldsymbol{X} - \boldsymbol{X}^{\text{o}}) + K(\boldsymbol{P} - \hat{\boldsymbol{P}})^T (\boldsymbol{P} - \hat{\boldsymbol{P}}) \tag{10}$$

where B is the background error covariance matrix, $\boldsymbol{X}^{\text{o}}$ is the set of observation whose error covariance is O, and H is the observation operator which computes the model equivalent $H\boldsymbol{X}$ of the observation $\boldsymbol{X}^{\text{o}}$.

2.4.1. *Typical cost functional for parameter estimation*

$$J(p) = J_{\text{h}}(p) + J_{\text{d}}(p) + J_{\text{f}}(p) + J_{\text{r}}(p) \tag{11}$$

$J_{\text{h}}(p)$ weighted least-squares term between 'measured' and model estimated parameters, with weights which are related to confidence in the data. They may be more reliable.

$J_f(p)$ weighted least squares error between 'measured' and model estimated parameters at final time of assimilation—J_f optimizing improvements in data measurements made at later times.

$J_d(p)$ a weighted prior data error term, which represents prior knowledge about the parameters with weights representing the confidence in the measured prior data.

$J_r(p)$ a Tichonov regularization term, which deals with instabilities in values of parameter estimates that appear to be closely related to noise in measured data. This term smooths parameter estimates, by imposing a penalty on oscillations in parameter estimates.

2.4.2. Reduction in the dimensionality of parameters

If the number of distributed parameters to be estimated is large—we can approximate them by constructing (simple) sequences using simple polynomial basis functions according to

$$P^e(X) = \sum_{j=1}^{N_1} \alpha_j X^{j-1} \tag{12}$$

where estimates $P^e(X) \to p(X)$ as $N_1 \to \infty$. Parameter identification is now reduced to identifying a finite number of coefficients α_j such that the cost functional Eq. (11) is minimized.

2.5. Ill-posedness of the problem of parameter estimation

Chavent (1974, 1979, 1983, 1991) was amongst the first to propose parameter estimation using the adjoint method in connection with the output least squares problem. He also recognized that this inverse problem is often ill-posed. This ill-posedness is characterized by

- nonuniqueness
- instability of the identified parameters

Instability here means that small errors in data will cause serious errors in the identified parameters.

The uniqueness problem in parameter identification is intimately related to identifiability.

- The concept of identifiability addresses the question of whether it is at all possible to obtain unique solutions of the inverse problem for unknown parameters of interest in a mathematical model from data collected in the spatial and time domains.
- The parameter estimation problem consists in finding an estimated value $\hat{\boldsymbol{P}}$ of the parameter $\boldsymbol{P}$ from knowledge of data $\boldsymbol{Z}$, of the parameter to output mapping ϕ and some 'a priori' knowledge on the parameter which is condensed in a set C of admissible parameters C_{ad}.

Due to measurement and forecast model errors the equation

$$\text{find } \hat{\boldsymbol{P}} \in C_{ad}, \ \phi(\hat{\boldsymbol{P}}) = \boldsymbol{Z} \tag{13}$$

has no analytic solution so the equation is solved approximatively using a least-squares formulation

$$\text{find } \hat{\boldsymbol{P}} \in C_{\text{ad}}, \text{ s.t.}$$

$$J(\hat{\boldsymbol{P}}) \leq J(\boldsymbol{P}) \ \forall \boldsymbol{P} \in C_{\text{ad}} \tag{14}$$

where $J(\hat{\boldsymbol{P}})$ is the output or measurement error criterion

$$\forall \boldsymbol{P} \in C_{\text{ad}}, \ J(X) = \|\phi(X) - \boldsymbol{Z}\|_F^2 \tag{15}$$

for some norm in data space F and one attempts to minimize Eq. (15) over the set C using his favourite optimization routine.

Here C_{ad} is the admissible parameter set which is chosen as small as possible by taking into account largest possible amount of 'a priori' information on the unknown parameter such as the following considerations: lower and upper bounds, trends and regularity.

In many instances of parameter estimation one finds that (a) instability occurs when the discretization of parameter is refined and (b) lack of uniqueness of the estimated parameter $\hat{\boldsymbol{P}}$, and/or the optimization algorithm gets stuck in a local minimum.

2.6. Identifiability, ill-posedness and regularization of parameter estimation

The key difficulty in developing successful numerical techniques for identifying spatially dependent parameters resides in the fact such problems are ill-posed.

Ill-posedness follows from the fact that the differentiation operator is not continuous with respect to any physically meaningful observation topology.

For instance, the problem of identifying spatially dependent coefficients appearing in the differential operator of a partial differential equation (PDE) is, in general, both nonlinear and ill-posed (Ewing and Lin, 1989).

Using the customary approach of output least-squares Chavent (1991) we minimize a cost functional

$$J_{\text{LS}} = \int_0^T \sum_{i=1}^{M} \left[U(x_i, y_i, \alpha, t) - Z_{\text{d}_i} \right]^2 \text{d}t \tag{16}$$

subject to model equations, initial and boundary conditions, where Z_{d_i} is a set of given observations (measurements) of $U(x_i, y_i, \alpha, t)$ at a set of discrete spatial locations, $i = 1, 2, \ldots, M$.

When the number of parameters is kept small, a well behaved solution results.

However, modeling error is significant, since corresponding subspace of parameter $\delta\alpha$ is too restricted to provide good approximation for an arbitrary α.

As number of parameters is increased, numerical instabilities appear, manifested by spatial oscillation in estimated α, frequency and amplitude of which are not consistent with expected smoothness of α.

A typical symptom is a flat global minimum of J_{LS} (Yakowitz and Duckstein, 1980).

To alleviate problem one may incorporate 'a priori' statistics concerning α by adding a Bayesian term to the cost functional.

Numerical instabilities and ill-posed nature of problem strongly suggest a regularization approach.

'Regularization' of a problem refers generally to solving a related problem, called the 'regularized problem', the solution of which is more regular in a sense than that of the original problem, but which approximates solution of the original problem (Tichonov, 1963, Tichonov and Arsenin, 1977).

When one refers to ill-posed problems, regularization is an approach to circumvent lack of continuous dependence on the data. The regularized problem is a well-posed problem whose solution yields a physically meaningful answer to the given ill-posed problem.

The idea of regularization was first proposed by Tichonov (1963) or earlier, and extended by Morozov (1968) and Morozov and Stessin (1993).

3. Maximum likelihood (ML) method for parameter estimation

In this inverse approach, each of the model parameters to be estimated is represented by a set of discrete 'model parameters'. Let us designate the 'true' values of these model parameters by p, and their prior estimates (or 'measured') values by p^*.

The purpose of the inverse model is to provide improved estimates of p, $\hat{p}$ by relying on model variables measurements—obtained at a set of observation locations at discrete time intervals.

The discrepancies between measured and true quantities such as $p^* - p$ are referred to as 'measurement errors' while differences between measured and computed quantities $p^* - p$ are called 'residuals'.

Maximum likelihood theory is developed in terms of 'prior errors' which are usually taken to be measurement errors. In practice estimation is performed by minimizing a criterion expressed in terms of the residuals which are a combination of measurement errors and errors arising from the numerical model. Thus the prior statistics entering into the estimation criterion should reflect both types of errors.

Since prior errors in the model parameters are affected by a variety of cases we can assume that the prior errors can be considered Gaussian with zero mean. The covariance C_i, the prior errors associated with parameter type p_i can be written as

$$C_i = \sigma_i^2 V_i \tag{17}$$

where σ_i^2 is either a known or unknown positive scalar and V_i is a known symmetric positive-definite matrix.

We assume that the prior estimates of various parameter types are naturally uncorrelated. The global covariance matrix of the model parameters to be estimated C_p is block diagonal, its diagonal components being C.

The prior errors can be written as

$$\epsilon^* = N(0, \sigma_i^2 V_i) \tag{18}$$

meaning ϵ^* is a Gaussian random vector with 0 mean and covariance matrix $\sigma_i^2 V_i$.

Let $Z^* = (\underline{x}^*, \underline{P}^*)$ be a vector incorporating model and parameter data and $\theta = (\sigma_x^2, \sigma_i^2, \ldots)^T$ be a vector of unknown statistical parameters characterizing prior errors. If $\beta = (p, \theta)^T$ is the vector of all the unknown parameters then the likelihood $L(\beta/Z^*)$ of a hypothesis regarding the value of β_1 given Z^* and a specific model structure such as numerical method parameterization etc. (see Carrera and Neuman, 1986a) is proportional to $f(Z^*/\beta)$, the probability density of observing Z^* if β was true.

For a given model structure, the optimum parameter estimates are taken to be those that maximize $L(\hat{\beta}/Z^*)$, where $\hat{\beta}$ is an estimate of β. Based on an hypothesis that all data have been properly transformed to yield Gaussian distributions of the prior errors it becomes obvious that the likelihood function can be written as

$$L(\beta^*/Z^*) = f(Z^*/\beta)$$
$$= (2\pi)^{-1/2} \det|C_x|^{-1/2} \exp\left[\frac{1}{2}(Z^* - Z)^T C_Z^{-1}(Z^* - Z)\right] \quad (19)$$

Here C_x is the covariance matrix of the prior errors

$$C_Z = \begin{bmatrix} C_x & 0 \\ 0 & C_p \end{bmatrix} \quad (20)$$

where C_x and C_p are the covariance matrices of the prior model and model parameters errors $(C_x = \sigma_x^2 V_x)$.

In practice maximum likelihood (ML) estimates are generally obtained by minimizing the 'log-likelihood'

$$S = -2\ln[L(\beta/Z^*)] \quad (21)$$

where we can explicitly write S as

$$S = \log\det C_z + \frac{1}{2}(Z^* - Z)^T C_z^{-1}(Z^* - Z) \quad (22)$$

This criterion has the desirable property that the log-likelihood of a hypothesis, given all the data, is the sum of the log-likelihood of the same hypothesis, given each separate set of data.

This allows one to introduce prior information about the parameters into the estimation scheme and to analyze data representing conditions created by a variety of initial and boundary conditions. An excellent survey of approximation of the ML method to parameter estimation of aquifer parameter is represented by Carrera and Neuman (1986b). Dee (1995) used ML for covariance parameter estimation.

The method was used in oceanography application by ten Brummelhuis and Heemink (1990) and ten Brummelhuis et al. (1993). Wahba (1990) and more recently Wahba et al. (1994, 1995) used more advanced statistical methods such as Gong et al. (1996) generalized the statistical parameter estimation using generalized cross validation (see also Wahba et al., 1995) as well as unbiased risk methods for adaptive tuning of parameters in one space and one time variables for the equivalent barotropic vorticity equation on a latitude circle, simultaneously tuning some smoothing, weighting and physical parameters.

4. Total variation as an L_1 regularization method for parameters with discontinuities

Using total variation (TV) regularization may turn out to be particularly useful to the meteorological community related to the potential of the method to deal with parameter functions which may have jump discontinuities.

This approach bypasses the difficulty of standard regularization techniques that they do not allow discontinuous solutions (Acar and Vogel, 1994).

The method was used by Santosa and Symes (1988) for reconstructing blocky impedance profiles in seismology and for parameter identification by Gutman (1990). Blocky profiles means functions that are piecewise constant and hence have sharply defined edges (Dobson and Santosa, 1994, 1995).

Since 1990 there has been a wide interest in the TV methods for recovering 'blocky', possibly discontinuous images, from noisy data. Significant work includes Rudin and Osher (1994), Rudin et al. (1992), Acar and Vogel (1994), Rudin (1987), Osher and Rudin (1990), Alvarez et al. (1992) and recent work by Dobson and Santosa (1994), Vogel and Oman (1995) and Vogel and Wade (1995), Vogel (1994), Vogel and Oman (1996).

Using total variation methods can indeed result in a significant contribution to problems in optimal parameter estimation where location and size of discontinuities are important.

The topic of total variation regularization was also the subject of two articles in *SIAM News* (December 1993 by Rudin and July 1994 by Santosa) where the work of Rudin et al. (1994) and Osher as well as the work of for denoising and deblurring images was described and illustrated.

Consider solving ill-posed operator equation $AU = Z$. Solving the problem of constrained minimization

$$\min_{U} J(U) \tag{23}$$

Subject to $\|AU - \mathbf{Z}\|^2 = \sigma^2$. This will achieve stability with the requirement that the solution be of bounded variation rather than smooth. Here A is a linear operator from $L^p(\Omega)$ into a Hilbert space Z containing data vector $\mathbf{Z}$.

Data $\mathbf{Z}$ and operator A are inexact.

σ^2 is an estimate of the size of the error in the data and $J(\boldsymbol{U})$ is a bounded variation (BV) norm or semi-norm of $\boldsymbol{U}$.

$$J_\beta(\boldsymbol{U}) = \int_\Omega \sqrt{|\nabla \boldsymbol{U}|^2 + \beta}\, \mathrm{d}x \tag{24}$$

$$\beta \geq 0 \tag{25}$$

$$\|\boldsymbol{U}\|_{\beta V} = \|\boldsymbol{U}\|_{L_1(\Omega)} \tag{26}$$

Another approach is to solve the regularized minimization problem

$$\min_{U} \|AU - \mathbf{Z}\|^2 + \alpha J(\boldsymbol{U}) \tag{27}$$

This can be viewed as a penalty method approach for solving the constrained optimization problem. The penalty parameter $\alpha > 0$ controls trade off between goodness of fit to data measured by $\|AU - Z\|^2$ and variability at jumps in approximate solution as measured by $J(U)$.

The final form of the total variational penalized least-squares functional is

$$\tfrac{1}{2}\|AU - Z\| + \alpha \int_{\Omega} \sqrt{|\nabla U|^2 + \beta} \tag{28}$$

Under mild conditions on the operator A Eq. (24) has a unique minimizer U^* and U^* depends continuously on the parameters α, β and the data Z and on the operator A.

The choice of the penalty parameter α in Eq. (24) depends on factors like signal-to-noise ratio.

When the error level is very low, α tends to be very small. A number of standard minimization methods can be applied for an unconstrained minimization of the modified TV penalty functional.

A nonlinear preconditioned conjugate gradient algorithm will perform well or one can use augmented Lagrangian methods.

The application of total variation problems to meteorological parameter estimation problems will address the issue of parameters which have jump discontinuities—allowing us to bypass difficulty of standard regularization technique that do not allow discontinuous solutions.

5. Parameter estimation by extended Kalman filter

The model system is given by

$$W_k^{\mathrm{f}} = \psi_{k-1}(W_{k-1}^{\mathrm{a}}) \tag{29}$$

W_{k-1}^{a}—best estimation of model state; ψ_{k-1}—state transition function.

$$W_k^{\mathrm{t}} = \psi_{k-1}(W_{k-1}^{\mathrm{t}}) + b_{k-1}^{\mathrm{t}} \tag{30}$$

b_{k-1}^{t} is a Gaussian white noise sequence.

$$Eb_k^{\mathrm{t}}(b_l^{\mathrm{t}})^T = Q_k \delta_{k,l} \tag{31}$$

Q_k—model error covariance.

Evolution of system is measured by observation contaminated by error b_k^{o},

$$W_k^{\mathrm{o}} = h_k(W_k^{\mathrm{t}}) = b_k^{\mathrm{o}} \tag{32}$$

$$Eb_k^{\mathrm{o}}(b_l^{\mathrm{o}})^T = R_k \delta_{k,l} \tag{33}$$

superscripts t, a, f denote truth, analysis and forecast. A linear estimation of the model state based on observation (despite the fact that dynamic model ψ_k and observation h_k may be nonlinear) is sought:

$$W_k^{\mathrm{a}} = W_k^{\mathrm{f}} + K_k(W_k^{\mathrm{o}} - W_k^{\mathrm{f}}) \tag{34}$$

K_k—weight assigned to 'observations'.

Weights are optimized by minimizing the function

$$J_k = E\left[\left(W_k^{\mathrm{a}} - W_k^{\mathrm{t}}\right)^T\left(W_k^{\mathrm{a}} - W_k^{\mathrm{t}}\right)\right] = \mathrm{trace}\left[P_k^{\mathrm{a}}\right] \tag{35}$$

P_k^{a} being the analysis error covariance.

EKF linearly approximates ψ_k and h_k at each step along the trajectory resulting from a continuous model update with 'observations'.

The EKF algorithm when 'observations' are available at time-step k is:

$$W_k^{\mathrm{f}} = \psi_{k-1}\left(W_{k-1}^{\mathrm{a}}\right) \tag{36}$$

$$P_k^{\mathrm{f}} = \Psi_{k-1}\left(W_{k-1}^{\mathrm{a}}\right)P_{k-1}^{\mathrm{a}}\Psi_{k-1}^T\left(W_{k-1}^{\mathrm{a}}\right) + Q_{k-1} \tag{37}$$

Updating step

$$W_k^{\mathrm{a}} = W_k^{\mathrm{f}} + K_k\left(W_k^{\mathrm{o}} - h_k\left(W_k^{\mathrm{f}}\right)\right) \tag{38}$$

$$P_k^{\mathrm{a}} = \left[I - K_k H_k\left(W_{k-1}^{\mathrm{a}}\right)\right]P_k^{\mathrm{f}} \tag{39}$$

$$K_k = P_k^{\mathrm{f}}H_k^T\left(W_k^{\mathrm{f}}\right)\left[H_k\left(W_k^{\mathrm{f}}\right)P_k^{\mathrm{f}}H_k^T\left(W_k^{\mathrm{f}}\right) + R_k\right]^{-1} \tag{40}$$

where

$$\Psi\left(W_k^{\mathrm{a}}\right) = \left.\frac{\partial\psi_k(W)}{\partial W}\right|_{W=W_k^{\mathrm{a}}} \tag{41}$$

$$H_k\left(W_k^{\mathrm{f}}\right) = \left.\frac{\partial h_k(W)}{\partial W}\right|_{W=W_k^{\mathrm{f}}} \tag{42}$$

K_k is called the Kalman gain representing optimal weight given by observations.

Ψ_k is the state transition matrix—a linear propagator of forecast error covariance in time.

Hao (1994), Hao and Ghil (1994) as well as Hao et al. (1996), Hao and Ghil (1995) and Ghil (1996) applied EKF method to estimate model parameters by incorporating the unknown parameter in the scheme such that it is treated as an additional state variable.

$$W_k^{\mathrm{f}} = \psi_{k-1}\left(W_{k-1}^{\mathrm{t}}, \mu_{k-1}^{\mathrm{t}}\right) + b_{k-1}^{\mathrm{t}} \tag{43}$$

$$\mu_k^{\mathrm{f}} = U_{k-1}^{\mathrm{t}} + b_{k-1}^{\mu\mathrm{t}} \tag{44}$$

$$W_k^{\mathrm{o}} = h_k\left(W_k^{\mathrm{t}}\right) + b_k^{\mathrm{o}} \tag{45}$$

where W_k and μ_k can be combined into a composite state vector W_k'.

Issues of identifiability and regularity will require some attention and further research, as well the fact that parameter estimation with EKF is trajectory dependent. The number of observations required for successful optimal parameter estimation is found to be proportional to the size of the errors in the assimilated state fields when the parameter error is left uncorrected, while correct estimation of 'important' model parameters, results in improved estimation of model state. Hao and Ghil (1995) addressed the issue of identifiability and show that the scheme in Eqs. (43)–(45) can be extended to a set of parameters—where μ will serve as a vector.

The dimension of the system increases by the number p of parameters to be estimated, which is usually small, i.e. $p \ll n$. Their research also shows that the number of observations required to estimate a given parameter is proportioned to the sensitivity of the state-estimation errors to the parameter value.

Parameter estimation using Kalman filtering is addressed in a number of textbooks such as Goodwin and Sin (1984, Haykin (1986) and Ljung (1987).

6. Sensitivity analysis as a tool in parameter estimation

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 $VR(X^0, \boldsymbol{\alpha}^0, \boldsymbol{h}_x, \boldsymbol{h}_\alpha)$ of a specific response $R(\boldsymbol{X}, \boldsymbol{\alpha})$

$$R(\boldsymbol{x},\boldsymbol{\alpha}) = \int_{t_0}^{t_a} r(t;\boldsymbol{x},\boldsymbol{\alpha})\mathrm{d}t \tag{46}$$

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

$$VR(\boldsymbol{x}^0,\boldsymbol{\alpha}^0;\boldsymbol{h}_x,\boldsymbol{h}_\alpha) = \int_{t_0}^{t_R} r'_x \cdot \boldsymbol{h}_x \mathrm{d}t + \int_{t_0}^{t_R} r'_\alpha \cdot \boldsymbol{h}_\alpha \mathrm{d}t \tag{47}$$

$$r'_x = \left[\left(\frac{\partial r}{\partial x_1},\ldots,\frac{\partial r}{\partial x_P}\right)\right]_{(x^0 \alpha^0)} \tag{48}$$

$$r'_{\boldsymbol{\alpha}} = \left[\left(\frac{\partial r}{\partial \alpha_1},\ldots,\frac{\partial r}{\partial \alpha_N}\right)\right]_{(x^0,\alpha^0)} \tag{49}$$

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 nth parameter the corresponding variation $h^n_{\boldsymbol{\alpha}}$ of the parameter vector is

$$h^n_{\boldsymbol{\alpha}} = (0,\ldots,h^n_\alpha,\ldots,0)^T \tag{50}$$

and the corresponding sensitivity is VR^n.

The relative sensitivity S_n is the dimensionless quantity

$$S_n = \frac{VR^n}{R}\left(\frac{\boldsymbol{h}^n_\alpha}{\boldsymbol{\alpha}^0_n}\right)^{-1} \tag{51}$$

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 (Zou et al., 1993a) 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 Cacuci (1981, 1988) and Zou et al. (1993a) we can carry out 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 Errico et al., 1993a,b). 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 (Rabier et al., 1992, 1993).

7. Summary and conclusions

Parameter estimation using adjoint of full physics NWP models or oceanographic models will pose serious challenges which were not fully encountered in either meteorology and oceanography when only few parameters were estimated. The issues of identifiability, taken usually for granted in optimal parameter estimation with adjoint method for either atmospheric or ocean general circulation models should be carefully addressed. The mathematical framework for assessing uniqueness and identifiability in optimal parameter estimation using adjoint optimal control methods was outlined. This issue has particular importance since the potential for benefits using variational data assimilation with adjoint methods is highly accrued if one can optimally estimate key parameters in aforesaid oceanic and atmospheric models. Basic research results, using recent advances in issues related to identifiability, regularization stability and total variation regularization were shown to be required to establish robust parameter estimation procedures.

Relative sensitivity analysis will serve as a guide to identify most important parameters whose changes impact most chosen relevant responses for the atmospheric or oceanographic models employed.

The rich experience accumulated by research workers from other fields such as water-resources and seismology as exemplified by the few references provided in this paper can serve as a guide for efforts of reliable optimal parameter estimation in meteorology and oceanography.

Methods of large-scale constrained minimization for the solution of optimal parameter estimation for range-bounded parameters will have to employ sequential quadratic programming and projected gradient minimization methods.

When the optimal parameter estimation is done for a full physics model and his adjoint, as is the case in atmospheric NWP (Numerical Weather Prediction) models, methods of stratification of parameters by physical package, seasonality and other considerations (see Louis and Živković, 1994) should be tested.

Acknowledgements

The author would like to thank the reviewers and the Associate Editor of *Dynamics of Atmospheres and Oceans* for their incisive remarks and constructive criticism which contributed to improvement in the presentation of this paper. This work has been supported by NSF grant ATM-9413050 managed by Dr Pamela Stephens. Additional support was provided by SCRI at Florida State University, which is partially funded by the Department of Energy trough Contract DE-FC0583ER250000.

Appendix A. Rigorous mathematical framework for identifiability and regularization

There has been a serious mathematical effort to understand in depth parameter estimation starting with the seminal works of Richard Bellman and collaborators in 1965 (Bellman et al., 1965a,b) while in the last 20 years works closely related to the use of the adjoint method include contributions of Chavent (1979, Kitamura and Nakarigi (1977, Kubrusly (1977) and more recently works of Banks and Rosen (1987, Banks and Kunish (1989, Omatu and Seinfeld (1989) and Banks (1992a,b) and references therein. It is therefore incumbent upon the meteorological community at both the operational and university level to acquaint itself with these new results.

Let $\mathscr{A}$, U and $\mathscr{F}$ be Banach spaces, where $\mathscr{A}$ represents a space of partial differential operators, U represents the space of solutions and $\mathscr{F}$ the space of right hand sides.

Consider a system represented by

$$\Psi(A,U)=f \tag{A.1}$$

where Ψ is a mapping not necessarily linear from $\mathscr{A}\times\mathsf{U}$ into $\mathscr{F}$.

We assume

(a1) Ψ is of C^k class ($k\geq 1$).

(a2) There is an open subset $\mathscr{A}_c$ of $\mathscr{A}$ and an open subset U_c of U s.t. $\forall A\in\mathscr{A}_c$. Eq. (A.1) admits a unique solution $U\in\mathsf{U}_c$.

(a3) $\forall A\in\mathscr{A}_c$, $\forall u\in\mathsf{U}_c$, $(\partial\psi/\partial u)\ A(u)$ is a linear homeomorphism of U on to $\mathscr{F}$.

Thus one can define an implicit function

$$U=\Phi(A) \tag{A.2}$$

as the solution of Eq. (A.1). Φ is of C^k class from $\mathscr{A}_c$ into U_c. Furthermore, consider that A depends on a set of parameters p belonging to the Banach space $\wedge$. The set of physically admissible parameters p is $\wedge_{ad}$. We assume

(a4) A, $\wedge\rightarrow\mathscr{A}$ is of C^k class ($k\geq 1$).

(a5) $\wedge_{ad}$ is a norm-closed convex subset of $\wedge$.

(a6) $\mathscr{A}(\wedge_{ad})\leq\mathscr{A}_c$.

The parameter identification problem can be posed as follows:

Knowing mappings Ψ, $\mathscr{A}\times\mathsf{U}\rightarrow\mathscr{F}$ and $\mathscr{A}$; $\wedge\rightarrow\mathscr{A}$ and the element $f\in\mathscr{F}$ and given an observation Z_d of U, find $p\in\wedge_{ad}$ to satisfy Eq. (A.1).

Thus the identification problem can be viewed as solving in $\wedge_{ad}$ the nonlinear operator equation

$$(C\cdot\Phi\cdot A)(p)=Z_d \tag{A.3}$$

If this operator $C\cdot\Phi\cdot A$: $\wedge_{ad}\rightarrow\mathscr{H}$ has an unique inverse, and the inverse is continuous, one can apply the least-squares method (adjoint method).

It consists of minimizing over $\wedge_{ad}$ the functional

$$J_{LS}(p)=\|C(\Phi(A(p)))-Z_d\|^2_{\mathscr{H}} \tag{A.4}$$

Often the problem of solving Eq. (A.3) is ill-posed. Hence minima of $J_{LS}(p)$ over $\wedge_{ad}$ will not depend continuously on the data Z_d.

To regularize parameter p, we introduce a more regular space $\mathscr{R}$ for which we assume,

(b1) $\mathscr{R}$ is a Hilbert space.

(b2) $\mathscr{R}$ is densely imbedded in $\wedge$.

(b3) The embedding operator from $\mathscr{R}$ into $\wedge$ is compact.

Define $\mathscr{R}_{ad}=\mathscr{R}\cap\wedge_{ad}$. It follows that $\mathscr{R}_{ad}$ is a norm-closed convex subset of $\mathscr{R}$. Introduce now stabilizing functional (regularization).

$$J_s(p)=\|p\|^2_{\mathscr{R}},\quad p\in R_{ad} \tag{A.5}$$

and smoothing functional

$$J_\beta(p) = J_{LS}(p) + \beta J_s(p) = \|C(\Phi(A(p))) - Z_d\|^2_{\mathscr{H}} + \beta \|p\|^2_{\mathscr{P}}, \quad p \in \mathscr{P}_{ad} \tag{A.6}$$

Identification of parameters by regularization proceeds as follows (Nakarigi, 1983, Kravaris and Seinfeld, 1986).

Given observation $Z_d \in \mathscr{H}$ find $p_\beta \in \mathscr{P}_{ad}$ so as to minimize cost functional $J_\beta(p)$.

Upon modification of the algorithms by regularization and by using appropriate constraints, the estimated parameters improve significantly.

When 'a priori' knowledge about constraints for certain problem is available a constrained optimization algorithm may be an efficient choice for solving the parameter estimation problem (Chavent, 1991, Kunish and Sachs, 1992) otherwise the use of penalty-regularization terms in the cost function may be preferable.

Several key issues need to be examined:

1. Whether Eq. (A.3) admits a unique solution (identifiability). This means that there is a chance for the optimally estimated parameter to be close to the true unknown parameter.
2. Whether solution of Eq. (A.3) depends continuously on the observations Z_d (stability). The stability of the estimated parameter is with respect to perturbations on the observations. (See Chavent, 1979, Kravaris and Seinfeld, 1985, 1986).
3. Relationship between dimensionality of state space, observation space and parameter space.

References

Acar, R., Vogel, C.R., 1994. Analysis of bounded variation penalty method for ill-posed problems. Inverse Problems 10 (6), 1217–1229.

Alvarez, L., Lions, P.L., Morel, J.-M., 1992. Image selective smoothing and edge detection by nonlinear diffusion. SIAM J. Numer. Anal. 29 (3), 845–866.

Banks, H.T., 1992a. Control and estimation in distributed parameter systems. In: Frontiers in Applied Mathematics, vol. 11. SIAM, Philadelphia, 227 pp. edited by H.T. Banks.

Banks, H.T., 1992b. Computational issues in parameter estimation and feedback control problems for partial differential equation systems. Physica D 60, 226–238.

Banks, H.T., Kunish, K., 1989. Estimation Techniques for Distributed Parameter Systems. Birkhauser, Boston, 315 pp.

Banks, H.T., Rosen, I.G., 1987. Numerical schemes for the estimation of functional parameters in distributed models for mixing mechanisms in lake and sea sediment cores. Inverse Problems 3, 1–23.

Bellman, R., Jaquez, J., Kalaba, R., Schwimmer, S., 1965a. Quasi-linearization and the estimation of chemical rate constants from raw kinetic data. The Rand Corporation, RM-4721-NIH.

Bellman, R., Kagiwada, H., Kalaba, R., Ueno, S., 1965b. Inverse problems in radiative transfer: layered media. Icarus 4, 119.

Bennett, A.F., 1992. Inverse Methods in Physical Oceanography. Cambridge University Press, Cambridge, 346 pp.

Bennett, A.F., McIntosh, P.C., 1982. Open ocean modeling as an inverse problem: tidal theory. J. Phys. Occeanogr. 12, 1004–1018.

Cacuci, D.G., 1981. Sensitivity theory for nonlinear systems: II. Extensions to additional classes of responses. J. Math. Phys. 22, 2803–2812.

Cacuci, D.G., 1988. The forward and adjoint methods of sensitivity analysis. In: Ronen, Y. (Ed.), Uncertainty Analysis. CRC Press, Inc., 282 pp.
Carrera, J., Neuman, S.P., 1986a. Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. Water Resour. Res. 22 (2), 199–210.
Carrera, J., Neuman, S.P., 1986b. Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability and solution algorithms. Water Resour. Res. 22 (2), 211–227.
Carrera, J., Neuman, S.P., 1986c. Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data. Water Resour. Res. 22 (2), 228–242.
Chavent, G., 1974. Identification of functional parameters in partial differential equations. In: Goodson R.E., Polis, M. (Eds.), Identification of Parameters in Distributed Systems, ASME Conference on Automatic Control, 17–21 June 1974, Austin, TX. pp. 31–48. American Society of Mechanical Engineers, New York, NY 10017.
Chavent, G., 1979. 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. Pergamon Press, Oxford, pp. 85–97.
Chavent, G., 1983. Local stability of the output least square parameter estimation technique. Mat. Appl. Comp. 2, 3–22.
Chavent, G., 1991. On the theory and practice of non-linear least-squares. Adv. Water Resour. 14 (2), 55–63.
Chavent, G., Lemonnier, P., 1974. Identification de la non-linearite d'une equation parabolique quasi-lineaire. Appl. Math. Opt. 1, 121–162.
Chavent, G., Dupuy, M., Lemonnier, F., 1975. History matching by use of optimal theory. Soc. Pet. Eng. J. 15, 74–86.
Chertok, D.L., Lardner, R.W., 1996. Variational data assimilation for a nonlinear hydraulic model. Appl. Math. Modeling 20, 675–682.
Courtier, P., 1986. Le modele adjoint, outil pour des experiences de sensibility. Note de travail no 166, EERM, Paris, France, 36 pp. (available from Meteo-France).
Courtier, P., 1987. Application du contrôle optimal á le prévision numerique en Météorologie. These de Doctorat de l'Universite Paris 6.
Cushman-Roisin, M., 1986. The hydrological inverse problem: reconsideration and application to the micro-compute. Ph.D. Dissertation, Florida State University, Tallahassee, FL, USA.
Daley, R., 1991. Atmospheric Data Analysis. Cambridge University Press, 497 pp.
Das, S.K., Lardner, R.W., 1990. On the estimation of parameters of hydraulic models by assimilation of periodic tidal data. J. Geophys. Res. (Oceans) 96, 15 187–15 196.
Das, S.K., Lardner, R.W., 1992. Variational parameter estimation for a two dimensional numerical tidal model. Int. J. Numer. Meth. Fluids 15, 313–327.
Dee, D.P., 1995. On-line estimation of error covariance parameter for atmospheric data assimilation. Mon. Weather Rev. 123, 1128–1145.
Dobson, D.C., Santosa, F., 1994. An image enhancement technique for electrical impedance tomography. Inverse Problems 10, 317–334.
Dobson, D.C., Santosa, F., 1995. Recovery of blocky images from noisy and blurred data. SIAM J. Appl. Math. Vol. 56 (4), 1996, p. 1181–1198.
Errico, R.M., Vukićević, T., Raeder, K., 1993a. Examination of the accuracy of a tangent linear model. Tellus 45A, 462–477.
Errico, R.M., Vukićević, T., Raeder, K., 1993b. Comparison of initial and lateral boundary condition sensitivity for a limited area model. Tellus 45A, 539–557.
Ewing, R.E., Lin, T., 1989. Parameter identification problems in single phase and two-phase flow. In: F. Kappel, W. Kunish and W. Shappacher (Eds.), Control and Estimation of Distributed Parameter Systems. Birkhauser Verlag, Boston, 434 pp. Full Series On Numerical Maturities, Vol. 91.
Fukumori, I., Benveniste, J., Wunsch, C., Haidvogel, D.B., 1993. Assimilation of sea surface topography into an ocean circulation model using a steady state smoother. J. Phys. Ocean. 23, 1831–1855.
Ghil, M., 1996. Advances in sequential estimation for atmospheric and oceanic flows. J. Meteor. Jpn. Vol. 75, No. 1B, pp. 289–304, (Special issue "Data Assimilation in Meteorology and Oceanography: Theory and Practice, Ghil, M., Ike, K., H. Bennett, P. Courtier, M. Kimoto, H. Nagata, M. Saiki and N. Sato (Eds.).

Ghil, M., Malanotte-Rizzoli, P., 1991. Data assimilation in meteorology and oceanography. In: Saltzman, B. (Ed.), Advances in Geophysics, vol. 33. Academic Press, pp. 141–266.
Gong, J., Wahba, G., Johnson, D.R., Tribbia, J., 1996. Adaptive tuning of numerical weather prediction models: simultaneous estimation of weighting, smoothing and physical parameters. Mon. Weather Rev. (submitted).
Goodson, R.E., Polis, M.P., 1979. Parameter identification in distributed systems: a synthesizing overview in identification of parameters in distributed systems. In: Symposium of Annual Aut. Control. Am. Soc. Mech. Eng., New York, pp. 1–31.
Goodwin, G.C., Sin, K.S., 1984. Adaptive Filtering Prediction and Control. Prentice-Hall, Englewood Cliffs, NJ, 540 pp.
Gunson, J.G., 1995. Estimating open ocean boundary conditions: sensitivity studies. Doctoral Thesis, MIT and Woods Hole Oceanographic Institution, Joint Program in Physical Oceanography, 149 pp.
Gutman, S., 1990. Identification of discontinuous parameters in flow equations. SIAM J. Control Optim. 28, 1049–1060.
Haidvogel, D.B., Robinson, A.R. (Eds.), 1989. Data Assimilation, Special Issue. Dyn. Atmos. Oceans 13 (3–4).
Hao, Z., 1994. Data assimilation for interannual climate-change prediction. Ph.D. Thesis, Dept of Atmospheric Sciences, UCLA, 224 pp.
Hao, Z., Ghil, M., 1994. Data assimilation in a simple tropical ocean model with wind stress errors. J. Phys. Oceanogr. 24 (10), 2111–2128.
Hao, Z., Ghil, M., 1995. Sequential parameter estimation for a coupled ocean–atmosphere model. In: WMO/TD No. 651, Second Int. Symposium on Assimilation of Observations in Meteorology and Oceanography, 13–17 March 1995, Tokyo, Japan, vol. 1. pp. 181–186. World Meteorological Organization, Geneva, Switzerland.
Hao, Z., Ghil, M., Neelin, J.D., 1996. Data assimilation for a coupled ocean atmosphere model, Part I: Sequential state estimation. J. Phys. Oceanogr. (submitted).
Haykin, S., 1986. Adaptive Filter Theory. Prentice-Hall, Englewood Cliffs, NJ, 590 pp.
Heemink, A.W., 1987. Two-dimensional shallow water flow identification. Appl. Math. Model. 12, 109–118.
Hogg, N.G., 1996: Oceanographic data for parameter estimation. In Modern Approaches to data Association in Ocean Modeling. Vol. 61. Elsevier Oceanography Series 53–77, 455 pp.
Kitamura, S., Nakarigi, S., 1977. Identifiability of spatially-varying and constant parameters in distributed systems of parabolic type. SIAM J. Control Optim. 15, 785–802.
Kravaris, C., Seinfeld, J.H., 1985. Identification of parameters in distributed parameters systems by regularization. SIAM J. Control Optim. 23, 217–241.
Kravaris, C., Seinfeld, J.H., 1986. Identification of spatially varying parameters in distributed parameters systems by discrete regularization. J. Math. Anal. Appl. 119, 128–152.
Kubrusly, C.S., 1977. Distributed parameter system identification: a survey. Int. J. Control 26 (4), 509–535.
Kunish, K., Sachs, E.W., 1992. Reduced SQP methods for parameter estimation problems. SIAM J. Numer Anal. 29 (6), 1793–1820.
Lardner, R.W., 1993. Optimal control of open boundary conditions for a numerical tidal model. Comput. Methods Appl. Mech. Eng. 102 (3), 367–387.
Lardner, R.W., Das, K.S., 1994. Optimal estimation of eddy viscosity for a quasi-three dimensional numerical tidal and storm surge model. Int. J. Num. Methods Fluids 18, 295–312.
Lardner, R.W., Song, Y., 1995. Optimal estimation of eddy viscosity and friction coefficients for a quasi-3-dimensional numerical tidal model. Atmos.-Ocean 33, 581–611.
Lardner, R.W., Al-Rabeh, A.H., Gunay, N., 1993. Optimal estimation of parameters in a two dimensional hydrodynamical model of the Arabian Gulf. JGR Oceans 98, C10, 18, 229–18, 242.
Le Dimet, F.X., Navon, I.M., 1988. Variational and optimization methods in meteorology: a review. Technical Report SCRI 144, Supercomputer Computations Research Institute, Florida State University, Tallahassee, FL, 83 pp.
Ljung, L., 1987. System Identification. Theory for the User. Prentice-Hall, Englewood Cliffs, NJ, 519 pp.
Louis, J.-F., Živković, M., 1994. Optimizing a local weather forecast system. AER, Inc., Final Report, Cambridge, MA.

Malanotte-Rizzoli, P. (Ed.), 1996. Modern Approaches to Data Assimilation in Ocean Modeling, vol. 61. Elsevier Oceanography Series, 455 pp.

Malanotte-Rizzoli and E. Tziperman, 1996. The Oceanographic data Assimilation Problem: Overview, Motivation and Purposes. In Modern Approaches to data assimilation in Ocean Modeling. Vol. 61. Elsevier Oceanography Series. 3–17, 455 pp.

Marotzke, J., 1992. The role of integration time in determining a steady state through data assimilation. J. Phys. Oceanogr. 22, 1556–1567.

Marotzke, J., Wunsch, C., 1993. Finding the steady state of a general circulation model trough data assimilation. J. Geophys. Res. 98, 20149–20167.

Menke, W., 1984. Geophysical Data Analysis: Discrete Inverse Theory. Academic Press, 289 pp.

Miller, R.N., 1987. Theory and practice of data assimilation for oceanography. Harvard Tech. Rep. 26, 59 pp.

Morozov, V.A., 1968. The error principle in the solution of operator equations by the regularization method. USSR Comput. Math. Math. Phys. 892, 63–87.

Morozov, V.A., Stessin, M., 1993. Regularization Methods for Ill-Posed Problems. CRC Press, Boca Raton, FL, 257 pp.

Nakarigi, S., 1983. Identifiability of linear systems in Hilbert spaces. SIAM J. Control Optim. 21, 501–530.

Nash, S.G., Sofer, A., 1996. Linear and Nonlinear Programming. McGraw-Hill Companies Inc., 692 pp.

Navon, I.M., Legler, D.M., 1987. Conjugate-gradient methods for large-scale minimization in meteorology. Mon. Weather Rev. 115, 1479–1502.

Navon, I.M., Zou, X., Berger, M., Phua, P.K.H., Schlick, T., Le Dimet, F.X., 1992a. Numerical experience with limited memory quasi-Newton and truncated Newton methods. In: K.H. Phua, C.M. Wang, W.Y. Yeong, T.Y.Y. Leong (Eds.), Optimization Techniques and Applications, vol. 1. World Scientific Publishing Co., 1232 pp.

Navon, I.M., Zou, X., Berger, M., Phua, P.K.H., Schlick, T., Le Dimet, F.X., 1992b. Testing for reliability and robustness of optimization codes for large scale optimization problems. In: K.H. Phua, H.T. Loh, K.C. Tau and F.S. Chou (Eds.), Optimization Techniques and Applications, vol. 1. World Scientific Publishing Co., 1232 pp.

Nocedal, J., 1991. Theory of algorithms for unconstrained minimization. Acta Numerica 1, 199–247.

Omatu, S., Seinfeld, J.H., 1989. Distributed Parameter Systems: Theory and Applications. Oxford Mathematical Monographs, Oxford University Press, 430 pp.

Osher, S., Rudin, L.I., 1990. Feature oriented image enhancement using shock filters. SIAM J. Numer. Anal. 27 (4), 919–940.

Panchang, V.G., O'Brien, J.J., 1989. On the determination of hydraulic model parameters using the strong constraint formulation. In: Davies, A.M. (Ed.), Modeling Marine Systems, I. CRC Press, Inc., pp. 5–18.

Prevost, C., Salmon, R., 1986. A variational method for inverting hydrographic data. J. Mar. Sci. 44, 1–34.

Rabier, F., Courtier, P., Talagrand, O., 1992. An application of adjoint models to sensitivity. Beitr. Phys. Atmos. 65 (3), 177–192.

Rabier, F., Courtier, P., Hervou, M., Strauss, B., Persson, A., 1993. Sensitivity of forecast error to initial conditions using the adjoint model. ECMWF Tech. Memo. 197, 29 pp.

Richardson, J.E., Panchang, V.G., 1992. A modified adjoint method for inverse eddy viscosity estimation in coastal circulation models. In: Malcolm L. Spaulding, Keith Bedford, Alan Blumberg, Ralph Cheng and Craig Swanson (Eds.), Estuarine and Coastal Modeling, Proc. 2nd Int. Conf. Am. Soc. Civil Eng., New York, pp. 733–745.

Rudin, L.I., 1987. Images, numerical analysis of singularities and shock filters. Ph.D. Thesis, Computer Sci. Dept. Caltech, Pasadena, CA, Tech. Rep 5250: TR-87.

Rudin, L.I., Osher, S., 1994. Total variation based image restoration with free local constraints. In: IEEE International Conference on Image Processing, vol. 1. IEEE Signal Processing Society (Eds.), Los Alamitos, CA, USA, pp. 31–35.

Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithm. Physica D. 60, 259–268.

Rudin, L.I., Osher, S., Fu, C., 1994. Total variation based restoration of noisy blurred images. SIAM J. Numer. Anal. (in press).

Santosa, F., Symes, W., 1988. Reconstruction of blocky impedance profiles from normal incidence reflection seismograms which are band-limited and miscalibrated. Wave Motion 10, 209–230.

Seinfeld, J.H., Kravaris, C., 1982. Distributed parameter identification in geophysics–petroleum reservoirs and aquifers. In: Tzafestas, S.G. (Ed.), Distributed Parameter Control Systems. Pergamon Press, Oxford, pp. 367–390.

Smedstad, O.M., O'Brien, J.J., 1991. Variational data assimilation and parameter estimation in an equatorial Pacific Ocean model. Progr. Oceanogr. 26, 179–241.

Stauffer, D.R., Bao, J., 1993. Optimal determination of nudging coefficients using the adjoint equations. Tellus 45A, 358–369.

Sun, N.-Z., Yeh, W.W.-G., 1990a. Coupled inverse problem in groundwater modeling, 1, Sensitivity and parameter identification. Water Resour. Res. 26 (10), 2507–2525.

Sun, N.-Z., Yeh, W.W.-G., 1990b. Coupled inverse problem in groundwater modeling, 2, Identifiability and experimental design. Water Resour. Res. 26 (10), 2527–2540.

Sun, N.-Z., Yeh, W.W.-G., 1992. A stochastic inverse solution for transient groundwater flow: parameter identification and reliability analysis. Water Resour. Res. 28, 3269–3280.

Tarantola, A., 1987. Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation. Elsevier Science, 613 pp.

ten Brummelhuis, P.C.J., Heemink, A.W., 1990. Parameter identification in tidal models with uncertain boundary conditions. Stochastic Hydrol. Hydraul. 4, 193–208.

ten Brummelhuis, P.C.J., Heemink, A.W., van den Boogaard, H.P.F., 1993. Identification of shallow-sea models. Int. J. Numer. Meth. Fluids 17, 637–665.

Tichonov, A.N., 1963. Regularization of ill-posed problems. Dokd. Akad. Nauk, USSR 153, 49–52.

Tichonov, A.N., Arsenin, V.Y., 1977. Solutions of Ill-Posed Problems. Halsted Press, New York, 258 pp.

Tziperman, E., Thacker, W.C., 1989. An optimal control/adjoint equations approach to studying the oceanic general circulation. J. Phys. Oceanogr. 19, 1471–1485.

Tziperman, E., Thacker, W.C., Long, R.B., Hwang, S.-M., 1992. Oceanic data analysis using a general circulation model. Part 1: Simulations. J. Phys. Ocean. 22, 1434–1485.

Vogel, C.R., 1994. Non-convergence of the L-curve regularization parameter selection method. Inverse Problems, Vol. 12, 1996, 535–547.

Vogel, C.R., Oman, M.E., 1995. Fast numerical methods for total variation minimization in image reconstruction. In: SPIE Proceedings Vol. 2563, "Advanced Signal Processing Algorithms", July 1995. Society of Photo Optical Instrumentation Engineers, Franklin Luk (Ed.).

Vogel, C.R., Oman, M.E., 1996. Iterative methods for total variation denoising. SIAM J. Sci. Comput. 17 (1), 227–238.

Vogel, C.R., Wade, J.G., 1995. Analysis of costate discretizations in parameter-estimation for linear evolution-equations. SIAM J. Control Optim. 33 (1), 227–254.

Wahba, G., 1990. Spline models for observational data. In: CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 59. Society for Industrial and Applied Mathematics, 169 pp., SIAM eds., Philadelphia, PA.

Wahba, G., Johnson, D.R., Gao, F., Gong, J., 1994. Adaptive tuning of numerical weather prediction models: Part I: randomized GCV and related methods in three and four dimensional data assimilation. Tech. Rep. 920, Dept of Statistics, University of Wisconsin, Madison, WI, 34 pp.

Wahba, G., Johnson, D.R., Gao, F., Gong, J., 1995. Adaptive tuning of numerical weather prediction models, randomized GCV in three and four dimensional data assimilation. Mon. Weather Rev. 123, 3358–3369.

Wang, Z., 1993. Variational data assimilation with 2-D shallow water equations and 3-D FSU global spectral models. Ph.D. Thesis, Dept. of Mathematics, Florida State University, Tallahassee, FL, 235 pp.

Wang, Z., Navon, I.M., Zou, X., Le Dimet, F.X., 1995. A truncated Newton optimization algorithm in meteorology applications with analytic Hessian/vector products. Comput. Optim. Appl. 4, 241–262.

Wergen, W., 1992. The effect of model errors in variational assimilation. Tellus 44A, 297–313.

Wunsch, C., 1996. The Ocean Circulation Inverse Problem. Cambridge University Press, 442 pp.

Yakowitz, S., Duckstein, L., 1980. Instability in aquifer identification: theory and case studies. Water Resour. Res. 16, 1045–1064.

Yeh, W.W.-G., 1986. Review of parameter estimation procedures in groundwater hydrology. Water Res. Rev. 22, 95–108.

Yu, L., O'Brien, J.J., 1991. Variational estimation of the wind stress drag coefficient and the oceanic eddy viscosity profile. J. Phys. Oceanogr. 21, 709–719.

Yu, L., O'Brien, J.J., 1992. On the initial condition in parameter estimation. J. Phys. Oceangr. 22, 1361–1364.

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

Zou, X., Navon, I.M., Barcilon, A., Whittaker, J., Cacuci, D., 1993a. An adjoint sensitivity study of blocking in a two-layer isentropic model. Mon. Weather Rev. 121 (10), 2833–2857.

Zou, X., Navon, I.M., Berger, M., Phua, M.K., Schlick, T., Le Dimet, F.X., 1993b. Numerical experience with limited-memory, quasi-Newton methods for large-scale un-constrained nonlinear minimization. SIAM J. Optim. 3 (3), 582–608.