Data assimilation for Numerical Weather Prediction : a review

I. Michael Navon

School of Computational Science and Department of Mathematics, The Florida State University, Tallahassee, Florida 32306-4120

Abstract

During the last 20 years data assimilation has gradually reached a mature center stage position at both Numerical Weather Prediction centers as well as being at the center of activities at many federal research institutes as well as at many universities.

The research encompasses now activities which involve, beside meteorologists and oceanographers at operational centers or federal research facilities, many in the applied and computational mathematical research communities. Data assimilation or 4-D VAR extends now also to other geosciences fields such as hydrology and geology and results in the publication of an ever increasing number of books and monographs related to the topic.

In this short survey article we provide a brief introduction providing some historical perspective and background, a survey of data assimilation prior to 4-D VAR and basic concepts of data assimilation.

I first proceed to outline the early 4-D VAR stages (1980-1990) and address in a succinct manner the period of the 1990's that saw the major developments and the flourishing of all aspects of 4-D VAR both at operational centers and at research Universities and Federal Laboratories. Computational aspects of 4-D Var data assimilation addressing optimization methods, parameter estimation and computational burdens as well as ways to alleviate them are briefly outlined

Brief interludes are provided for each period surveyed allowing the reader to have a better perspective A brief survey of different topics related to state of the art 4-D Var today is then presented and we conclude with what we perceive to be main directions of research and the future of data assimilation and some open problems. We will strive to use the unified notation of Ide et al.[167]

Key words: 4-D VAR data assimilation, 3-D VAR, parameter estimation, minimization methods
PACS:

1 Introduction

Data assimilation in atmospheric sciences started from the fact that NWP is an initial value problem. This since we start at whatever constitutes the present state and use the NWP model to forecast its evolution. Early works by Richardson [310] and Charney [55] were based on hand interpolations [176]. This in order to combine present and past observations of the state of the atmosphere with results from the model (also referred to as "mathematical" model). Since this was a rather tedious procedure efforts to obtain "automatic" objective analysis-the first methods have been developed by Panofsky [286], Gilchrist and Cressman [142] , Cressman [76] , Barnes [16,17]. Use of prior information to supplement rather insufficient data was pioneered by Bergthorsson and Döös[26] , Cressman [76] followed by the comprehensive work of Lev Gandin [122].

Early reviews of data assimilation whose purpose is that of" using all available information (data) to determine as accurately as possible the state of the atmospheric (or oceanic) flow" (Talagrand)[350] were provided by Le Dimet and Navon[203], an in-depth survey of Ghil and Malanotte-Rizzoli[137] as well as by the outstanding book of Daley " Atmospheric Data Analysis"[81].

A collection of papers in "Data Assimilation in Meteorology and Oceanography: Theory and Practice[138] summarizes state of the art of data assimilation for that period. See also a short survey by Zupanski and Kalnay[407] along with the excellent book of Kalnay[176] "Atmospheric Modeling, Data Assimilation and Predictability". An early effort linking Optimal Interpolation (O.I.) with the variational method was done by Sasaki [315] [316] and in more final form by Sasaki [317–321] which can be viewed as a 3-D VAR approach. It was Lorenc[222] that showed that OI and 3-D VAR were equivalent provided the cost functional assumes the form:

$$J = \frac{1}{2}\{[y^o - H(x)]^T R^{-1}[y^o - H(x)] + (x - x^b)^T B^{-1}(x - x^b)\} \tag{1}$$

The first term measures the distance of forecast field x to observations y^o and the second term measures the distance to background x^b.

The analysis x is obtained by adding the innovation to the model forecast with weights W based on estimated statistical error covariances of forecast and observations.

$$x = x^b + W[y^o - H(x^b)] \tag{2}$$

Related theoretical material related to the set-up that led to modern data as-

similation may be found in the " Inverse Problem Theory" of Tarantola[351], the optimal control book of Lions[219], the " Perspectives in Flow Control and Optimization" by Max Gunzburger[153] along with "Inverse Modeling of the Ocean and Atmosphere" by Andrew Bennett [23] and "Dynamic Data Assimilation: A Least Squares Approach" by John Lewis et al.[213] ,Cacuci(2003)[46] and Cacuci et al.(2005) [48].

In this brief review we first provide some historical background to the data assimilation effort along with some basic concepts of data assimilation. We then proceed to survey in a section the early stages (1980-1990) of 4-D VAR data assimilation with brief interludes summarizing and providing perspectives as we go along . In the following section we address some computational aspects of data assimilation such as issues of automatic differentiation, suitable and robust large-scale unconstrained minimization algorithm and parameter estimation applications along with the incremental method which alleviated the computational burden of 4-D VAR and made it operationally viable at large operational NWP centers. A short section is dedicated to state-of the art of data assimilation at present time and we close with a short section outlining open problems and directions of development of 4-D VAR in the future.

Relationship between OI and 3-D VAR

The terminology of 4-D VAR (4-dimensional data assimilation) was originally used in research centers in the context of using continuous data assimilation satellite data leading to the First Global Atmosphere Research Program (GARP) Global Experiment [262,263], Charney, Halem and Jastrow [56].

Insertion of observations directly into primitive equations models excited spurious inertia-gravity oscillations in the model and required the use of damping schemes [240] for damping the high-frequency components. A full-account of these techniques and the history of continuous data assimilation is provided in the seminal book of Daley [81] . This review will survey some aspects of variational data assimilation while only providing a brief outline of methodologies that prevailed prior to the 1980's. We will rely on work of Kalnay[176], Daley[81], Talagrand[350] , Zupanski and Kalnay[407] , Ghil et al. (Eds)[138], works of the present author and his collaborators, the review of Ghil and Malanotte-Rizzoli[137] and an early review that remained an unpublished technical report [203].

Panofsky[286] is credited for pioneering the first objective analysis based on 2-D polynomial interpolation. It was followed by Gilchrist and Cressman[142] who put forward an interpolation scheme for geopotential field as a quadratic polynomial in x and y

$$E(x, y) = a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2, \tag{3}$$

then minimizing mean square difference between polynomial and observations within a radius of influence of the closest grid point,

$$\min_{a_{ij}} E = \min_{a_{ij}} \{ \sum_{k=1}^{K_v} p_v \cdot (E_0^v - E(x_v, y_v))^2 + \sum_{k=1}^{K_v} q_v \cdot \{ [u_v^0 - u_g(x_v, y_v)]^2 + [v_v^0 - v_g(x_v, y_v)]^2 \} \} \tag{4}$$

where p_v and q_v were empirical weights and u_g and v_g the components of the geostrophic wind obtained from the gradient of geopotential height $E(x, y)$ at observation point k. K was total number of observations within the radius of influence. The introduction of first guess estimate is credited to have been introduced by Bergthorsson and Döös[26]. Usually either climatology or a combination of it with first guess was used in the analysis cycle. See also the influential work of Gandin[122], translated from Russian by the Israeli program of Translations in 1965.

2 Successive correction method

The first analysis method in 4DDA was the successive correction method developed by Bergthorsson and Döös[26] and by Cressman[76]. The field of background was chosen as a blend of forecast and climatology with a first estimate given by the first guess field

$$f_i^0 = f_i^b. \tag{5}$$

f_i^b background field estimated at the i-th grid point, f_i^0 being the zeroth iteration estimate of gridded field. This is hence followed by new iteration obtained by "successive corrections"

$$f_i^{n+1} = f_i^n + \sum_{k=1}^{K_i^n} w_{ij}^n (f_k^0 - f_k^n) + \sum_{k=1}^{K_i^n} w_{ik}^n + \epsilon^2 \tag{6}$$

f_i^n- n-th iteration estimate at ith grid point,

f_k^0- k-th observation surrounding grid point,

f_i^n- value of n-th field estimate calculated at observation point k derived by interpolation from nearest grid points,

ϵ^2- estimate of ratio of observation error variance to background error variance.

The important ingredient is constituted by the weights w_{ik}^n which are related to a radius of influence. Cressman[76]proposed the following weights in the SCM (Successive corrections method).

$$w_{ik}^n = \frac{R_n^2 - r_{ik}^2}{R_n^2 + r_{ik}^2} \quad \text{if } r_{ik}^2 \leq R_n^2 \tag{7}$$

$$w_{ik}^n = 0 \text{ if } r_{ik}^2 > R_n^2 \tag{8}$$

r_{ik}^2 square of distance between observation point r_k and a grid point at r_i.

The controlling parameter is the radius of influence R_n, allowed to vary between iterations while K_i^n is the number of observations within a radius of R_n of the grid point i. If one reduces the radius of influence, this results in a field reflecting large scales after first iteration -and tends towards smaller scales after additional iterations. For additional technical details see Daley[81], Kalnay[176].

Cressman[76]took the coefficient ϵ^2 to be zero. For noisy data with errors it may lead to erroneous analysis. Taking $\epsilon^2 > 0$ i.e. assuming observations with errors, allows some impact to the background field. Barnes[16] defined the weights to follow a Gaussian or normal distribution

$$w_{ij} = \begin{cases} exp-(\frac{r_{ik}^2}{d^2}) & \text{if } r_{ik} \leq d \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where d is the radius of influence.

It uses an adaptive version where the radius of influence changes by a factor γ

$$0 < \gamma < 1. \tag{10}$$

It was shown by Bratseth[37] that with an appropriate choice of weights these SCM iterative method analysis increments can be made to be the same as those obtained using optimal interpolation (OI). Lewis et al.[213] quote also similar independent work done by Franke and Gordon[120], Franke[121] and Seaman[328].

3 The nudging method

The nudging method also referred to as Newtonian relaxation belongs to the class of methods called continuous or dynamic 4DDA. Data are introduced into the assimilation system (4-dimensional data assimilation) at each time step of model integration during the assimilation time period. This empirical method is called Newtonian relaxation [177,163]. An illustrative example of nudging is [176] one where a forcing term is added to the x-momentum equation whereby in the preforecast period the model variables are driven towards the observation by the addition of extra forcing terms in the equations.

This goes on as long as the actual initial time has not yet been reached.

$$\frac{\partial u}{\partial t} = -v\nabla u + fv - \frac{1}{\rho}\frac{\partial p}{\partial x} + \frac{u - u_{obs}}{\tau}, \tag{11}$$

where in the forcing term, τ (depends) is the time scale of the relaxation and depends on the variable and is chosen empirically. If τ is too small, the solution will converge fast towards the observations. In general τ should be chosen such that the forcing term is similar in size to the less dominant terms to avoid rebounding effect that slows down assimilation process but large enough to impact the assimilation. For additional work on nudging methods see also Ramamurthy and Carr[306]. Zou et al.[398] used adjoint optimal parameter estimation to estimate best nudging time scale.

The nudging method is also used in meso-scale models to assimilate small-scale observations in the absence of statistical data. The method has also been used on global scale by Lyne et al.[231].

The variational calculus approach

It was introduced in meteorology by Yoshi Sasaki in his Ph.D Thesis[315] and later extended by him to include dynamic model laws [317–320]. He proposed three basic types of variational formalism in the numerical variational analysis method . The basic formalisms are categorized into three areas: (1)"timewise localized" formalism, (2) formalism with strong constraint, and (3)a formalism with weak constraint. Exact satisfaction of selected prognostic equations were formulated as constraints in the functionals for the first two formalisms. This approach is now generically referred to as 3-D VAR.

In 3-D VAR one defines a cost function proportional to the square of the distance between analysis and both background and observations, and it was showed by Lorenc [221], [222] that the OI and the 3-D VAR approaches are

equivalent provided the cost function is defined as

$$J = \frac{1}{2}[y^o - H(x)]^T R^{-1}[y^o - H(x)] + (x - x^b)B^{-1}(x - x^b). \tag{12}$$

where

B is the background error covariance,

R is the observation error covariance,

H is an interpolation operator (or observation operator),

x^b is the first guess or background,

y^o is the observation,

$y^o - H(x^b)$ are the observational increments

$$x^a = x^b + W[y^o - H(x^b)] \tag{13}$$

W is a weight matrix based on statistical error covariances of forecast and observations.

4 Variational methods

The start of variational methods is originally attributed to the work of Euler [113,114] and Lagrange [188,189] in the seventeenth and eighteenth century. The Euler-Lagrange equation, developed by Leonhard Euler and Joseph-Louis Lagrange in the 1750s, is the major formula of the calculus of variations. It provides a way to solve for functions which extremize a given cost functional. It is widely used to solve optimization problems, and in conjunction with the action principle to calculate trajectories. Variational calculus has had a broad appeal due to its ability to derive behavior of an entire system without details related to system components. Broadly speaking variational calculus involves finding stationary points of functionals written as integral expressions. The general theory is rigorously explained in the work by Lanczos[190] and Courant and Hilbert[309].

Basic to the constrained minimization theory is the method of undetermined Lagrange multipliers where

$$\lambda = (\lambda_1, \ldots, \lambda_n)^T \tag{14}$$

is a vector of n unknowns for the solution of

$$\min f(x) \in R^n \tag{15}$$

$$\text{subject to } g(x) = 0 \quad x \in R^m \tag{16}$$

and using the first-order conditions for a minimum we obtain using the first derivatives of the Lagrangian function

$$L(\lambda, x) = f(x) + \lambda^T g(x) \tag{17}$$

$$\nabla_x L(x, \lambda) = \frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} \tag{18}$$

$$\nabla_\lambda L(x, \lambda) = g(x) \tag{19}$$

The Lagrange multiplier λ can be viewed as measuring sensitivity of value of function f at a stationary point to changes in the constraint (see also Nocedal and Wright[280].

One can show formally (see any text book on variational methods) that finding in a given domain of admissible functions $u(x)$ the continuous first derivatives of a functional I for which $I(u(x))$ is a stationary value (i.e. any function which extremizes the cost functional) must also satisfy the ordinary differential equation called the Euler-Lagrange equation

$$\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\frac{\partial F}{\partial u'} = 0 \tag{20}$$

where

$$I(u(x)) = \int_{x_a}^{x_b} F(u(x))dx \quad x_a \leq x \leq x_b \tag{21}$$

$$u' = \frac{\partial u}{\partial x} \tag{22}$$

As an example of a typical application of variational methods, consider work of Sasaki[323–325]. Navon[264] used it to enforce conservation of total enstrophy, total energy and total mass in one and two-dimensional shallow water equations models on a rotating place. This is an aspect less discussed of Sasaki's[323–325] contribution related to the use of the variational approach to conserve integral invariants.

Let us consider here at enforcing potential enstrophy conservation

$$Z = \frac{1}{2}\int_0^L \int_0^D \frac{Q^2}{h} dxdy = \frac{1}{2}\int_0^L \int_0^D (\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} + f)^2 h^{-1} dxdy \tag{23}$$

for the 2-D shallow water equations over a limited-area domain discretized by finite-differences resulting in the functional

$$\begin{aligned} J = & \sum_{j=1}^{N_x}\sum_{k=1}^{N_y}[\tilde{\alpha}(u-\tilde{u})^2 + \tilde{\alpha}(v-\tilde{v})^2 + \tilde{\beta}(h-\tilde{h})^2]_{jk} \\ & + \lambda_z\{\sum_{j=1}^{N_x}\sum_{k=1}^{N_y}[(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} + f)^2_{jk} h^{-1}_{jk}] \cdot \Delta x \Delta y - Z_0\} \end{aligned} \tag{24}$$

where

$(\tilde{u}, \tilde{v}, \tilde{h})$ are model predicted variables,

(u, v, h) are values adjusted by the variational method,

λ_Z is a Lagrange multiplier constant w.r.t. space but varying with time.

Using algebraic commutation rules between finite-difference and variational operators, we obtain a set of nonlinear coupled partial differential equations.

$$2\tilde{\alpha}(u-\tilde{u}) + (\lambda_Z/h)(\Delta s)^2(\nabla_x \nabla_y v - \nabla_y^2 u) = 0 \tag{25}$$

$$2\tilde{\alpha}(v-\tilde{v}) + (\lambda_Z/h)(\Delta s)^2(\nabla_y \nabla_x v - \nabla_x^2 u) = 0 \tag{26}$$

$$2\tilde{\beta}(h-\tilde{h}) + (\lambda_Z/2h^2)(\Delta s)^2(\nabla_x v - \nabla_y u + f)^2 = 0 \tag{27}$$

$$\frac{1}{2}\sum_{j=1}^{N_x}\sum_{k=1}^{N_y}[(\nabla_x v - \nabla_y u + f)^2 h^{-1}(\Delta s)^2]_{jk} - Z_0 = 0 \tag{28}$$

where $\Delta x = \Delta y,\ (\Delta s)^2 = \Delta x \Delta y,\ Z_0$ is the value of discretized enstrophy at time $t = 0$, i.e initial potential enstrophy.

$\nabla x, \nabla y$-finite difference operators.

The coupled nonlinear equations require that numerical solutions of u, v, h and Lagrange multiplier λ_z be obtained using an iterative technique. Substitution of iterates $u_{jk}^{(\nu+1)}, v_{jk}^{(\nu+1)}$ and $h_{jk}^{(\nu+1)}$ into equation for potential enstrophy yields a highly nonlinear equation for λ_Z solved iteratively by a Newton iteration.

One can also consider a generalized functional including simultaneous constraints of total mass, total energy and potential enstrophy. For a discussion on these issues see also work of Takacs[346] on the issue of using a staggered mesh à la Arakawa methods and shortcomings of enforcing 'a-posteriori' integral invariants conservation.

5 First interlude

5.1 Situation in data-assimilation at beginning of 1980's

Charney, Halem and Jastrow[56] proposed that numerical models be used to assimilate newly available asynoptic data. The idea was to insert asynoptic temperature information obtained from satellite-born radiometers into the model at its true (asynoptic) time. Continuous data assimilation referred to frequent insertion of asynoptic data. Charney et al.[56] experiment suggested continuous data assimilation [345]. Problems of real data insertion soon emerged in the form of an inertia-gravity wave shock [80](Daley and Puri) leading to essential rejection by the model of the information of real observational data. A remedy for continuous data assimilation of real data was to reduce the insertion interval to the time step of the model[247] (Miyakoda et al.).

Other approaches were via geostrophic wind correction outside the tropics or nudging also referred to as Newtonian relaxation (Hoke and Anthes)[163], Davis and Turner[90]. See also work of Talagrand[347,253,348]. Ghil, Halem and Atlas[133], Mcpherson[242]. Mcpherson[242] viewed data assimilation as "a process by which something is absorbed into something else". During 1974 Marchuk[235] proposed application of adjoint method in meteorology (Russian article of 1967 [234]) and during 1976 Penenko and Obratsov [289]-used these methods to study linear adjoint sensitivity.

In 1969, Thompson[360] had already put forward the idea that incorrect analyses at two successive times may be optimally adjusted to maintain dynamical consistency with a given prediction model. This may be viewed as a precursor to variational data assimilation. Since 1958 Marchuk[233] and collaborators used adjoint methods for linear sensitivity analysis problems . Atmospheric issues were also addressed in the same fashion (see Marchuk[235]) Adjoint operators have been introduced by Lagrange[188,189] and have been used in modern times since Wigner[385] and many others in different application domains.

The advent of optimal control theory of partial differential equations is attributed to Bellman starting in the late 50's (see Bellman[18]) i.e. the Hamilton-

Jacobi-Bellman equation and to Pontryagin (Pontryagin's minimum principle)[292]. The major impetus in this area came from the monograph of Lions[218]on optimal control of partial differential equations. It was to be that a former doctoral student of Lions, Francois Le Dimet, introduced the concepts of optimal control to the meteorological community starting in the early 1980's.

One major work which impacted in a serious way the adjoint sensitivity analysis was the work of Cacuci et al.[41], D.G. Cacuci[42,43]. Historically one can trace back linear adjoint sensitivity to work of Wiener (1940-1942). See Cacuci 2004 lecture and related work of his [47],[44],[45]. Wiener was the first to interpret physically the adjoint functions (see also Lewins, 1962) as importance functions. As mentioned above Cacuci (1979-1981) [382] presented a rigorous theory for adjoint sensitivity of general nonlinear systems of equations.

LeDimet[196] was then preparing his technical report at Clermont-Ferrand introducing for the first time optimal control methodology with variational adjustment to the meteorological community , that led to the seminal paper by LeDimet and Talagrand (1986), [199]

6 Emergence of early data assimilation works

LeDimet[197], Lewis and Derber[98], Courtier [66], Le Dimet and Talagrand(1986)[199] were the first to work on adjoint data assimilation. Cacuci(1981a,1981b) [42,43] extended adjoint sensitivity analysis to the fully nonlinear case. Lagrange multiplier methods were presented in detail by Bertsekas[28], while Navon and De Villiers[265] exhibited the method in detail applied to enforcing conservation of integral invariants.

7 Optimal interpolation (OI) methods

Lev Gandin[122] coined the term (OI) but the technique of statistical interpolation can be traced back to Kolmogorov[209] and Wiener[384] and the terminology of optimal interpolation was apparently due to Wiener[384].

A review of the work of these two mathematicians is provided in the Yaglom[389] book on stochastic processes (see Lewis et al.[213]). In atmospheric sciences use of statistical interpolation goes back to Eliassen[109], while Krige[182], used it in the mining industry.

Use of least-squares method, a mathematical optimization technique which,

when given a series of measured data, attempts to find a function which closely approximates the data (a "best fit"), i.e. to obtain best estimate of state of the atmosphere by combining prior information which can consist of either a first guess or a background with observations which have errors. The concept of background field goes back to Gauss[125,126]. We wish to carry out a minimum variance estimation.

In a general form the optimal least-squares estimation is defined by the following interpolation equations

$$X^a = X_b + K(y - H[X_b]), \tag{29}$$

where K is a linear operator referred to as gain or weight matrix of the analysis and is given by

$$K = BH^T(HBH^T + R)^{-1}, \tag{30}$$

where X^a is the analysis model state,

H- an observation operator,

B- covariance matrix of the background errors $(X_b - X)$,

X- being the time model state,

X_b- background model state,

R- covariance matrix of observation errors.

The analysis error covariance matrix is

$$A = (I - KH)B(I - KH)^T + KRK^{-1} \tag{31}$$

If K is optimal least-squares gain, A becomes

$$A = (I - KH)B \tag{32}$$

(see proof in Bouttier and Courtier[35]).

One can show that the best linear unbiased estimator [350,35](Talagrand(1997), Bouttier and Courtier(1999)) may be obtained as the solution of the following variational optimization problem.

$$\min J = (X - X_b)^T B^{-1}(X - X_b) + (y - H(X))^T R^{-1}(y - H(X)) \tag{33}$$

$$= J_b(X) + J_o(X)$$

One notes that if the background and observation error probability functions are Gaussian then X_a is also the maximum likelihood estimation of X_t (time). Probability density function represents a probability distribution in terms of integrals, being non-negative everywhere with an integral from $-\infty$ to $+\infty$ being equal to 1. More exactly a probability distribution has density $f(x)$, if $f(x)$ is a non-negative Lebesgue integrable function from $R \rightarrow R$ such that the probability of the interval $[a, b]$ is given by $\int_a^b f(x)dx$ for any two numbers a and b.

For a comprehensive examination of OI in meteorology we refer to Lorenc[221] and Lorenc[222]. The most important advantage of using statistical interpolation schemes such as OI and 3-D VAR instead of empirical schemes such as SCM [176] is the fact that they are taking into account the correlation between observational increments.

How to estimate the prior error covariances B and R and the observation operator H? A difficult issue with observation operator is the case of satellite products such as radiances ,a piece of information which cannot be directly used. The observation operator performs both interpolation from model grid to satellite observation location and then uses physical theory (such as in the case of radiances) to convert model column of temperature to synthetic radiances. Observation error covariance matrix R is obtained from instrument error estimates which, if independent, mean that the covariance matrix R will be diagonal. This can facilitate computations.

Assume that background and observation error (covariances) are uncorrelated, the analysis error covariance matrix is given as

$$A = (I - KH)B(I - KH)^T + KRK^T. \tag{34}$$

Solution of minimum covariance requires

$$\frac{\partial}{\partial K}(\text{trace}(A)) = 0 \tag{35}$$

$$\frac{\partial}{\partial A}(\text{trace}BAC) = B^T C^T \tag{36}$$

$$\frac{\partial}{\partial t}(\text{trace}ABA^T) = A(B + B^T) \tag{37}$$

$$\begin{aligned}\frac{\partial}{\partial K}(\text{trace}(A)) &\equiv (I-KH)(B+B^T)H^T + K(R+R^T) \\ &= -2(I-KH)BH^T + 2KR \\ &= -2BH^T + 2K(HBH^T + R) \\ &= 0\end{aligned} \tag{38}$$

from which we obtain the optimal weight K

$$K = BH^T(HBH^T + R)^{-1}. \tag{39}$$

8 Estimating background error covariances

The background error covariance is both the most difficult error covariance to estimate and it has a most important impact on results (Kalnay)[176], Navon et al.[275]. This, since it is primarily the background error covariance that determines the spread of information as well as allowing observations of wind field to enrich information about the mass field and vice-versa.

In order to render modelling of B practically feasible some compromises had to be made with respect to statistical aspects of the covariance matrix such as anisotropy, flow dependence and baroclinicity [117]. The first approach by Hollingsworth and Lönnenberg[164] concerned statistics of innovations, namely observation - minus -background (in short forecasts) and rawinsonde observations. The assumption made was that observation errors are spatially uncorrelated and they assigned spatial correlations of innovations to the background error. Hidden in this method of use of innovation statistics is the implicit assumption of a dense homogeneous observing network.

For 3-D VAR the most popular and universally adopted method does not depend on measurements but rather uses differences between forecasts of different time-lengths which verify at the same time. It is known as the " NMC" (now NCEP) method having been introduced by Parrish and Derber[288]. In an operational numerical weather prediction they use

$$B \approx \alpha E\{[X_f(48h) - X_f(24h)][X_f(48h) - X_f(24h)]^T\} \tag{40}$$

This provides a multivariate global forecast difference covariance. If this time interval is longer than the forecast used to generate background fields then the covariances of the forecast difference will be broader than those of the background error.

A new method based on ensemble of analyses to estimate the background

errors is described in detail in Fisher[117]. The ensemble is generated by perturbing all inputs to the analysis system resulting in a perturbed analysis and a forecast is run from this perturbed analysis. By running the analysis forecast twice for same period and perturbing both runs using statistically independent perturbations-the difference between these pairs of background fields acquire statistical characteristics of differences between background error fields. For a description of novel developments including a spectral method for covariance modelling which includes inhomogeneity see Fisher[117]. For a detailed description on construction of covariance matrices see Gaspari and Cohn(1999,2006) [123,124].

A square-root factorization of the background error covariance is based on formulations provided by Weaver and Courtier[381] and Derber and Bouttier[104] and avoiding requiring the availability of the inverse of B via the transformation

$$J_b = \frac{1}{2}\delta_X^T B^{-1}\delta_X = \frac{1}{2}\delta_X^T (B^{\frac{1}{2}}B^{\frac{T}{2}})^{-1}\delta_X = \frac{1}{2}V^T V \tag{41}$$

where $\delta_X = X(t_0) - X_b,\ V = B^{-\frac{1}{2}}\delta_X$, i.e $\delta_X = B^{\frac{1}{2}}V$.

For a shallow-water equations model, the model variables are thus partitioned into balanced and unbalanced components. The balancing operator K_b acts on the unbalanced components of model variables and we have

$$K_b = K_b' + I \tag{42}$$

where K_b' is formulated based on geostrophic balance written in spherical coordinates.

$$K_b = K_b' + I = \begin{pmatrix} I & 0 & 0 \\ -\frac{g}{f}\frac{1}{a}\frac{\partial}{\partial\theta} & I & 0 \\ \frac{g}{f}\frac{1}{a\cos\theta}\frac{\partial}{\partial X} & 0 & I \end{pmatrix} \tag{43}$$

which is a lower triangular matrix for control vector $(h, u, v)^T$.

$$B = K_b B_u K_b^T \tag{44}$$

B_u is block diagonal error covariance for the unbalanced component of the variables

$$B_u = \sum\nolimits_b C \sum\nolimits_b \tag{45}$$

$\sum_b$ block diagonal matrix of background error variances at every grid point. C is a symmetric matrix assumed block diagonal and we have square root factorization

$$C = C^{\frac{1}{2}} C^{\frac{T}{2}} \tag{46}$$

The square root factorization of B ensuring it is symmetric and positive definite is provided by

$$\begin{aligned} B &= K_b B_u K_b^T = K_b (\sum\nolimits_b C \sum_b) K_b^T \\ &= K_b (\sum_b C^{\frac{1}{2}} C^{\frac{T}{2}} \sum_b) K_b^T \\ &= (K_b \sum_b C^{\frac{1}{2}})(C^{\frac{T}{2}} \sum_b K_b^T) \\ &= B^{\frac{1}{2}} B^{\frac{T}{2}} \end{aligned} \tag{47}$$

9 Framework of Variational data Assimilation

The objective of variational 4-D Var is to find the solution to a numerical forecast model that best fits a series of observational fields distributed in space over a finite time interval. We are assuming that the model of the atmosphere can be written as

$$B \frac{dX}{dt} + A(X) = 0 \tag{48}$$

with B being identity for a dynamical model or the null operator for a steady state model. A can be a linear or nonlinear operator. We have U defined as a control variable which may consist of initial conditions, boundary conditions and/or model parameters.

U should belong to a class admissible controls U_{ad}. We are looking for a unique solution $X(U)$ of (48). The major step consists in formulating the cost function J which measures distance between model trajectory and observations as well as the background field at initial time during a finite time-interval, referred to as the time window.

Typically in meteorology (see LeDimet and Talagrand[199], Rabier[304]).

$$J(X_0) = \frac{1}{2}(X_0 - X_b)^T B^{-1} (X_0 - X_b) \tag{49}$$

$$+\frac{1}{2}\sum_{i=0}^{N}(H_i(X_i)-y_i)^T R_i^{-1}(H_i(X_i)-y_i)$$

where

X_0 is the NWP model state as time t_0,

X_b-background state at time t_0, typically a $6h$ forecast from a previous analysis,

B-the background error covariance matrix,

y_i-the observation vector at time t_i,

H_i-observation operator,

$X_i = M_{i,0}(X_0)$ model state at time t_i,

R_i-observation error covariance matrix at time t_i.

where an alternative to writing the NWP model is

$$X_{i+1} = M_{i+1,i}(X_i) \tag{50}$$

$M_{i+1,i}$ is the nonlinear NWP model from time t_i to time t_{i+1}.

The minimization of the cost functional can be viewed both in the perspective of finding its gradient in (a) Lagrangian approach, (b) adjoint operator approach and (c) a general synthesis of optimality conditions in the framework of optimal control theory approach. Requiring the gradient of the cost to vanish with respect to initial conditions control variable X_0 yields

$$\nabla_{X_0} J(X_0) = B^{-1}(X_0 - X_b) + \sum_{i=0}^{N} \mathbf{M}_{i,0}^T \mathbf{H}_i^T R_i^{-1}[H_i(X_i) - y_i] \tag{51}$$

where we substitute the dynamical constraint

$$X_{i+1} = M_{i+1,i}(X_i) \tag{52}$$

while perturbations of the atmospheric state are obtained by linearizing the nonlinear model (52) as

$$\delta X_{i+1} = \mathbf{M}_{i+1,i}(X_i)\delta X_i \tag{53}$$

yielding

$$\nabla_{X_0} J(X_0) = B^{-1}(X_0 - X_b) + \sum_{i=0}^{N} \mathbf{M}_{i,0}^T \mathbf{H}_i^T R_i^{-1}[H_i(X_i) - y_i] \tag{54}$$

where $\mathbf{H}_i$ is the tangent linear operator of the observation operator H_i and $\mathbf{H}_i^T$ is the adjoint operator and

$$\mathbf{M}_{i,0}^T = \mathbf{M}_{1,0}^T \mathbf{M}_{2,1}^T \cdots \mathbf{M}_{i,i-1}^T \tag{55}$$

is the adjoint model consisting of a backward integration from time t_i to time t_0.

The minimization of the cost functional is obtained using a gradient-based minimization algorithm. Starting from a first guess

$$X^0(t_0) = X_b(t_0) \tag{56}$$

while at each iteration step $k = 1, 2, \cdots, N$

we compute and store both first guess trajectory and the observation departures $H_i(X_i) - y_i$

by integrating forward in time the nonlinear model

$$X^k(t_i) = M(t_i, t_0)(X^k(t_0)) \tag{57}$$

Start with initializing the adjoint variable at time t_N

$$\delta' X^k(t_N) = 0 \tag{58}$$

integrating the adjoint model backwards in time from final time t_N to initial time t_0. and whenever observations are encountered a forcing term

$$\mathbf{H}_i^T R_i^{-1}(H_i(X_i) - y_i) \tag{59}$$

is added to $\delta' X^k(t_i)$.

Finally one can show that

$$\delta' X^k(t_0) + B^{-1}[X^k(t_0) - X_b] \tag{60}$$

is the gradient ∇J^k with respect to the control variable $X^k(t_0)$.

If

$$||\nabla_{X_0} J^{k+1}|| \leq \epsilon \max\{1, ||X_k||\} \tag{61}$$

(where ϵ is a predetermined adequately chosen tolerance.) If above criterion is satisfied then stop.

If the above criterion is not satisfied then use a stepsize search algorithm using, say, a cubic interpolation usually provided by the gradient based minimization algorithm.

One then updates the first guess, namely

$$X^{k+1}(t_0) = X^k(t_0) - \rho^k \nabla J^k \tag{62}$$

where ρ is a step-size in the direction of descent and find the next minimization iterate using a gradient based minimization algorithm.

All the time we assume that the nonlinear cost function has a unique minimum and avoid temporarily addressing the complex issue of the presence of multiple minima.

10 Variational formalism

10.1 The Lagrangian approach

One can consider a model given as in LeDimet and Talagrand[199] by

$$F(U) = 0 \tag{63}$$

where U denotes meteorological fields being considered. Suppose we have observations $\hat{U}$ occurring at an irregular set of points distributed in both space and time.

We wish to solve the problem of finding a solution that minimizes a cost function

$$J(U) = \int ||U - \hat{U}||^2 dxdydt \tag{64}$$

where $||\quad||$ is a suitable norm and $\hat{U}$ consists of discrete observations hence the integral is replaced by suitable finite sums. Here we view the model equation

$$F(U) = 0 \tag{65}$$

as a strong constraint on cost function J. Using classical Lagrange multiplier technique a Lagrangian of (64) subject to model strong constraint allows us to convert this constrained minimization into an unconstrained minimization problem by defining a Lagrangian (see Bertsekas[28]) as

$$L(U, \lambda) = J(U) + (\lambda, F(U)) \tag{66}$$

for an adequately defined inner product for a functional space in which $F(U)$ also belongs.

Then finding minima of $J(U)$ subject to

$$F(U) = 0 \tag{67}$$

is equivalent to finding the minima of

$$\nabla_\lambda L = 0 \quad \text{and} \tag{68}$$

$$\nabla_U L = 0 \tag{69}$$

which taking into account boundary conditions turns out to be the Euler-Lagrange equations of the problem. Since the Euler-Lagrange equations can seldom be solved directly, we are interested in practical algorithms for solving the minimization of cost functional subject to strong model constraint by transforming it into a sequence of unconstrained minimization problems.

There are many constrained minimization algorithms-but the simplest and most robust of them are the penalty and the multiplier (or duality) algorithms. These are presented in many numerical minimization text books, (Nocedal and Wright[280], Nash and Sofer[261]) For shortcomings of penalty and duality algorithms see Bertsekas[28] and Navon and De Villiers[265].

In the augmented Lagrangian algorithm(where the constrained problem is converted into a sequence of unconstrained minimization problems) we have

$$L(\rho, U, \lambda) = J(U) + \lambda^T F(U) + \rho||F(U)||^2 \tag{70}$$

This algorithm was initially proposed by Hestenes[161] and independently by Powell[293]. Here $\rho > 0$ is a penalty coefficient.

The solution is obtained by performing a sequence of unconstrained minimization where (U_K, λ_K, ρ_K) are known. We minimize $L_{\rho_K}(U_K, \lambda_K)$ and obtain U_{K+1}.

The updating of the multipliers and the penalty in the iterative procedure of the Augmented Lagrangian algorithm is done via first order multiplier update

$$\lambda_{K+1} = \lambda_K + \rho_K F(U_{K+1}) \tag{71}$$

where

$$\rho_{K+1} = \beta\gamma_K \tag{72}$$

where β positive and large than 1. Here k is the iteration index.

The full description of the methodology and application of the augmented Lagrangian is available in Bertsekas[28], Fortin and Glowinski[119] and described in detail in Navon and De Villiers[265]. A gradient based descent algorithm is used for the unconstrained minimization.

11 Optimal control view point

In optimal control of partial differential equations developed by Lions[218,219] the Lagrange multiplier is viewed as an adjoint variable. The adjoint method of optimal control allows computing the gradient of a cost J with respect to the control variables.

Consider as in Gunzburger[153] a second order nonlinear elliptic PDE

$$-\nabla(a\nabla\phi) + b\cdot\nabla\phi + \phi^3 = \sum_{k=1}^{K} \alpha_K f_K \tag{73}$$

in domain Ω with boundary conditions

$$\phi = 0 \quad \text{on} \qquad \Gamma \tag{74}$$

a, b and f_K are given functions defined on Ω.

We define a cost as

$$J(\phi, \alpha_1, \cdots, \alpha_K) = \frac{1}{2}\int_{\Omega}(\phi - \Phi)^2 d\Omega + \frac{\sigma}{2}\sum_{k=1}^{K}(\alpha_K)^2 \tag{75}$$

Φ is a given function and σ a penalty parameter. We introduce a Lagrange multiplier (here adjoint variable) ζ and define a Lagrangian

$$L(\phi, g, \zeta) = J(\phi, g) - \zeta^T F(\phi, g) \tag{76}$$

We aim to find controls g, states ϕ and adjoint states ζ such that the Lagrangian is stationary and we obtain as in the Augmented Lagrangian approach

$$\frac{\partial L}{\partial \zeta} = 0, \quad \text{constraint} \tag{77}$$

$$\frac{\partial L}{\partial \phi} = 0, \quad \text{adjoint equation} \tag{78}$$

$$\frac{\partial L}{\partial g} = 0, \quad \text{optimality condition} \tag{79}$$

Taking a first order variation of L with respect to the Lagrange multiplier we obtain a variation in the state yielding an optimality condition

$$(\frac{\partial F}{\partial \phi}|_{(\phi,g)})^T \zeta = (\frac{\partial J}{\partial \phi}|_{(\phi,g)})^T \tag{80}$$

which yields the optimality condition.

12 Situation of data assimilation-The early period (1980-1987) of 4-D Var

Efforts in early adjoint applications following Francois LeDimet[197] early technical report consisted of work of Lewis and Derber[212] and LeDimet and Talagrand[199] as well as Courtier[66]. These pioneering efforts started the meteorological optimal control application called the adjoint operator approach. Work of Navon and De Villiers[265] on Augmented Lagrangian methods used to enforce conservation of integral invariants related to the same topic and are referred to in early work of LeDimet and Talagrand[199].

John Lewis and John Derber[212] were the first to present adjoint method having read the report of Francois le Dimet (1982) and inspired by earlier work of Thompson[360]). Lorenc[222] presented a detailed account of state of theory in data assimilation for that period. It become soon apparent that size and complexity of atmospheric equations is such that enormous resources

were required-limiting applications of 4-D VAR and requiring it to undergo drastic approximations for actual operation forecast circumstances. Penenko and Obratsov[289] used adjoint data assimilation to perform simple experiments on a linear model (see Talagrand and Courtier[349]) , while Derber[98] used it in his Ph.D thesis to adjust analysis to a multi-level quasi-geostrophic model. Hoffmann[162] was the next one to use 4-D VAR (even though he used a simplified primitive equation model and in order to estimated the gradient he perturbed in turn all the components of the initial state.)

Talagrand and Courtier[349] presented a more in-depth general exposition of the theory of adjoint equations in the framework of variational assimilation and applied it to the inviscid vorticity equation and to the Haurwitz wave. Their results are presented in Courtier and Talagrand[68].

13 Illustrative example

Our purpose is to illustrate how to derive the adjoint of the shallow water equations model explicitly.

The shallow water equations model may be written as

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} + fv - \frac{\partial \phi}{\partial x}, \tag{81}$$

$$\frac{\partial v}{\partial t} = -u\frac{\partial v}{\partial x} - v\frac{\partial v}{\partial y} - fu - \frac{\partial \phi}{\partial y}, \tag{82}$$

$$\frac{\partial \phi}{\partial t} = -\frac{\partial (u\phi)}{\partial x} - \frac{\partial (v\phi)}{\partial y}, \tag{83}$$

where u, v, ϕ and f are the two components of the horizontal velocity , geopotential fields and the Coriolis factor, respectively.

We shall use initial conditions due to Grammeltvedt [145]

$$h = H_0 + H_1 tanh\frac{9(y-y_0)}{2D} + H_2 sech\frac{9(y-y_0)}{D}\sin\frac{2\pi x}{L}, \tag{84}$$

where $H_0 = 2000m$, $H_1 = -220m$, $H_2 = 133m$, $g = 10msec^{-2}$, $L = 6000km$, $D = 4400km$, $f = 10^{-4}sec^{-1}$, $\beta = 1.5 \times 10^{-11}sec^{-1}m^{-1}$. Here L is the length of the channel on the β plane, D is the width of the channel and $y_0 = D/2$ is the middle of the channel. The initial velocity fields were derived from the

initial height field via the geostrophic relationship, and are given by

$$u = -\frac{g}{f}\frac{\partial h}{\partial y}, \tag{85}$$

$$v = \frac{g}{f}\frac{\partial h}{\partial x}. \tag{86}$$

The time and space increments used in the model are $\Delta x = 300km$, $\Delta y = 220km$, $\Delta t = 600s$, which mean that there are 21×21 grid point locations in the channel and the number of the components of initial condition vector $(u, v, \phi)^t$ is 1083. Therefore the Hessian of the cost function in our test problem has a dimension of 1083×1083. The southern and north boundaries are rigid walls where the normal velocity components vanish, and it is assumed that the flow is periodic in the west-east direction with a wave length equal to the length of the channel.

Let us define

$$\vec{X} = (u, v, \phi)^T, \tag{87}$$

$$F = -\begin{pmatrix} u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + \frac{\partial \phi}{\partial x} \\ u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + \frac{\partial \phi}{\partial y} \\ \frac{\partial (u\phi)}{\partial x} + \frac{\partial (v\phi)}{\partial y} \end{pmatrix}. \tag{88}$$

It is easy to verify that

$$\frac{\partial F}{\partial \vec{X}} = -\begin{pmatrix} \frac{\partial (u(\cdot))}{\partial x} + v\frac{\partial (\cdot)}{\partial y} & (\cdot)\frac{\partial u}{\partial y} - f(\cdot) & \frac{\partial (\cdot)}{\partial x} \\ (\cdot)\frac{\partial v}{\partial x} + f(\cdot) & u\frac{\partial (\cdot)}{\partial x} + \frac{\partial (v(\cdot))}{\partial y} & \frac{\partial (\cdot)}{\partial y} \\ \frac{\partial (\phi(\cdot))}{\partial x} & \frac{\partial (\phi(\cdot))}{\partial y} & \frac{\partial (u(\cdot))}{\partial x} + \frac{\partial (v(\cdot))}{\partial y} \end{pmatrix}. \tag{89}$$

The adjoint of an operator L, L^* is defined by the relationship

$$< L\vec{X}, \vec{Y} > = < \vec{X}, L^*\vec{Y} >, \tag{90}$$

where

$$< \cdot, \cdot > \tag{91}$$

where D is the spatial domain. Using the definition (91), the adjoint of (90)

can be derived as

$$[\frac{\partial F}{\partial \vec{X}}]^* = -\begin{pmatrix} -u\frac{\partial(\cdot)}{\partial x} - \frac{\partial(v(\cdot))}{\partial y} & (\cdot)\frac{\partial v}{\partial x} + f(\cdot) & -\phi\frac{\partial(\cdot)}{\partial x} \\ (\cdot)\frac{\partial u}{\partial y} - f(\cdot) & -v\frac{\partial(\cdot)}{\partial y} - \frac{\partial(u(\cdot))}{\partial x} & -\phi\frac{\partial(\cdot)}{\partial y} \\ -\frac{\partial(\cdot)}{\partial x} & -\frac{\partial(\cdot)}{\partial y} & -u\frac{\partial(\cdot)}{\partial x} - v\frac{\partial(\cdot)}{\partial y} \end{pmatrix}. \quad (92)$$

Therefore the first order adjoint model with the forcing terms may be written as

$$-\frac{\partial u^*}{\partial t} = -(-u\frac{\partial u^*}{\partial x} - \frac{\partial(vu^*)}{\partial y} + v^*\frac{\partial v}{\partial x} + fv^* - \phi\frac{\partial \phi^*}{\partial x}) + W_u(u - u^o), \quad (93)$$

$$-\frac{\partial v^*}{\partial t} = -(u^*\frac{\partial u}{\partial y} - fu^* - v\frac{\partial v^*}{\partial y} - \frac{\partial(uv^*)}{\partial x} - \phi\frac{\partial \phi^*}{\partial y}) + W_v(v - v^o), \quad (94)$$

$$-\frac{\partial \phi^*}{\partial t} = -(-\frac{\partial u^*}{\partial x} - \frac{\partial v^*}{\partial y} - u\frac{\partial \phi^*}{\partial x} - v\frac{\partial \phi^*}{\partial y}) + W_\phi(\phi - \phi^o), \quad (95)$$

with final conditions

$$u(T) = 0,\ v(T) = 0,\ \phi(T) = 0, \quad (96)$$

where $P = (u^*, v^*, \phi^*)^t$ is the first order adjoint variable vector , W_u, W_v, W_ϕ are weighting factors which are taken to be the inverse of estimates of the statistical root-mean-square observational errors on geopotential and wind components respectively. In our test problem, values of $W_\phi = 10^{-4}m^{-4}s^4$ and $W_u = W_v = 10^{-2}m^{-2}s^2$ are used.

14 OI, 3-D VAR, PSAS

Lorenc[222] showed that the optimal weight matrix W that minimizes the matrix of analysis error covariance solution may be posed in terms of a variational assimilation problem, namely that of finding the optimal analysis field X_a that minimizes a cost function. The cost function measures the distance between the field variables X and the background X_b (the background term of the cost)-plus another term, namely the distance to the observations y^o weighted by the inverse of the observation error covariance matrix R

$$J(X) = \frac{1}{2}(X - X_b)^T B^{-1}(X - X_b) + [y^o - H(X)]^T R^{-1}[y^o - H(X)] \quad (97)$$

where H is the forward observational operator. The cost function (97) can also be derived based on a Bayesian approach.

A formalism allowing viewing the assimilation algorithms of O-I, 3-D VAR, PSAS and 4-D VAR as a sequence of corrections to a model state can be derived from the work of Lorenc[222], Kalnay[176] and Courtier[75]. See also research work of Da Silva et al.[86] who first proposed the physical space statistical analysis system (PSAS) (see also report of Aarnes[1]).

We are considering incrementing background model state X_b with additional information from the observation z where

$$X_a = X_b + K(z - HX_b). \tag{98}$$

Here H is an observation operator mapping the model state on space and time locations of the observation, X_a is the analysis and K is the gain matrix weighting the contributions from the new information according to the reliability of the observation relative to respective reliability of the model state. Following Kalnay[176], Lorenc[222] OI, 3-D VAR, 4-D VAR and PSAS are mathematically equivalent but 3-D VAR and related PSAS have the advantage w.r.t. OI by virtue of the fact that one can minimize the cost function J with global unconstrained minimization algorithms for 3-D VAR hence all the approximation made in OI are not necessary. Other advantages of 3-D VAR are enumerated in Kalnay[176].

To show equivalence of 3-D VAR and OI we start from the matrix system

$$\begin{pmatrix} R & H \\ H^T & -B^{-1} \end{pmatrix} \begin{pmatrix} W \\ X_a - X_b \end{pmatrix} = \begin{pmatrix} z - HX_b \\ 0 \end{pmatrix} \tag{99}$$

where R and B are the error observation error and background error covariance matrices, respectively, assumed to be symmetric and positive-definite. The equivalence between OI and 3-D VAR statistical problems was proven by Lorenc[222], Kalnay [176] and using suggestion of Jim Purser(see Kalnay [176])

$$W = K_{OI} = BH^T(R + HBH^T) \tag{100}$$

To see the equivalence between OI and the PSAS scheme where minimization is performed in the space of observations rather than in the model space (since the number of observation is usually much smaller than the dimension of model space-PSAS may turn out to be more efficient than 3-D VAR for obtaining similar results)we note that

$$\begin{pmatrix} R & H \\ H^T & -B^{-1} \end{pmatrix} \begin{pmatrix} W \\ X_a - X_b \end{pmatrix} = \begin{pmatrix} z - HX_b \\ 0 \end{pmatrix} \tag{101}$$

is equivalent to

$$\begin{pmatrix} W & 0 \\ H^T & -B^{-1} \end{pmatrix} \begin{pmatrix} W \\ X_a - X_b \end{pmatrix} = \begin{pmatrix} z - HX_b \\ 0 \end{pmatrix} \tag{102}$$

yielding

$$w = W^{-1}(z - HX_b) \tag{103}$$

and

$$X_a - X_b = BH^T W^{-1} \tag{104}$$

One first solves the linear system

$$W_w = z - HX_b \tag{105}$$

and then interpolates solution onto model space as

$$X_a = X_b + BH^T w \tag{106}$$

In PSAS one solves the first step by minimizing the cost functional

$$J(w) = \frac{1}{2} w^T W w - w^T (Z - HX_b) \tag{107}$$

thus allowing a better conditioning of the minimization due to smaller dimension of W i.e

$$\dim(W) \leq \dim(B) \tag{108}$$

Courtier[75] has shown that there is a duality between 3-D VAR and the physical space statistical analysis system (PSAS). He also showed that the temporal extension of 3-D VAR leads to 4-D VAR while the temporal extension of PSAS,4-D VAR PSAS is achieved using an algorithm related to the representer technique (Bennett[23]), which is a practical algorithm for decoupling the Euler-Lagrange equations associated with the variational problem with weak constraint. (see Amodei[3])

15 4-D VAR developments in early 1990's

A comprehensive list of adjoint applications to meteorological problem is provided by Courtier[72]. The early 1990's were characterized by publication of many research efforts related to extending 4-D VAR data assimilation to multilevel primitive-equation models using analyses as observations along with other work using synthetic observations. See for instance Thepaut and Courtier[356], Navon et al.[271], Zupanski[404]. Thepaut et al.[357] used real observations while Rabier and Courtier[298] studied the performance of 4-D VAR in the presence of baroclinic instability. Courtier et al.[74] introduced an incremental formulation of the 4-D VAR, a major achievement allowing the 4-D VAR method to become computationally feasible on that period's computers.

It was perceived rather early by Derber[100] that the perfect model hypothesis is a weakness of 4-D VAR. In the above seminal paper he assumed the model error to be fully correlated in time and solved the problem by including the bias in the control variable. Wergen[383] and Miller et al.[245] illustrated how serious the problem is.

At universities research in 4-D VAR data assimilation proceeded to address issues such as the impact of incomplete observations on 4-DVAR (see Zou et al.[397]), while at the suggestion of Francois Le Dimet, Zhi Wang completed a doctoral thesis on second order adjoint methods (Wang[376]), as well as a first paper on second order adjoint data assimilation.(Wang et al.)[375] Initial work on 4-D VAR data assimilation with the semi-implicit semi Lagrangian (SLSI)models in 2-D and 3-D was using both shallow-water and a NASA multilevel model.(see Li et al.[215–217]) Basic work on optimization methods suitable for 4-D VAR was carried out by Zou et al.[399] based on Navon et al.[272,273]. Application of 4-D VAR to a finite-element model of the shallow-water equations was carried out by Zhu et al.[392] while a novel Hessian preconditioning method based on an idea of Courtier et al.[74] was written by W. Yang et al.[390] Aspects of 4-D VAR dealing with boundary conditions as control variables were dealt amongst others in the work of Zou et al. (1995).

16 Large scale unconstrained minimization algorithms suitable for 4-D VAR variational data assimilation

The operational implementation of the 4-D VAR method hinges crucially upon fast convergence of the large-scale unconstrained minimization algorithm for the minimization of the cost functional. Since problems in oceanography

and meteorology contain many degrees of freedom ($\approx 10^6 - 10^7$), conjugate-gradient methods (Navon and Legler[268]) and limited-memory quasi-Newton (LMQN) methods along with Truncated-Newton methods are the only ones being considered due to the fact that only information from the first few iterations has to be saved. Thus is due to huge memory requirements of 4-D VAR taxing the capability of present day computers. All these methods have in common the fact that they require storing only a few vectors in memory. The studies of Gilbert and Lemarechal[141] and of Liu and Nocedal[220]indicate that L-BFGS (LMQN) and its French equivalent M1QN3 are among the best LMQN methods available to date . Indeed these methods are now exclusively used in 4-D VAR implementations at operational centers. The aim of this section is just to present some of the most useful algorithms mostly, but not only, for large-scale unconstrained minimization, which are considered and implemented in 4-D VAR applications.

In view of further 4-D VAR applications we present here a very short preview of constrained optimization methods is also presented. Since there are problems where fundamental questions exist concerning the uniqueness of the retrieved solution (Li[214]; Gauthier[127]), global minimization tools may also be required in future research. This topic is however outside the scope of this review. Access to user-friendly subroutine libraries is required, along with availability of minimization software such as present in IMSL, NAG and Harwell software libraries along with the newly developed LANCELOT library for large-scale nonlinear constrained optimization (Conn, Gould and Toint[65]) Following Nocedal[279], Nocedal and Wright[280]I will enumerate below a selection of most useful algorithms for unconstrained minimization, stressing those that are useful for solving large-scale unconstrained minimization problems.

1. Conjugate-gradient (C-G) methods are very useful for solving very large problems and can be efficiently implemented on multiprocessor machines. A survey of their properties and application to problems in meteorology was provided by Navon and Legler[268]. Amongst the C-G methods we mention:

- The Powell C-G algorithm with restarts[294]. This version of the C-G method with restarts is subroutine VA14 of the Harwell subroutine library (1977).
- CONMIN of Shanno and Phua[330], which is a robust extension of the C-G method requiring a few-more vectors of storage.

Quasi-Newton Methods

BFGS - variable metric or quasi-Newton methods are very efficient but are not good candidates for large-scale optimization required in 4-D VAR data assimilation due to the fact that storage of the approximation to the Hessian matrix is required. This despite the fact that the BFGS QN method is both

fast and robust.[280]

Limited Memory Quasi-Newton Methods

2. LMQN (Limited Memory Quasi-Newton) methods for large-scale optimization. We will mainly focus on the limited memory L-BFGS , an LMQN, which resembles the BFGS method but avoids storage of matrices. The Liu and Nocedal version[220] is implemented in the Harwell routine VA15 while a very similar version, M1QN3 was developed by Gilbert and Lemarechal[141] at INRIA in the library MODULOPT. These methods proved to be the main stay of 4-D VAR optimization in the last few years (Navon et al. [271]; Zou et al.[397,398]; Zou et al.[399]) Thepaut and Courtier[356]; etc.)

The most efficient method employed operationally is the L-BFGS method of Liu and Nocedal[220] referred by the French as M1QN3 (Gilbert and Le Lemarechal [141]. It uses an economic number of vectors(between 5 and 7) and does not require any matrix storage thus being suitable for operational implementation of 4-D VAR at operational centers.

3. Newton's method.

The NAG library has a good line search implementation of the Newton method while the IMSL library has a trust region implementation based on Dennis and Schnabel[97], Gay[130]. This method, while quadratically convergent, however due to its requirement of Hessian matrix storage is not recommended for large-scale minimization in the 4-D VAR context.

Truncated Newton (T-N) methods.

These methods are very suitable for large-scale minimization and require only function and gradients storage. Both the Nash T-N method Nash(1984b)[257] as well as the more recent TNPACK algorithm of Schlick and Fogelson)[326,327]provide robust codes for T-N . Work with Wang [378,379] as well as work by Le Dimet, Navon and Daescu [204] using second order adjoint Hessian /vector products showed that they are very useful if second order adjoint information is available. Their applicability to 4-D VAR assimilation with operational models in 3-D space and time remains still to be verified.

Hybrid methods combining the L-BFGS with the Truncated Newton method have been recently proposed by Morales and Nocedal[250],Morales and Nocedal[251] and have been in data assimilation tested in 2-D models by Daescu and Navon[77] and Das, Meirovitch and Navon[88] with certain amount of success.

17 Spectrum of the Hessian and rate of convergence of unconstrained minimization

Hessian information is crucial in many aspects of both constrained and unconstrained minimization. All minimization methods start by assuming a quadratic model in the vicinity of the minimum of a multivariate minimization problem. For the problem

$$\min_{X \in R^n} F(X) \tag{109}$$

the necessary condition for X^* to be a stationary point is

$$\nabla F(X^*) = 0. \tag{110}$$

The eigenvalues of the Hessian matrix predict the behavior and convergence rate for unconstrained minimization. To show this, let us consider again the multivariate nonlinear function $F(X)$ of (109)and let X^* denote a local minimizer point that satisfies the condition

$$F(X^*) \leq F(X) \tag{111}$$

for all X such that

$$|X - X^*| < \varepsilon \tag{112}$$

where ε is typically a small positive number whose value may depend on the value of X^*. We define $F(X^*)$ as an acceptable solution of (109).

If F is twice continuously differentiable, and X^* is an absolute minimum then

$$\nabla F(X^*) = 0 \tag{113}$$

and the Hessian $G(X^*)$ of F at X^* is positive-definite, i.e.

$$\mathbf{P}^T G(X^*)\mathbf{P} > 0, \quad \forall \mathbf{P} \in R^n \tag{114}$$

Let us expand F in a Taylor series about X^*

$$\begin{aligned} F(X) &= F(X^* + h\mathbf{P}) \\ &= F(X^*) + \frac{1}{2}h^2 \mathbf{P}^T G(X^*)\mathbf{P} + O(h^2), \quad (\text{since} \nabla F(X^*) = 0) \end{aligned} \tag{115}$$

where

$$||\mathbf{P}||^2 = 1 \quad \text{and} \quad h = |X - X^*| \tag{116}$$

For any acceptable solution we obtain

$$h^2 = |X - X^*|^2 \approx \frac{2\varepsilon}{\mathbf{P}^T G(X^*)\mathbf{P}} \tag{117}$$

substantially affects size of $|X - X^*|$ i.e., rate of convergence of the unconstrained minimization (Gill[143]).

If $G(X^*)$ is ill-conditioned, the error in X will vary with the direction of the perturbation $\mathbf{P}$.

If $\mathbf{P}$ is a linear combination of eigenvectors of $G(X^*)$ corresponding to the largest eigenvalues, the size of $|X - X^*|$ will be relatively small, while if, on the other hand is a linear combination of eigenvectors of $G(X^*)$ corresponding to the smallest eigenvalues, the size of $|X - X^*|$ will be relatively large, i.e., slow convergence. For details see LeDimet et al.[202].

18 Non differentiable minimization

If the function F to be minimized is non smooth then methods of non differentiable optimization are required. They can be divided into two main classes: subgradient methods and bundle methods.

Since the gradient of a non smooth function F exists only almost anywhere we have to replace the gradient by the generalized gradient of the form

$$\begin{aligned} \partial\, F(X) = \mathbf{conv}\{g | &\textbf{there exists sequence } (X_i)_{i\in N} \textbf{ such that} \\ &\lim_{i\to\infty} X_i = X, F \textbf{ differentiable at} X_i,\ i \in N, \textbf{and } \lim_{i\to\infty} \nabla F(X_i) = g\} \end{aligned} \tag{118}$$

where "conv" stands for convex hull and it is defined as the closure of the set which contains all convex linear combinations of subgradients (an element of the generalized gradient is called subgradient).

The non smooth optimization methods are based on the assumptions that the function F is locally Lipschitz continuous and we can evaluate the function and its arbitrary subgradient at each point.

18.1 The subgradient methods

The history of subgradient methods starts in the 60s with Shor[331], Polyak[291] and Ermolev[110] to mention but a few.

The main idea is to employ only one subgradient $\xi_k \in \partial F(X_k)$ instead of the gradient $\nabla F(X_k)$. Hence the natural generalization of gradient method is to replace the gradient by the normalized gradient in the formula for d_k defined above:

$$d_k = -\xi_k / ||\xi_k|| \tag{119}$$

The above strategy of generating d_k does not ensure descent and hence minimizing line searches becomes unrealistic. Also the standard stopping criterion can no longer be applied, since an arbitrary subgradient contains no information on the optimality condition $0 \in \partial F(X)$.

Due to these facts we are forced to use an a- priori choice of step sizes t_k to avoid line searches and the stopping criterion. Thus we define the next iteration point by

$$X_{k+1} = X_k - t_k \frac{\xi_k}{||\xi_k||} \tag{120}$$

where $\xi_k \in \partial F(X_k)$ and a suitable $t_k > 0$ was chosen.

In order to accelerate the rate of convergence we may try to generalize smoother methods than the gradient method. The most efficient methods presently available are based on generalized Quasi-Newton methods: ellipsoid and space dilation algorithms by Shor [331] and the variable metric method by Uryasev[369]

18.2 The bundle methods

The guiding principle behind them is to exploit the previous iterations by gathering the subgradient information into a bundle of subgradients. The pioneering bundle method, the ε -steepest descent method, was developed by Lemarechal [205]. The main difficulty in Lemarechal's method is the a priori choice of an approximation tolerance which controls the radius of the ball in which the bundle model is thought to be a good approximation of the objective function. For the application of non-differentiable minimization methods in 4-D VAR with discontinuous cost functions see Zhang et al.[391] and Homescu and Navon[166]).

18.3 Constrained Minimization

In as far as constrained minimization is concerned its use in data assimilation pertains mainly to use of penalty and barrier functions for adjoint parameter estimation problems where the parameters are known to be subject to upper and lower bounds (Zhu and Navon[394]).

Penalty algorithms have been also used by Zou et al.[397] for issues related to data assimilation with incomplete observations.

18.4 Open problems in minimization

Use of stochastic minimization of the type of simulated annealing, genetical algorithms and neural networks is in its infancy (See work of Krasnopolsky et al.[180,181] on neural networks. When one has a problem characterized by multiple minima , use of simulated annealing, genetic algorithms or other methods in determining the global minimum will have to be tested in the framework of 4-D Var data assimilation. Issues of computational cost will be paramount when considering such methods in future operational implementation.

19 Adjoint parameter estimation in meteorology

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

Zou et al.[398] estimated the magnitude of the nudging coefficient in the NMC adiabatic version of the spectral MRF (Medium Range Forecast) model, while Wang[376] estimated the same coefficient using the FSU adiabatic spectral model. Stauffer and Bao (1993) also performed a parameter estimation of nudging coefficients in a 1-D linearized shallow water equations model. Wergen[383] 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 observations the parameters were recovered to an acceptable degree of accuracy. Louis and Zivković[208] 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 meteorol-

ogy can be viewed to be an extension of the 4-D VAR approach for controlling initial or initial and boundary conditions. For a detailed survey of the state-of-the-art of parameter estimation in meteorology and oceanography, covering all aspects of various techniques employed for parameter estimation see Navon[274]. 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 with respect 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 upper and lower 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[261] and Nocedal and Wright[280]) for details). 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 only address this issue briefly at the end of this section. For details related to identifiability see Navon[274].

Stratification of groups of parameters to be optimally identified may proceed in geosciences by either seasonal stratification or by 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.

19.1 Adjoint parameter estimation : Implementation details

While a sizable amount of research on adjoint parameter estimation was carried out in the last twenty years in fields such as groundwater hydrology and petroleum reservoirs for instance by Carrera and Neuman[50–52] Yeh[388], Seinfeld and Kravaris[329] matched by a major effort of the mathematical community such as Chavent and Lemonnier[57] and Chavent et al. [58] 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 obtained in above mentioned research fields. See Sun et al [342,343] for state of the art surveys on inverse methods for parameter estimation.

Some detailed text books have been written in the last few years addressing inverse parameter estimation issues. We will mention the recent book by Aster, Borchers and Thurber[7], Tarantola[352] and Sun[341] to quote but a few. We will focus here on adjoint parameter estimation in the framework of 4-D VAR data assimilation.

A typical cost functional in adjoint parameter estimation takes the form (see for instance, Zou, Navon, and Le Dimet[398]

$$J(\mathbf{X},\mathbf{P}) = \int_{t_0}^{t_R} < W(\mathbf{X}-\mathbf{X}^{obs}), (\mathbf{X}-\mathbf{X}^{obs}) > dt \tag{121}$$
$$+ \int_{t_0}^{t_R} K < \mathbf{P}-\hat{\mathbf{P}}, \mathbf{P}-\hat{\mathbf{P}} > dt$$

where vector $\mathbf{P}$, represents model parameters, $\hat{\mathbf{P}}$ is the vector of estimated parameters, K are specified weighting matrices, $\mathbf{X}^{obs}$ is the observation vector, $\mathbf{X}$ 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 model equation is schematically

$$\frac{\partial \mathbf{X}}{\partial t} = \mathbf{F}(\mathbf{X}) + K(\mathbf{P}-\hat{\mathbf{P}}) \tag{122}$$

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

$$\nabla_P J = 2K(\mathbf{P}-\hat{\mathbf{P}}) - \int_{t_0}^{t_r} < \mathbf{P}, (\mathbf{X}-\mathbf{X}^{obs}) > dt \tag{123}$$

and the adjoint model equation is

$$\frac{\partial \mathbf{Q}}{\partial t} + [\frac{\partial \mathbf{F}}{\partial x}]^T \mathbf{Q} - \mathbf{P}^T \mathbf{Q} = W(\mathbf{X}-\mathbf{X}^{obs}) \tag{124}$$

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

$$-\mathbf{P}^T \mathbf{Q} \tag{125}$$

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

$$\delta J =< \nabla_P J, \delta \mathbf{P} > \tag{126}$$

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

In a more general set up the cost function is

$$J(\mathbf{X}, \mathbf{P}) = \frac{1}{2}(\mathbf{X} - \mathbf{X}_b)^T B^{-1}(\mathbf{X} - \mathbf{X}_b) \tag{127}$$

$$+\frac{1}{2}(H\mathbf{X} - \mathbf{X}^o)^T O^{-1}(H\mathbf{X} - \mathbf{X}^o)$$

$$+K(\hat{\mathbf{P}} - \mathbf{P}) \tag{128}$$

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

19.2 Typical cost functional for parameter estimation

The cost functional assumes the form

$$J(P) = J_h(P) + J_D(P) + J_f(P) + J_r(P) \tag{129}$$

where $J_h(P)$ - 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 - 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 and

$J_r(P)$ - A Tichonov[361,362] regularization term - which deals with instabilities in values of the parameter estimates that appear to be closely related to noise in measured data. This term smooths the parameter estimates, by

imposing a penalty on oscillations in the parameter estimate.For a modern approach to regularization of discrete ill-posed problems see Hansen (1998) [158].

20 Adjoint parameter estimation with constrained parameters

Here we aim to perform optimal parameter estimation in a variational approach setting, i.e. to obtain an optimal value of the parameter α such that

$$J(\alpha^0) < J(\alpha) \quad \forall \alpha \tag{130}$$

where J is a cost function which measures the discrepancy between the observations and the corresponding model forecast variables. Hence, the optimal parameter can be retrieved by fitting the model forecast fields to the observations. Given constrained parameters, i.e., parameters whose values vary between certain bounds, for instance, when the parameter α satisfies $\alpha \in [a, b]$, here a and b denote the lower and upper boundary respectively, the cost function for parameter estimation may assume the following form:

$$J(\mathbf{X}, \alpha) = \int_{t_0}^{t_R} < W(\mathbf{X} - \mathbf{X}^{obs}), (\mathbf{X} - \mathbf{X}^{obs}) > dt + \lambda g(\alpha) \tag{131}$$

where the vector α denotes the vector of model parameters, λ is the penalty coefficient, $\mathbf{X}$ represents the state variable vector, $\mathbf{X}^{obs}$ the observation vector. The second term consists of a penalty function, which is defined as:

$$g(\alpha) = \begin{cases} \frac{1}{2}(x-b)^2 & \text{if } x \geqslant b \\ 0 \quad \text{if } a < x < b \\ \frac{1}{2}(x-a)^2 & \text{if } x \leqslant a \end{cases} \tag{132}$$

where $g(\alpha)$ is a function only of the violated constraints. The first derivative of this function is:

$$\frac{\partial g(\alpha)}{\partial \alpha} = \begin{cases} (x-b) & \text{if } x \geqslant b \\ 0 \quad \text{if } a < x < b \\ (x-a) & \text{if } x \leqslant a \end{cases} \tag{133}$$

Another type of penalty effective in transforming a constrained optimization problem into an unconstrained one is the barrier method which imposes a penalty for reaching the boundary of an inequality constraint. Typically, we will use a logarithmic barrier function of the form:

$$J(\mathbf{X},\alpha) = \int_{t_0}^{t_R} < W(\mathbf{X}-\mathbf{X}^{obs}), (\mathbf{X}-\mathbf{X}^{obs}) > dt - \mu \sum_{i=1}^{m} \log h_i(\alpha) \qquad (134)$$

where μ is the barrier coefficient and h is the constraint function. The barrier methods are strictly feasible methods, i.e., the iterates lie in the interior of the feasible region, and create a "barrier" keeping iterates away from boundaries of the feasible region (Nash and Sofer)[261]. The choice of the penalty parameters is requiring some care and numerical experience in constrained optimization. The readers are referred to the excellent book of Bertsekas[29]. There are two different purposes for the inclusion of the second term in equations (131) or (134), one being to ensure that the retrieved parameter lies within the boundaries, and a penalty term of the form in Equation (131) or a logarithm of the box constraints as in Equation (134) is efficient. The other purpose is to increase the convexity of the cost function by adding a positive value to the Hessian matrix, thereby increasing its positive-definiteness. Suppose that the forward model is given in the form

$$\frac{\partial \mathbf{X}}{\partial t} = \mathbf{F}(\mathbf{X},\alpha,t) \qquad (135)$$

Its corresponding tangent linear model is defined as

$$\frac{\delta \mathbf{X}}{\partial t} = (\frac{\mathbf{F}(\mathbf{X},\alpha,t)}{\partial \mathbf{X}})\delta\mathbf{X} + (\frac{\mathbf{F}(\mathbf{X},\alpha,t)}{\partial \alpha})\delta\alpha \qquad (136)$$

The adjoint model derived is expressed in the form

$$\frac{\partial P}{\partial t} - (\frac{\partial \mathbf{F}}{\partial \mathbf{X}})^* P = W(\mathbf{X}-\mathbf{X}^{obs}) \qquad (137)$$

where P represents the adjoint variables. The gradients of the cost function with respect to the initial condition and that of the parameter α are assuming the following form,

$$\nabla_{X_0} J = P(0), \qquad (138)$$

$$\nabla_{\alpha} J = \int_{t_0}^{t_R} (\frac{\partial \mathbf{F}}{\partial \mathbf{X}})^* P dt + \lambda \frac{\partial g}{\partial \alpha} \qquad (139)$$

respectively. The adjoint model is of the same form as that where only the initial conditions are considered as the control variables. Hence, the problem of parameter estimation via the adjoint method when the number of parameters to be estimated is small does not result in an additional computational effort. We may expect that the parameter estimation process will provide us with both optimally determined parameters and initial conditions simultaneously. The gradient of the cost function with respect to both the initial conditions and the parameter is written as:

$$\nabla J = (\nabla_{X_0} J, \nabla_\alpha J)^T. \tag{140}$$

21 Illustrative algorithmic flowchart of constrained adjoint parameter estimation (Zhu and Navon (1999)[394]

1. The 6-hr forecasts starting from the initialized analysis at 1800UTC Sept. 2, 1996 was taken as the initial guess of the initial condition.

Given a positive definite initial approximation to the inverse Hessian matrix H_0 (generally taken as the identity matrix I), we integrated the full-physics FSU GSM 6 hours, then calculated the cost function using Equation (131). Since further study needs to be carried out to determine the observational error covariance, we simply take the inverse of the maximum square of the difference between the two time level observations as the weight matrix W in Equation (131) (See Navon, et al.)[271]. The initial guess of α , the penalty coefficient value λ and the upper and lower boundaries where the parameters may vary were specified here.

We then integrate the full-physics adjoint model of the FSU GSM backward in time to obtain the gradient of J with respect to the control variable $\mathbf{Y} = (\mathbf{X}_0, \alpha)^T$,

$$g_0 = g_{\mathbf{Y}} = (\nabla_{X_0} J, \nabla_\alpha J)^T \tag{141}$$

and the search direction

$$d_0 = -H_0 g_0 \tag{142}$$

3. For $k = 0, 1, 2, \cdots$, minimize $J(\mathbf{Y}_k, \beta_k d_k)$ with respect to $\beta \geqslant 0$ to obtain as $\mathbf{Y}_{k+1}$

$$\mathbf{Y}_{k+1} = \mathbf{Y}_k + \beta_k d_k \tag{143}$$

where β_k is a positive scalar, the step-size being obtained by a line search so as to satisfy a sufficient decrease (See Gill, et al.)[143].

4. Compute

$$g_{k+1} = \nabla J(\mathbf{Y}_{k+1}) \tag{144}$$

5. Compute a new search direction

$$d_{k+1} = -H_{k+1} g_{k+1} \tag{145}$$

6. Check whether the solution converges. If the convergence criterion

$$g_{k+1} \leqslant \varepsilon' \mathbf{Max}(1, \mathbf{Y}_{k+1}) \tag{146}$$

is satisfied, where ε' is a user supplied small number, then the algorithm terminates with $\mathbf{Y}_{k+1}$ as the optimal solution; otherwise go back to Step 3.

22 Open problems in adjoint parameter estimation

Some parameters are not easily identifiable (see Navon[274]. Other render the estimation problem ill-posed and require a Tichonov [362] regularization.

The issue is even more difficult in climate models since classical 4-D VAR does not extend to climate models.

Climate models are designed to simulate the full multi-dimensional multiscale complexity of the climate system. Although climate models are based upon the sound principles of conservation of mass, momentum, energy, and water, computational limitations necessitate a finite grid and hence the parameterization of processes that cannot be explicitly resolved. These subgrid parameterizations have limited physical bases, and hence employ a number of parameters whose values are uncertain. The climate simulated by global climate models is therefore sensitive to the values of these model parameters characterized by multiple scales and varying uncertainties. The parameter values can be adjusted to improve a variety of aspects of the climate simulation. Current practice is to adjust values of one parameter at a time to correct problems with one particular aspect of the climate simulation, often with unintended adverse impacts on other aspects of the climate simulation.

Parameter estimation in climate models poses a huge challenge due to importance of decisions related to correct parameter estimation that are affecting global warming results and related issues, see Stainforth et al.[335],

Stocker[340] to cite but a few. Usual methodologies of variational adjoint parameter estimation in NWP do not carry over to climate simulation Lea et al.[192,193] and a combination of approaches is necessary to account for nonlinear feedbacks and uncertainties when tuning multiple multiscale parameters against observations.

23 Model Error in 4-D VAR

Numerical weather prediction (NWP) models are imperfect, since they are discretized, dissipative and dispersion errors arise, and, moreover subgrid processes are not included. In addition, most of the physical processes and their interactions in the atmosphere are parameterized and a complete mathematical modeling of the boundary conditions and forcing terms can never be achieved. Usually all of these modeling drawbacks are collectively addressed by the term model error (ME). The model equations do not represent the system behavior exactly and model errors arise due to lack of resolution as well as inaccuracies occurring in physical parameters, boundary conditions and forcing terms. Errors also occur due to numerical discrete approximations. A way to take these errors into account is to use the weak constraint 4D-Var.

Variational data assimilation is based on the minimization of:

$$J(\mathbf{x}) = [H(\mathbf{x}) - y]^T R^{-1} [H(\mathbf{x}) - y] + (\mathbf{x}_0 - \mathbf{x}_b)^T B^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \Phi(\mathbf{x})^T C^{-1} \Phi(\mathbf{x}) \tag{147}$$

Here $\mathbf{x}$ is the 4D state of the atmosphere over the assimilation window, H is a 4D observation operator, accounting for the time dimension. Φ represents remaining theoretical knowledge after background information has been accounted for (such as balance relations or digital filtering initialization introduced by Lynch and Huang[230]). One can see that model M verified exactly although it is not perfect.

23.1 Weak constraint 4D-Var

The model can be imposed as a constraint in the cost function, in the same way as other sources of information:

$$\Phi_i(\mathbf{x}) = \mathbf{x}_i - \mathbf{M}_{i-1} \tag{148}$$

Model error η is defined as: $\eta_i(\mathbf{x}) = \mathbf{x}_i - \mathbf{M}_{i-1}$.

The cost function becomes:

$$J(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{n}(\mathcal{H}(\mathbf{x}_i) - y_i)^T R_i^{-1}(\mathcal{H}(\mathbf{x}_i) - y_i) + \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T B^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\sum_{i=1}^{n}\eta_i^T Q_i^{-1}\eta \tag{149}$$

Another issue requiring attention is that model error covariance matrix Q has to be defined. Strong constraint 4D-Var is obtained when $\Phi_i(\mathbf{x}) = 0$ i.e. $\eta = 0$ (perfect model).

Studies indicate that model error (ME)can severely impact forecast errors, see for instance Boer[34];Dalcher and Kalnay[79]; Bloom and Shubert[33] and Zupanski[408].

For early methods on estimating modeling errors in operational NWP models see Thiébaux and Morone[359] and Saha[314]. Thus giving up the assumption that the model is perfect, in the context of strong constraint VDA leads us to weak constraint formulation of VDA, and if we include time evolution of the variables, we could say we have a weak constraint 4D-Var (time plus three space dimensions).

Comparing the strong and weak constraint VDA, in the formulation of former, it is assumed that η has mean and model error covariance matrix $Q = E(\eta(t)\eta^T(t')) = 0, \quad \forall\ t$ and t' and model error covariance matrix, $E[\cdot]$ is the mathematical expectation operator. It should be noted that if the mean and (co)variance of a random vector are prescribed to be equal to zero, then all realizations of that random vector are identically equal to zero, thus, $\eta \equiv 0$. Whereas in the weak constraint version of VDA, the mean and covariance of ME have to be specified. However exact statistical details of ME are difficult to obtain (Daley[82,83]; Dee and Da Silva[92] and Zhu and Kamachi[393]) a fact which led researchers to suggest a variety of assumptions to approximate the ME.

Early efforts to model the systematic component of ME were pioneered by Derber[100]. He suggested a simplified approach to model η to be equal to $\lambda(t)\phi$. The temporal part, $\lambda(t)$ is a specified function of time alone, while ϕ is a spatially dependent, control variable. Three different forms of λ were considered, namely, parabolic, delta function and constant in time. It was observed that the parabolic variation of λ provided results comparable to a constant in time λ. Using a similar approach (Wergen[383]; Zupanski[404]) it was shown that inclusion of ME allowed significant reduction in forecast RMSE.

For dynamically evolving systems such as discrete NWP models, ME is expected to depend on the model state and should be evolving in time (Griffith and Nichols[148,150]). Various simple forms of evolution of ME in time were considered by Griffith and Nichols[150,277], At any time step, t_k, the evolution of ME is

$$\eta_k = T_k(\mathbf{e}_k) + \mathbf{q}_k \tag{150}$$

where T_k describes the distribution of systematic errors in the NWP model equations, and $\mathbf{q}_k$, (stochastic component) is an unbiased, serially correlated, normally distributed random vector, with known covariance. The evolution of $\mathbf{e}_k$, is in-turn modeled by assuming that it depends on the state vector,$\mathbf{x}_k$,

$$\mathbf{e}_{k+1} = g_k(\mathbf{x}_k, \mathbf{e}_k). \tag{151}$$

23.2 Systematic Model error and State Augmentation

In order to take into account systematic components in the model errors, we assume that the evolution of the errors is described by the equations

$$\eta_k = T_k(\mathbf{e}_k) + \mathbf{q}_k \tag{152}$$

$$\mathbf{e}_{k+1} = \mathbf{g}_k(\mathbf{x}_k, \mathbf{e}_k) \tag{153}$$

where $\mathbf{q}_k \in R^n$ is unbiased, serially uncorrelated, normally distributed random vectors with known covariance matrices and the vectors $\mathbf{e}_k \in R^r$ represent time-varying systematic components of the model errors. The distribution of the systematic errors in the model equations is defined by the function $T_k : R^r \to R^n$. The functions $\mathbf{g}_k : R^n \times R^r \to R^r$ describing the systematic error dynamics, are to be specified. In practice little is known about the form of the model errors and a simple form for the error evolution that reflects any available knowledge needs to be prescribed. Examples of simple forms of the error evolution includes:

constant bias error: $\mathbf{e}_{k+1} = \mathbf{e}_k,\ T_k = I.$

This choice allows for a constant vector $\mathbf{e} = \mathbf{e}_0$ of unknown parameters to be found, which can be interpreted as statistical biases in the model errors. This form is expected to be appropriate for representing the average errors in source terms or in boundary conditions.

Evolving error: $\mathbf{e}_{k+1} = F_k\mathbf{e}_k,\ T_k = I.$

Here $F_k \in R^{n\times n}$ represents a simplified linear model of the state evolution. This choice is appropriate, for example, for representing discretization error in models that approximate continuous dynamical processes by discrete time systems.

Spectral form: $\mathbf{e}_{k+1} = \mathbf{e}_k,\ T_k = (I, \sin(k/N\tau)I, \cos(k/N\tau)I)$.

In this case the constant vector $\mathbf{e} \equiv \mathbf{e}_0$ is partitioned into three components vectors, $\mathbf{e}^T = (\mathbf{e}_1^T, \mathbf{e}_2^T, \mathbf{e}_3^T)$ and τ is a constant determined by the timescale on which the model errors are expected to vary, for example, a diurnal timescale. The choice approximates the first order terms in a spectral expansion of the model error.

The weak constraint VDA doubles the size of the optimization problem (compared to strong constraint VDA), in addition if the stochastic component is included in the ME formulation, then one would have to save every random realization at each model time step, which amounts to tripling the size of the optimization problem. The computational results in [150] were provided by neglecting $\mathbf{q_k}$, the stochastic component of ME and using the constant and evolving forms of the systematic component, see [150] for additional details. Similar approaches for modeling the systematic component of ME was considered by Martin et al. (2002) and reduction of ME control vector size by projecting it on to the subspace of eigenvectors corresponding to the leading eigenvalues of the adjoint-tangent linear operators was illustrated by Vidard et al. (2000).

Other choices can be prescribed, including piecewise constant error and linearly growing error (see Griffith[151], Martin et al.[237], Griffith et al.[150] and Griffith and Nichols[150]). These techniques have been applied successfully in practice to estimate systematic errors in an equatorial ocean model (Martin et al.[238]) Zupanski et al. (2005)[409] provided results obtained using the NCEP' s regional weather prediction system in weak constraint VDA framework. Akella and Navon (2005)[2] studied in depth the nature of modeling errors and suggested a decreasing, constant and increasing in time forms of ME. Implementation of these forms in a weak constraint VDA framework yielded a further reduction in forecast errors. For highly accurate advection schemes such as the Piecewise Parabolic Method (PPM) scheme, they found that the increasing form of ME is the best (when tested in the framework of a twin experiment). They considered three different forms of ME using high resolution advection schemes in the presence of non-linear advection terms were studied in both strong and weak constraint VDA framework.

When the number of observations is considerably smaller, the method of representers (Bennett 1992)[23] provides a computationally efficient (in storage/ space requirements) formulation of VDA. The incorporation of ME in such a

framework has been shown by Bennett et al. (1993, 1996, and 1997)[22,20,21] and Uboldi and Kamachi (2000)[368].

Very little is known with certainty about ME spatio-temporal structure since MEs are not observable, contrary to forecast errors. The common practice is to assume that MEs are white. Daley (1992) [82]suggested use of a first order (in time) linear model for MEs. That approach was implemented by Zupanski (1997) [406] in its simplest form; the inevitable simplicity is due to the absence of empirical estimates of parameters and even structural features of the ME model. DelSole and Hou (1999)[96] considered the state-dependent part of ME and proposed a respective estimator.

Mitchell and Daley (1997)[246] considered the discretization part of ME and its effect on data assimilation. Menemenlis and Chechelnitsky (2000)[244] estimated the spatial structure of an ME white-noise model for an ocean circulation model. ME models rely on hypotheses that have never been checked namely the applicability of a stochastic model driven by an additive (and not, say, multiplicative) noise, Gaussianity of ME, the white-noise or red-noise hypotheses. Tools needed to use the information on ME (Tsyrulnikov 2005) [367] structure in meteorology and oceanography are available such as ensemble forecasting, weak-constraint four-dimensional variational assimilation (4D-Var, e.g. Zupanski 1997 [406]; Xu et al. 2005) [387], and Kalman filtering (e.g. Cohn 1997) [63]. Empirical approaches have been used only in ensemble techniques but cannot be used in the weak-constraint 4D-Var, where one must specify an ME spatio-temporal stochastic model.

24 Bias correction in 4-D VAR data assimilation

The bias problem in data assimilation is caused by problems with the data being assimilated, approximations in the observation operators used to simulate the data as well as by different limitations of the assimilation model itself (See Dee (2005)[95]). The term bias includes any type of error that is systematic rather than random. Large persistent mean values of analysis increments or regularly recurring spatial structures are a symptom of bias along with detection of monitoring statistics of observed-minus-background residuals for different instruments collected over time.

Dee(2005)[95] proposed to work with an augmented control vector

$$Z^T = [X^T, \beta^T] \tag{154}$$

including model state X as well as a parameter β.

Then the usual 4-D VAR standard variational analysis minimizes the functional

$$J(X) = (X^b - X)^T B^{-1}(X^b - X) + [y - h(X)]^T R^{-1}[y - h(X)] \tag{155}$$

where the function $h(\cdot)$ denotes a set of observation operators used to express relation between model state and observations, is transformed by modifying the observation operator to account for bias

$$\tilde{h}(z) = \hat{h}(X, \beta) \tag{156}$$

We minimize instead

$$J(z) = (z_b - z)^T B_z^{-1}(z_b - z) + [y - \tilde{h}(z)]^T R^{-1}[y - \tilde{h}(z)] \tag{157}$$

As mentioned by Trémolet (2005)[365] variational bias correction of satellite radiances was first implemented at NCEP in the spectral statistical interpolation (SSI) analysis system (Derber and Wu 1998), [103]) and more recently by Dee(2004)[94].

He used

$$B_z = \begin{pmatrix} B_X & 0 \\ 0 & B_\beta \end{pmatrix}. \tag{158}$$

the background error covariances for the bias parameters since one needs to minimize the bias functional. To implement this method one needs the modified operator $\tilde{h}(X, \beta)$ as well as an effective preconditioner for the point minimization problem.

Different aspects of the state of the art of bias and data assimilation are discussed in Dee(2005) [95]. However, developing useful models for bias remains still a challenge. Tsyrulnikov(2005) has started addressing some aspects of the problem by developing advanced stochastic representations of model errors consistent with both spatial and temporal structures of the forecast errors. See also Dee and Todling[93] for modeling bias in background fields by assuming persistence and see also Chepurin et al. (2005)[59]) for forecast model bias correction in ocean data assimilation. See also work of Janjic and Cohn (2006) [170] on treatment of observation error due to unresolved scales in atmospheric data assimilation as an example of the issue of dealing with representativeness errors.

25 Automatic differentiation

Automatic differentiation (AD) is a set to techniques based on the mechanical application of the chain rule to obtain derivatives of a function given as a computer program adjoint equations resulting from differentiation the discretized model equation. Automatic differentiation exploits fact that a computer code executes a sequence of elementary arithmetic operations consisting of additions or elementary functions.

By applying the chain rule derivative repeatedly to these operations derivatives of any order can be computed automatically .Other classical methods that achieve the same goal but with inherent shortcomings are symbolic differentiation or use of finite-differences. Symbolic differentiation is slow, while finite differences suffer from round-off errors in the discretization process and cancellations. Automatic differentiation has the advantage of solving these problems.

There are essentially two modes of AD, namely forward accumulation and reverse accumulation. Forward accumulation is less useful for data assimilation while reverse accumulation allows efficient calculation of gradients.

The first powerful general purpose AD systems was developed at Oak Ridge National Laboratory (Oblow 1983)[281], later endowed with the adjoint variant ADGEN for reverse AD (1987) by Pin et al (1987)[290]. Later ADIFOR (Bischof et al.[31]) was developed at Argonne National Laboratory , Odyssee at INRIA (Rostaing-Schmidt 1993)[312] and TAMC by Giering and Kaminski (1997)[139]. In France the TAPENADE code is used (see Hascoet and Pascual (2004)[159]. There are many many more languages. Earlier books on AD are those by Rall[305] and Kagiwada et al. 1986.[172]

Checkpointing is a general trade-off technique, used in the reverse mode of AD, that trades duplicate execution of a part of the program in order to save memory space employed to save intermediate results. Checkpointing a code fragment amounts to running this fragment without storage of intermediate values, thus saving memory space. At a later stage, when the intermediate value is required, the fragment is run a second time to obtain the required values. Results and application studies of automatic differentiation have been published in proceedings of the international workshop on automatic differentiation held in Breckenridge (See Griewank and Corliss (1991). The most comprehensive book and work is that of Andreas Griewank (Berz et al.[30]), Griewank and Corliss[147] and the comprehensive book of Griewank[149]).

26 Second Order Adjoint methods

Behind most of the methods used in meteorology such as: optimal interpolation, variational methods, statistical estimation etc., there is a variational principle, i.e. the retrieved fields are obtained through minimization of a functional depending on the various sources of information. The retrieved fields are obtained through some optimality condition which can be an Euler or Euler-Lagrange condition if regularity conditions are satisfied. Since these conditions are first order conditions, it follows that they involve the first order derivatives of the functional which is minimized. In this sense, data assimilation techniques are first order methods. But first order methods provide only necessary conditions for optimality but not sufficient ones. Sufficient conditions require second order information. By the same token, from the mathematical point of view sensitivity studies with respect to some parameter can be obtained through Gateaux derivatives with respect to this parameter. Therefore if we seek the sensitivity of fields which have already been defined through some first order conditions we will have to go to an order of derivation higher and in this sense sensitivity studies require second order information.

Early work on second order information in meteorology includes Thacker[344] followed by work of Wang et al.[375,376] stimulated by advice and expertise of F.X. Le Dimet, Wang[376]. Wang et al.[378]and Wang et al.[380] considered use of second order information for optimization purposes namely to obtain truncated -Newton and Adjoint Newton algorithms using exact Hessian/vector products obtained via second order adjoint. Application of these ideas was presented in Wang et al.[379]. Kalnay et al.[175] introduced an elegant and novel pseudo-inverse approach and showed its connection to the adjoint Newton algorithm of Wang et al.[379]. (See Kalnay et al.[175], Pu and Kalnay[296], Park and Kalnay[287], Pu et al.[295]). Ngodock[276] applied second order information in his doctoral thesis in conjunction with sensitivity analysis in the presence of observations and applied it to the ocean circulation. Le Dimet et al.[201] presented the basic theory for second order adjoint analysis related to sensitivity analysis.

A comprehensive review paper on second order adjoint methods was written by Le Dimet et al.(2002) [202] considering all aspects of second order adjoint methods.

27 Computing the second order information

In what follows we follow closely the presentation in Le Dimet et al. (2002)[202].

In general we will assume that the model has the general form:

$$F(\mathbf{X}, \mathbf{U}) = 0 \tag{159}$$

where $\mathbf{X}$, the state variable, describes the state of the environment, $\mathbf{U}$ is the input of the model, i.e. an initial condition which has to be provided to the model to obtain from Eq. (159) a unique solution $\mathbf{X}(\mathbf{U})$. We will assume that $\mathbf{X}$ and $\mathbf{U}$ belong to a space equipped with an inner product.

The closure of the model is obtained through a variational principle which can be considered as the minimization of some functional:

$$J(\mathbf{X}, \mathbf{U}) \tag{160}$$

For instance, in the case of variational data assimilation, J may be viewed as representing the cost function measuring the discrepancy between the observation and the solution associated with the value $\mathbf{U}$ of the input parameter. Therefore the optimal input for the model will minimize J.

27.1 First order necessary conditions

If the optimal $\mathbf{U}$ minimizes J , then it satisfies the Euler equations given by

$$\nabla J(\mathbf{U}) = 0 \tag{161}$$

where ∇J is the gradient of J with respect to control variables.

The gradient of J is obtained in the following way:

(i) we compute the Gateaux (directional) derivative of the model and of F in some direction $\mathbf{u}$. We may write

$$\frac{\partial F}{\partial \mathbf{X}} \times \hat{\mathbf{X}} + \frac{\partial F}{\partial \mathbf{U}} \times \mathbf{u} = 0 \tag{162}$$

where $\hat{()}$ stands for the Gâteaux derivative. Let Z be an application from R^n into R^n with variable $\mathbf{U}$. We define the Gâteaux derivative of Z in the direction $\mathbf{u}$ when this limit exists. For a generic function Z it is given by:

$$\hat{Z}(\mathbf{U}, \mathbf{u}) = \lim_{\alpha \to 0} \frac{Z(\mathbf{U} + \alpha \mathbf{u}) - Z(\mathbf{U})}{\alpha} \tag{163}$$

If $\hat{Z}(\mathbf{U}, \mathbf{u})$ is linear in $\mathbf{u}$ we can write

$$\hat{Z}(\mathbf{U}, \mathbf{u}) =< \nabla Z(\mathbf{U}), \mathbf{u} > \tag{164}$$

where ∇Z is the gradient of Z with respect to $\mathbf{U}$. The Gateaux derivative is also called a directional derivative. Here $\frac{\partial F}{\partial \mathbf{X}}$ (or $\frac{\partial F}{\partial \mathbf{U}}$) is the Jacobian of F with respect to $\mathbf{X}$ (or $\mathbf{U}$)) and

$$\hat{J}(\mathbf{X}, \mathbf{U}, \mathbf{u}) =< \frac{\partial J}{\partial \mathbf{X}}, \hat{\mathbf{X}} > + < \frac{\partial J}{\partial \mathbf{U}}, \mathbf{u} > \tag{165}$$

where $< >$ stands for the inner product.

The gradient of J is obtained by exhibiting the linear dependence of $\hat{J}$ with respect to $\mathbf{u}$. This is done by introducing the adjoint variable P (to be defined later according to convenience).

Taking the inner product between (162) and P yields

$$< \frac{\partial F}{\partial \mathbf{X}} \times \hat{\mathbf{X}}, P > + < \frac{\partial F}{\partial \mathbf{U}} \times \mathbf{u}, P >= 0 \tag{166}$$

$$< (\frac{\partial F}{\partial \mathbf{X}})^{T} \times P, \hat{\mathbf{X}} > + < (\frac{\partial F}{\partial \mathbf{U}})^{T} \times P, \mathbf{u} >= 0 \tag{167}$$

Therefore using (165), if P is defined as the solution of the adjoint model

$$(\frac{\partial F}{\partial \mathbf{X}})^{T} \times P = \frac{\partial J}{\partial \mathbf{X}} \tag{168}$$

then we obtain

$$\nabla J(\mathbf{U}) = (\frac{\partial F}{\partial \mathbf{U}})^{T} \times P + \frac{\partial J}{\partial \mathbf{U}} \tag{169}$$

Therefore the gradient is computed by solving Eq. (168) to obtain , then by applying Eq. (169).

27.2 Second order adjoint

To obtain second order information we look for the product of the Hessian $G(\mathbf{U})$ of J with some vector $\mathbf{u}$. As before we apply a perturbation to Eqs. (159), (168), and from Eq. (168) and (169) we obtain

$$(\frac{\partial^2 F}{\partial \mathbf{X}^2} \times \hat{\mathbf{X}} + \frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{U}} \times \mathbf{u})^T \times P + (\frac{\partial F}{\partial \mathbf{X}})^T \times \hat{P} = \frac{\partial^2 J}{\partial \mathbf{X}^2} \times \hat{\mathbf{X}} + \frac{\partial^2 J}{\partial \mathbf{X} \partial \mathbf{U}} \times \mathbf{u} \tag{170}$$

and

$$\widehat{\nabla J(\mathbf{U})} = G(\mathbf{U}) \times \mathbf{u} = -(\frac{\partial^2 F}{\partial \mathbf{U}^2} \times \mathbf{u} + \frac{\partial^2 F}{\partial \mathbf{U} \partial \mathbf{X}} \times \hat{\mathbf{X}})^T \times P - (\frac{\partial F}{\partial \mathbf{U}})^T \times \hat{P} + \frac{\partial^2 J}{\partial \mathbf{U}^2} \times \mathbf{u} + \frac{\partial^2 J}{\partial \mathbf{X} \partial \mathbf{U}} \times \hat{\mathbf{X}} \tag{171}$$

We introduce here Q and R, two additional variables. To eliminate $\hat{\mathbf{X}}$ and P, we will take the inner product of Eq. (162) and (170) with Q and R respectively, then add the results. We then obtain

$$\begin{aligned} &< \hat{\mathbf{X}}, (\frac{\partial F}{\partial \mathbf{X}})^T \times Q > + < \mathbf{u}, (\frac{\partial F}{\partial \mathbf{U}})^T \times Q > + < P, (\frac{\partial^2 F}{\partial \mathbf{X}^2}) \times \hat{\mathbf{X}} \times R > \\ &+ < P, (\frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{U}}) \times \mathbf{u} \times R > + < \hat{P}, (\frac{\partial F}{\partial \mathbf{X}}) \times R > \\ &= < \hat{\mathbf{X}}, (\frac{\partial^2 J}{\partial \mathbf{X}^2})^T \times R > + < \mathbf{u}, (\frac{\partial^2 J}{\partial \mathbf{X} \partial \mathbf{U}})^T \times R > \end{aligned} \tag{172}$$

Let us take the inner product of Eq. (171) with $\mathbf{u}$, then we may write

$$\begin{aligned} < G(\mathbf{U}) \times \mathbf{u}, \mathbf{u} > = &< -(\frac{\partial^2 F}{\partial \mathbf{U}^2} \times \mathbf{u} + \frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{U}} \times \hat{\mathbf{X}})^T \times P, \mathbf{u} > + \\ &< \hat{P}, (-\frac{\partial F}{\partial \mathbf{U}}) \times \mathbf{u} > < \frac{\partial^2 J}{\partial \mathbf{U}^2} \times \mathbf{u}, \mathbf{u} > + < \hat{\mathbf{X}}, \frac{\partial^2 J}{\partial \mathbf{X} \partial \mathbf{U}})^T \times \mathbf{u} > \end{aligned} \tag{173}$$

From (173) we get

$$\begin{aligned} &< \hat{\mathbf{X}}, (\frac{\partial F}{\partial \mathbf{X}})^T \times Q + (\frac{\partial^2 F}{\partial \mathbf{X}^2} \times P) \times R - \frac{\partial^2 J}{\partial \mathbf{X}^2} \times R > + < \hat{P}, \frac{\partial F}{\partial \mathbf{X}} \times R > = \\ &< \mathbf{u}, -(\frac{\partial F}{\partial \mathbf{U}})^T \times Q - (\frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{U}} \times P)^T \times R + \frac{\partial^2 J}{\partial \mathbf{X} \partial \mathbf{U}} \times R > \end{aligned} \tag{174}$$

Therefore if Q and R are defined as being the solution of

$$(\frac{\partial F}{\partial \mathbf{X}})^T < \mathbf{u}, -(\frac{\partial F}{\partial \mathbf{U}})^T \times Q + (\frac{\partial^2 F}{\partial \mathbf{X}^2} < \mathbf{u}, -(\frac{\partial F}{\partial \mathbf{U}})^T \times P) \times R - \quad (175)$$
$$(\frac{\partial^2 J}{\partial \mathbf{X}^2})^T \times R = (\frac{\partial^2 J}{\partial \mathbf{X} \partial \mathbf{U}})^T \times \mathbf{u} - (\frac{\partial^2 F}{\partial \mathbf{U} \partial \mathbf{X}} \mathbf{u}) \times P$$

$$(\frac{\partial F}{\partial \mathbf{X}}) \times R = -\frac{\partial F}{\partial \mathbf{U}} \times \mathbf{u} \quad (176)$$

then we obtain:

$$G(\mathbf{U}) \times \mathbf{u} = -(\frac{\partial^2 F}{\partial \mathbf{U}^2} \times \mathbf{u}) \times P + \frac{\partial^2 J}{\partial \mathbf{U}^2} \times \mathbf{u} - (\frac{\partial F}{\partial \mathbf{U}})^T \times Q - \quad (177)$$
$$(\frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{U}} \times P) \times R + \frac{\partial^2 J}{\partial \mathbf{X} \partial \mathbf{U}} \times R$$

For equations(168-174) we took into account the symmetry of the matrix of second derivative, e.g.

$$\frac{\partial^2 F}{\partial \mathbf{X}^2} = (\frac{\partial^2 F}{\partial \mathbf{X}^2})^T \quad (178)$$

leading to some simplifications. The system (174-174) will be called the second order adjoint. Therefore we can obtain the product of the Hessian by a vector $\mathbf{u}$ by (i) solving the system (174-174). (ii) applying formula (177).

27.3 Remarks

a) The system (174-174) which has to be solved to obtain the Hessian/vector product can be derived from the Gateaux derivative (174) which is the same as (176). In the literature, the system (174-174) is often called the tangent linear model, this denomination being rather inappropriate because it implies the issue of linearization and the subsequent notion of range of validity which is not relevant in the case of a derivative.

b) In the case of an N-finite dimensional space the Hessian can be fully computed after N integrations of vector of the canonical base. Equation 174 differs from the adjoint model by the forcing terms which will depend on $\mathbf{u}$ and R.

c) The system (174-177) will yield the exact value of the Hessian/vector product. An approximation could be obtained by using the standard finite differences, i.e.,

$$G(\mathbf{U}) \times \mathbf{u} \approx \frac{1}{\alpha}[\nabla J(\mathbf{U} + \alpha \mathbf{u}) - \nabla J(\mathbf{U})] \quad (179)$$

where α is the finite-difference interval which has to be carefully chosen. In the incremental 3/4D-Var approach the Hessian/vector product can readily be obtained by differencing two gradients.

However several integrations of the model and of its adjoint model will be necessary in this case to determine the range of validity of the finite-difference approximation (Wang[378] and references therein).

27.4 Time dependent model

In the case of variational data assimilation the model F is a differential system on the time interval $[0, T]$. The evolution of $\mathbf{X} \in H \sqsubset [C(0,T)]^n$ between 0 and T is governed by the differential system,

$$\frac{\partial \mathbf{X}}{\partial t} = F(\mathbf{X}) + \mathbf{B} \times \mathbf{V} \tag{180}$$

The input variable is often the initial condition,

$$\mathbf{X}(0) = \mathbf{U} \in R^n \tag{181}$$

In this system F is a nonlinear operator which describes the dynamics of the model, $\mathbf{V} \in V \sqsubset [C(0,T)]^m$ is a term used to represent the uncertainties of the model which we assume to be linearly coupled through the (m, n) - dimensional matrix $\mathbf{B}$,$\mathbf{U}$ is the initial condition, and the criteria J is the discrepancy between the solution of (180)-(181) and observations

$$J(\mathbf{U}, \mathbf{V}) = \frac{1}{2} \int_0^T ||\mathbf{HX} - \mathbf{X}_{obs}||^2 dt \tag{182}$$

where $\mathbf{H}$ is the observation matrix, i.e., a linear operator mapping $\mathbf{X}$ into $\mathbf{X_{obs}}$. The problem consists in determining $\mathbf{U}$ and $\mathbf{V}$ that minimize J.

A perturbation $\mathbf{v}$ on $\mathbf{V}$ and $\mathbf{u}$ on $\mathbf{U}$ gives $\hat{\mathbf{X}}$ and $\hat{\mathbf{J}}$ the Gateaux derivatives of $\mathbf{X}$ and $\mathbf{J}$ as solution of

$$\frac{d\hat{\mathbf{X}}}{dt} = \frac{\partial F}{\partial \mathbf{X}} \times \hat{\mathbf{X}} + \mathbf{B} \times \mathbf{V} \tag{183}$$

$$\hat{\mathbf{X}}(0) = \mathbf{u} \tag{184}$$

$$\hat{J}(\mathbf{U}, \mathbf{V}, \mathbf{u}, \mathbf{v}) = \frac{1}{2} \int_0^T < \mathbf{HX} - \mathbf{X}_{obs}, \mathbf{H}\hat{\mathbf{X}} > dt \tag{185}$$

Let us introduce P the adjoint variable, we take the product of (183) with P after a summation on the interval $[0, T]$ and an integration by parts followed by identification of linearities with respect to $\mathbf{U}$ and $\mathbf{V}$ in (185), we conclude that of P is defined as the solution of

$$\frac{dP}{dt} = \frac{\partial F}{\partial \mathbf{X}}^T \times P + \mathbf{H}^T\mathbf{H}(\mathbf{X} - \mathbf{X}_{obs}) \tag{186}$$

$$P(T) = 0 \tag{187}$$

and the components of the gradient ∇J with respect to $\mathbf{U}$ and $\mathbf{V}$ are

$$\nabla J_{\mathbf{U}} = -P(0) \tag{188}$$

$$\nabla J_{\mathbf{V}} = -\mathbf{B}^T P \tag{189}$$

$\mathbf{V}$ is time dependent, its associated adjoint variable Q will be also time dependent. Let us remark that the gradient of J with respect to $\mathbf{V}$ will depend on time . From a computational point of view the discretization of $\mathbf{V}$ will have to be carried out in such a way that the discretized variable remains in a space of "reasonable" dimension.

The second derivative will be derived after a perturbation h on the control variables $\mathbf{U}$ and $\mathbf{V}$

$$h = \begin{pmatrix} h_U \\ h_V \end{pmatrix} \tag{190}$$

The Gateaux derivatives $\hat{\mathbf{X}}$, P of $\mathbf{X}$ and P in the direction of h, are obtained as the solution of the coupled system

$$\frac{d\hat{\mathbf{X}}}{dt} = \frac{\partial F}{\partial \mathbf{X}}\hat{\mathbf{X}} + \mathbf{B}h_V \tag{191}$$

$$\hat{\mathbf{X}}(0) = h_U \tag{192}$$

$$\frac{d\hat{\mathbf{P}}}{dt} + (\frac{\partial^2 F}{\mathbf{X}^2} \times \hat{\mathbf{X}})^T \times P + (\frac{\partial F}{\partial \mathbf{X}})^T \times P = \mathbf{H}^T\mathbf{H}\hat{\mathbf{X}} \tag{193}$$

$$\nabla J_{\mathbf{U}} = -\hat{P}(0) \tag{194}$$

$$\nabla J_{\mathbf{V}} = -\mathbf{B}^T\hat{P} \tag{195}$$

We introduce Q and R, second order adjoint variables. They will be defined later for ease use of presentations.

Taking the inner product of (191) with Q and of (193) with R, integrating from 0 to T, then adding the resulting equations, we may write:

$$\int_0^T [< \frac{d\hat{\mathbf{X}}}{dt}, Q > - < \frac{\partial F}{\partial \mathbf{X}} \times \hat{\mathbf{X}}, Q > - < \mathbf{B}h_V, Q > + < \frac{d\hat{P}}{dt}, R > + \quad (196)$$
$$< [\frac{\partial^2 F}{\partial \mathbf{X}^2} \times \hat{\mathbf{X}}]^T \times P, R > + < [\frac{\partial F}{\partial \mathbf{X}}]^T \hat{P}, R > - < \mathbf{H}^T \mathbf{H}\hat{\mathbf{X}}, R >]dt = 0$$

The terms in $\hat{\mathbf{P}}$ and $\hat{\mathbf{X}}$ are collected and after integration by parts and some additional transformations we obtain

$$\int_0^T < \hat{\mathbf{X}}, -\frac{dQ}{dt} - [\frac{\partial F}{\partial \mathbf{X}}]^T \times Q + [\frac{\partial^2 F}{\partial \mathbf{X}^2} \times P]^T \times R - \mathbf{H}^T \mathbf{H} R > dt + \quad (197)$$
$$\int_0^T < \hat{P}, -\frac{dR}{dt} + (\frac{\partial F}{\partial \mathbf{X}}) \times R > dt - \int_0^T < h_V, \mathbf{B}^T \times Q > dt +$$
$$< \hat{\mathbf{X}}(T), Q(T) > - < \hat{\mathbf{X}}(0), Q(0) > + < \hat{P}(T), R(T) > -$$
$$< \hat{P}(0), R(0) >= 0$$

Let $\mathbf{G}$ be the Hessian matrix of the cost J. We have

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{UU} & \mathbf{G}_{UV} \\ \mathbf{G}_{VU} & \mathbf{G}_{VV} \end{pmatrix} \quad (198)$$

Therefore if we define the second order adjoint as being the solution of

$$\frac{dQ}{dt} + [\frac{\partial F}{\partial \mathbf{X}}]^T \times Q = [\frac{\partial^2 F}{\partial \mathbf{X}^2} P]^T \times R - \mathbf{H}^T \mathbf{H} R \quad (199)$$

$$\frac{dR}{dt} = [\frac{\partial F}{\partial \mathbf{X}}] \times R \quad (200)$$

and

$$Q(T) = 0 \quad (201)$$

$$R(0) = h_U \quad (202)$$

then we finally obtain

$$< -h_U, Q(0) >=< \hat{P}(0), R(0) > \quad (203)$$

$$\hat{P}(0) = -Q(0) \tag{204}$$

We would like to point out that Eq. (204) follows directly from Eq. (203) by using Eq. (202). The product of the Hessian by a vector r is obtained exactly by a direct integration of (200) and (202) followed by a backward integration in time of (199) and (201).

One can obtain $\mathbf{G}$ by n integrations of the differential system:

$$\frac{dQ}{dt} + [\frac{\partial F}{\partial \mathbf{X}}]^T \times Q = [\frac{\partial^2 F}{\partial \mathbf{X}^2} \times P]^T \times R - \mathbf{H}^T \mathbf{H} R \tag{205}$$

$$\frac{dR}{dt} = [\frac{\partial F}{\partial \mathbf{X}}] R \tag{206}$$

with the conditions

$$Q(T) = 0 \tag{207}$$

$$R(0) = \mathbf{e}_i \tag{208}$$

where $\mathbf{e}_i$ are the n-vectors of R^n the canonical base of thus obtaining

$$\mathbf{G}_{UU} \mathbf{e}_i = Q(0) \tag{209}$$

$$\mathbf{G}_{UV} \mathbf{e}_i = \mathbf{B}^T \times Q \tag{210}$$

One then integrates m times the differential system

$$\frac{dQ}{dt} + [\frac{\partial F}{\partial \mathbf{X}}]^T \times Q = [\frac{\partial^2 F}{\partial \mathbf{X}^2} \times P]^T \times R - \mathbf{H}^T \mathbf{H} R \tag{211}$$

$$\frac{dR}{dt} - [\frac{\partial F}{\partial \mathbf{X}}] \times R = \mathbf{f}_j \tag{212}$$

with initial and terminal conditions

$$Q(T) = 0 \tag{213}$$

$$R(0) = 0 \tag{214}$$

where $\mathbf{f}_j$ are the m canonical base vectors of R^m obtaining

$$\mathbf{G}_{VV} \times \mathbf{f}_j = \mathbf{B}^T \times Q \tag{215}$$

The system defined by these equations is the second order adjoint model. The Hessian matrix is obtained via $n + m$ integrations of the second order adjoint. The second order adjoint is easily obtained from the first order adjoint - differing from it only by some forcing terms, in particular the second order term. The second equation is that of the linearized model (the tangent linear model).

One can also obtain the product of a vector of the control space, times the Hessian at cost of a single integration of the second order adjoint.

27.5 Example: The shallow-water equations

The shallow-water equations (SWE) represent the flow of an incompressible fluid whose depth is small with respect to the horizontal dimension. The SWE can be written in a Cartesian system

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + \frac{\partial \phi}{\partial x} = 0, \tag{216}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + \frac{\partial \phi}{\partial y} = 0, \tag{217}$$

$$\frac{\partial \phi}{\partial t} + \frac{\partial (u\phi)}{\partial x} + \frac{\partial (v\phi)}{\partial y}, \tag{218}$$

In this system of equations $X = (u, v, \phi)^T$ is the state variable, u and v are the components of the horizontal velocity, ϕ is the geopotential and f the Coriolis parameter. We aim to present this example in order to provide a didactic setup, thus we will make the strongest simplifications.

a) We neglect the model error which following the previous notations implies $\mathbf{B} \equiv 0$. We only control the initial conditions.

b) We impose periodic boundary conditions.

c) The observations are assumed continuous in both space and time, which is tantamount to assume $\mathbf{H} \equiv \mathbf{I}$, where $\mathbf{I}$ is the identity operator. Let $\mathbf{U}_0 = (u_0, v_0, \phi_0)^T$, i.e., the initial condition, then the cost function assume the form

$$J(\mathbf{U}_0) = \frac{1}{2}\int_0^T [(u - u_{obs})^2 + (v - v_{obs})^2 + \gamma(\phi - \phi_{obs})^2]dt \tag{219}$$

where γ is a non-unit weighting term.

We derive directly the tangent linear model (TLM). The barred variables $\bar{\mathbf{X}} = (\bar{u}, \bar{v}, \bar{\phi})^T$ are the directional derivatives in the direction of the perturbation $\mathbf{h} = (\mathbf{h}_u, \mathbf{h}_v, \mathbf{h}_\phi)^T$ applied to the initial condition and we obtain

$$\frac{\partial \bar{u}}{\partial t} + u\frac{\partial \bar{u}}{\partial x} + \bar{u}\frac{\partial u}{\partial x} + v\frac{\partial \bar{u}}{\partial y} + \bar{v}\frac{\partial u}{\partial y} - f\bar{v} + \frac{\partial \bar{\phi}}{\partial x} = 0, \tag{220}$$

$$\frac{\partial \bar{v}}{\partial t} + u\frac{\partial \bar{v}}{\partial x} + \bar{u}\frac{\partial v}{\partial x} + v\frac{\partial \bar{v}}{\partial y} + \bar{v}\frac{\partial v}{\partial y} + f\bar{u} + \frac{\partial \bar{\phi}}{\partial y} = 0, \tag{221}$$

$$\frac{\partial \bar{\phi}}{\partial t} + \frac{\partial(\bar{u}\phi)}{\partial x} + \frac{\partial(u\bar{\phi})}{\partial x} + \frac{\partial(\bar{v}\phi)}{\partial y} + \frac{\partial(v\bar{\phi})}{\partial y} = 0, \tag{222}$$

By transposing the TLM we obtain the adjoint model. Let $\mathbf{P} = (\tilde{u}, \tilde{v}, \tilde{\phi})^T$ be the adjoint variable, then the adjoint model satisfies

$$\frac{\partial \tilde{u}}{\partial t} + u\frac{\partial \tilde{u}}{\partial x} + v\frac{\partial \tilde{v}}{\partial y} + \tilde{u}\frac{\partial v}{\partial y} - \tilde{v}\frac{\partial v}{\partial y} - f\tilde{v} + \phi\frac{\partial \tilde{\phi}}{\partial x} = u_{obs} - u, \tag{223}$$

$$\frac{\partial \tilde{v}}{\partial t} + \tilde{u}\frac{\partial u}{\partial y} - u\frac{\partial \tilde{v}}{\partial x} + \tilde{v}\frac{\partial u}{\partial x} + v\frac{\partial \tilde{v}}{\partial y} + f\tilde{u} + \phi\frac{\partial \tilde{\phi}}{\partial y} = v_{obs} - v, \tag{224}$$

$$\frac{\partial \tilde{\phi}}{\partial t} + \frac{\partial \tilde{u}}{\partial x} + \frac{\partial \tilde{v}}{\partial y} + u\frac{\partial \tilde{\phi}}{\partial y} + v\frac{\partial \tilde{\phi}}{\partial y} = \gamma(\phi_{obs} - \phi), \tag{225}$$

To obtain the second order model we linearize the couple direct model and adjoint model, we then transpose and obtain the second order adjoint variable $\mathbf{Q} = (\hat{u}, \hat{v}, \hat{\phi})^T$ and the variable $\mathbf{R} = (\bar{u}, \bar{v}, \bar{\phi})^T$ defined by the TLM.

$$\frac{\partial \hat{u}}{\partial t} + u\frac{\partial \hat{u}}{\partial x} + v\frac{\partial \hat{v}}{\partial y} + \hat{u}\frac{\partial v}{\partial y} - \hat{v}\frac{\partial v}{\partial y} - f\hat{v} + \phi\frac{\partial \hat{\phi}}{\partial x} = \tilde{v}\frac{\partial \bar{v}}{\partial x} - \bar{u}\frac{\partial \tilde{u}}{\partial x} - \bar{v}\frac{\partial \bar{u}}{\partial y} + \tilde{u}\frac{\partial \bar{v}}{\partial y} - \bar{\phi}\frac{\partial \tilde{\phi}}{\partial x} - \bar{u}, \tag{226}$$

$$\frac{\partial \hat{v}}{\partial t} + \hat{u}\frac{\partial u}{\partial y} - u\frac{\partial \hat{v}}{\partial x} + \hat{v}\frac{\partial u}{\partial x} + v\frac{\partial \hat{v}}{\partial y} + f\hat{u} + \phi\frac{\partial \hat{\phi}}{\partial y} = \tilde{u}\frac{\partial \bar{u}}{\partial x} - \bar{u}\frac{\partial \tilde{v}}{\partial x} - \tilde{v}\frac{\partial \bar{u}}{\partial y} + \bar{u}\frac{\partial \tilde{v}}{\partial y} - \bar{\phi}\frac{\partial \tilde{\phi}}{\partial y} - \bar{v}, \tag{227}$$

$$\frac{\partial \hat{\phi}}{\partial t} + \frac{\partial \hat{u}}{\partial x} + \frac{\partial \hat{v}}{\partial y} + u\frac{\partial \hat{\phi}}{\partial y} + v\frac{\partial \phi}{\partial y} = -\bar{u}\frac{\partial \tilde{\phi}}{\partial x} - \bar{v}\frac{\partial \tilde{\phi}}{\partial y} - \gamma\bar{\phi}, \tag{228}$$

We see that formally the first and second order adjoint models differ only by second order terms, which contain the adjoint variables. The calculation of second order derivatives requires the storage of the model trajectory, the tangent linear model, and the adjoint model.

27.6 Use of Hessian of cost functional to estimate error covariance matrices

A relationship exists between the inverse Hessian matrix and the analysis error covariance matrix of either 3-D VAR or 4-D VAR (See Thacker[344], Rabier and Courtier[298], Yang et al.[390], Le Dimet et al.[201]).

Following Courtier et al.[74] we consider methods for estimating the Hessian in the weakly nonlinear problem when the tangent linear dynamics is a good approximation to nonlinear dynamics. As a consequence the cost function is near to being quadratic. If as Gauthier[127] we consider the observations as random variables and we look at variational analysis as attempting to solve the minimization problem

$$\min J(\mathbf{v}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(\mathbf{Hx}-\mathbf{y})^T\mathbf{O}^{-1}(\mathbf{Hx}-\mathbf{y}) \qquad (229)$$

where $\mathbf{x}_b$ is the unbiased background field and $\mathbf{y}$ the set of unbiased observations, both being realizations of random variables of covariances $\mathbf{B}$ and $\mathbf{O}$ respectively and where the operator $\mathbf{H}$ computes the model equivalent $\mathbf{Hx}$ of the observation $\mathbf{y}$. Then the Hessian J'' of the cost function J at the minimum is given by

$$J'' = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{O}^{-1}\mathbf{H} \qquad (230)$$

obtained by differentiating (229) twice.

Moreover the analysis error covariance matrix is the inverse of the Hessian as shown in Appendix B of Rabier and Courtier[298]. Calling $\mathbf{x_a}$ the result of the minimization (i.e. the analysis) and $\mathbf{x_t}$ the truth, one has

$$\mathbf{E}[(x_a - x_t)(x_a - x_t)^T] = (J'')^{-1} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{O}^{-1}\mathbf{H})^{-1} \qquad (231)$$

A requirement is that the background error and the observation error are uncorrelated (Rabier and Courtier[298], Fisher and Courtier[115]). See also work of Thepaut and Moll[355] pointing out that the diagonal of the Hessian is optimal among all diagonal preconditioners.

28 Hessian Singular Vectors (HSV)

Computing HSV's uses the full Hessian of the cost function in the variational data assimilation which can be viewed as an approximation of the inverse of the analysis error covariance matrix and it is used at initial time to define a norm. The total energy norm is still used at optimization time. See work by Barkmeijer et al.[13,14]. The HSV's are consistent with the 3-D VAR estimates of the analysis error statistics. In practice one never knows the full 3-D VAR Hessian in its matrix form and a generalized eigenvalue problem has to be solved as described below.

The HSV's are also used in a method first proposed by Courtier [72] and tested by Rabier et al.[302] for the development of a simplified Kalman filter fully described by Fisher[116] and compared with a low resolution explicit extended Kalman filter by Ehrendorfer and Bouttier[108].

Let $\mathbf{M}$ be the propagator of the tangent linear model,$\mathbf{P}$ a projection operator setting a vector to zero outside a given domain. Consider positive-definite and symmetric operators including a norm at initial and optimization time respectively. Then the SV's defined by

$$\frac{< \mathbf{P}\varepsilon(t), \mathbf{EP}\varepsilon(t) >}{< \varepsilon(t_0), \mathbf{C}\varepsilon(t_0) >} \tag{232}$$

under an Euclidean norm are solution of generalized eigenvalue problem.

$$\mathbf{M}^*\mathbf{P}^*\mathbf{EPMx} = \lambda\mathbf{Cx} \tag{233}$$

In HSV, the operator $\mathbf{C}$ is equal to the Hessian of the 3-D Var cost function. As suggested by Barkmeijer et al.[13], one can solve (130) by using the generalized eigenvalue algorithm (Davidson [89]). See also Sleijpen and Van der Vorst[333]. Using

$$\mathbf{C} \equiv \nabla^2 J = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{O}^{-1}\mathbf{H} \tag{234}$$

and carrying out a coordinate transformation

$$\mathbf{x} = \mathbf{L}^{-1}\mathbf{x},\ \mathbf{L}^{-1}\mathbf{L} = \mathbf{B} \tag{235}$$

Then we obtain a transformed operator

$$(\mathbf{L}^{-1})^T\mathbf{CL} \tag{236}$$

and the Hessian becomes equal to the sum of identity and a matrix with rank less or equal to the dimensions of the vector of observations (Fisher and Courtier(1995)[115]). Veerse[371] proposes to take advantage of this form of the appropriate Hessian in order to obtain approximations of the inverse analysis error covariance matrix, using the limited memory inverse BFGS minimization algorithm.

Let $\mathbf{H}$ be $(\nabla^2 J)^{-1}$ the inverse Hessian and $\mathbf{H}^+$ the updated version of the inverse Hessian.

$$\mathbf{s} = \mathbf{x}^{n+1} - \mathbf{x}^n \tag{237}$$

where $\mathbf{s}$ is the difference between the new iterate and the previous one in a limited-memory quasi-Newton minimization procedure.

$$\mathbf{y} = \mathbf{g}^{n+1} - \mathbf{g}^n \tag{238}$$

is the corresponding gradient increment. One has the formula

$$\mathbf{H}^+ = \mathbf{U}(\mathbf{H}, \mathbf{y}, \mathbf{s}) = (\mathbf{I} - \frac{\mathbf{s} \bigotimes \mathbf{y}}{< \mathbf{y}, \mathbf{s} >}) \frac{\mathbf{s} \bigotimes \mathbf{s}}{< \mathbf{y}, \mathbf{s} >} \tag{239}$$

where $<,>$ is a scalar product with respect to which the gradient is defined and $\bigotimes$ stands for the outer product.

The method is implemented by using the inverse Hessian matrix-vector product built in the minimization code and based on Nocedal's[278] algorithm. These methods are useful when the second order adjoint method is not available due to either memory or CPU limitations.

29 4-D VAR status today

4-D VAR data assimilation is available and implemented today at several operational numerical weather prediction centers starting with European Centre for Medium-Range Weather Forecasts (ECMWF), (Rabier et al.[303]), Klinker et al. [178], while a similar system was operational at Meteo-France in 2000 (Janiskova et al.[168], Gauthier and Thépaut[129], Desroziers et al.[106]). More recently 4-D VAR was implemented at UK Met office, Japan and Canada.

The impact of adopting 4-D VAR was qualified as a substantial, resulting in an improvement in NWP quality and accuracy(see Rabier[304] and see the special Issue of QJRMS 2005). 4-D VAR combined with improvement in error

specifications and with a large increase in a variety of observations has led to improvements in NWP accuracy (Simmons and Hollingsworth[332]).

Hollingsworth et al. [165] show how observing system improvements led to improvements of forecast scores while Bouttier and Kelly[36] show that the improvement of forecast scores for the southern hemisphere are due to satellite data. Also, error statistics for different sources of observation constitutes an active field of research aimed mainly at obtaining better representation of the specific observation operators.

30 The algorithmic developments of note for 4-D VAR

Following an idea of Derber, Courtier et al. [74] proposed and developed the incremental 4-D VAR algorithm, where minimization is carried out at reduced resolution in the inner iteration and on a linear model. The 4-D VAR incremental algorithm minimizes the following cost function (Rabier[304])

$$J(\delta_{w_0}) = \frac{1}{2}\delta_{w_0}^T B^{-1}\delta_{w_0} + \frac{1}{2}\sum_{I=1}^{N}(H_i\delta_{X_i} - d_i)^T R_i^{-1}(H_i\delta_{X_i} - d_i) \tag{240}$$

with $\delta_{w_0} = s(X_0 - X_b)$.

Simplified increment at initial time t_0

$$d_i = y_i^o - H_i(X_i) \tag{241}$$

is the observation increment at time t_i. The solution resulting from minimization of the cost function is added to the background X_b to obtain analysis at t_0 i.e

$$X_0^a = X^b - S^{-I}\delta_{w_0}^a \tag{242}$$

where S^{-I} is the generalized inverse of operator S which projects from high to low resolution (i.e S^{-I} projects from low to high resolution). In an outer loop one updates the high resolution reference trajectory and observation departures. A refinement of the incremental 4-D VAR was proposed as a multi-incremental algorithm by Veerse and Thépaut[370]. Physical parameterizations that have been modified to allow use in the linear models used in the incremental procedure were implemented by Janiskova et al. [169], Lopez and Moreau[207].

31 Impact of observations

In view of high density of some observations horizontal thinning is performed on data sets, and optimal observation density is found by trial and error.

Another approach called "super-obbing",i.e. it averages neighboring observations. A new advance concerns the information content of the data. While usual method of estimating data impact in a forecasting system consists in performing observing system experiments (OSE) which turn out computationally expensive. However, another diagnostic called the "degrees of freedom for signal (DFS)" has been used by Rodgers[311], Fisher[118] and Cardinali et al.[49].

Given an analysis x_a, background x_b and observation y^o we have

$$x_a = x_b + (B^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} (y^o - \mathcal{H}(x_b)) \tag{243}$$

which can be written compactly as

$$x_a = x_b + Kd \tag{244}$$

B-being the background error covariance matrix, R the observation error covariance, H-linearized observation operator of $\mathcal{H}$. K is called the Kalman gain matrix and d innovation vector $d = y^o - \mathcal{H}(x_b)$.

The DFS is defined as

$$DFS = Tr(HK) \tag{245}$$

where the trace of the matrix HK measures the gain in information due to the observations or how an assimilation system extracts information signal from the background.(See Rabier (2005) [304]. One way to calculate DFS is the use of estimation the Hessian of the cost function provided. Fisher[118] and Cardinali et al.[49] used estimation of Hessian of the cost function provided by the minimization algorithm. Chapnik et al.[54] use evaluation of trace of the KH matrix, using a method put forward by Desroziers and Ivanov[105] to evaluate trace of KH.

Computing sensitivity of forecast to the observations can be carried out by considering the adjoint of data assimilation together with the adjoint of the forecast model. This allows use of adaptive observations to be addressed in next section (Berliner et al.(1999), Baker and Daley[27]) Daescu and Navon[78], Langland and Baker [185]

32 Singular Vectors

Lorenz[228] was the first to have introduced the concept of tangent linear model. If one denotes by $c(t)$ the state vector of a generic autonomous system, the evolution equations of which can be written as

$$\frac{\partial c}{\partial t} = A(c) \tag{246}$$

(adopting the notation of Buizza[40])

The integration of (246) from time t_0 to time t generates a trajectory from initial point c_0 to a point $c_(t)$. The evolution of a small perturbation x, around the time dependent trajectory $c(t)$ neglecting quadratic and higher order terms in the perturbation x can be written as

$$\frac{\partial x}{\partial t} = A_L x \tag{247}$$

where A_L is the tangent linear operator computed at the trajectory point $x(t)$ (see Lorenz(1965)[228]). If we denote by $L(t_0, t)$ the resolvent or propagator of the tangent linear model taking an initial perturbation at time t_0 to the final perturbation at time t_1. Defining an inner or Euclidean scalar product and let E denote a weight matrix (Li et al.[206]),

$$||x||^2 = (x, x) =< x, Ex > \tag{248}$$

where

$$< x, y >= \sum_{i=1}^{N} x_i y_i \tag{249}$$

is the canonical Euclidean scalar product.

Let L^T be the adjoint of L with respect to the inner product $< , >_E$

$$< L^T x, y >_E =< x, Ly >_E . \tag{250}$$

The adjoint of L with respect to the inner product defined by E in terms of adjoint L^* defined with respect to canonical Euclidean scalar product is

$$L_E^T = E^{-1} L^T E \tag{251}$$

In order to compute the leading singular vectors (see Golub and Van Loan[144]) that yield the fastest growing perturbations during the optimization period $[t_0, t]$ for which

$$\frac{||x(t_0)||^2}{||x(t)||^2} \tag{252}$$

is maximum.

We consider

$$||x(t)||_E^2 =< x(t_0), L_E^T L(x(t_0)) >_E . \tag{253}$$

We have to search for the eigenvectors $v_i(t_0)$ such that

$$L_E^T L v_i(t_0) = \lambda_i^2 v_i(t_0) \tag{254}$$

with the largest eigenvalues λ_i^2. If we were to use as Molteni et al.[248], Buizza and Palmer[38] the energy norm (see Li et al.[206]), we can relate the energy of the perturbation x at time t to its initial value at time t_0

$$\begin{aligned} \frac{||x_t||^2}{||x_0||^2} &= \frac{< x_t, E_t x_t >}{< x_0, E_0 x_0 >} \\ &= \frac{< Ax_0, E_t A x_0 >}{< x_0, E_0 x_0 >} \\ &= \frac{< x_0, A^T E_t A x_0 >}{< x_0, E_0 x_0 >} = \lambda^2 \end{aligned} \tag{255}$$

and we solve eigenvalue problem

$$A^T E_t A v_i = \lambda_i^2 E_0 v_i. \tag{256}$$

If $E_t = E_0$, then λ is called the energy amplification factor.

Due to the large dimension of the system and since in practice one needs only a small number of singular values (SVs) compared with huge dimension of system one uses the Lanczos methods (ARPACK library, documented in Lehoucq et al. 1998). In 1992 and 1995, Barkmeijer[12] and Buizza and Palmer[38] introduced projection operators to make the SV's more relevant to limited-area models. The local projection operator P, sets the model variables to zero outside a focused area, i.e. for all grid points or spectral components outside the geographic area of interest.

The definition of the amplification factor is generalized as

$$\lambda^2 = \frac{< Px_t, E_t Px_t >}{< x_0, E_0 x_0 >} \tag{257}$$

where adjoint operators A^T and P^T are available the eigenvalue problem can be solved using Lanczos-type algorithms. These algorithms require only evaluation of matrix-vector products and avoid computationally expensive manipulation of large matrices. Usually one uses the popular ARPACK package developed specifically for problems of large dimensions (Lehoucq et al.[194]).

An enormous effort and a large number of research works has centered on the issue of singular vectors led by researchers at ECMWF and the list of contributions is too long to enumerate. Important work on this topic includes that of Errico and Vukicevic[111] showing sensitivity of singular vectors to both choice of norm and length of the time optimization interval, Errico[112], Rabier et al.[301]. We will dwell in this survey only on one aspect of singular vectors and 4-D VAR namely its application for targeted observations.

33 Targeting observation with a 4-D VAR data assimilation system

Adaptive observation are denoting a set of strategies aiming at improving forecast skill of numerical weather prediction by identifying optimal locations where targeted observations must be collected.

Mathematically this is related to sensor-actuator or location problem but here we illuminate briefly only approaches where either the adjoint sensitivity approach or the use of singular vectors is concerned See (Rabier et al.[301], Palmer et al.[284], Berliner et al.[27], Baker and Daley[8], Bergot and Doernbecher[25],Leutbecher(2003)[195], Langland and Baker[185].

Optimal deployment of targeted observations was also addressed by Lorenz and Emanuel[229], Bergot[24], Morss et al.[252].

Daescu and Carmichael[77] and Daescu and Navon[78] account for the dynamical interaction between forecast sensitivity field and the sensitivity field associated to all additional data available to the assimilation procedure.

34 The singular vectors approach to targeting

Use of singular vectors for targeting observations is related to work of Palmer et al.[284], Buizza and Montani[39] and Daescu and Navon[78] to cite but a few.

One searches for directions where errors in the state vector at targeting time will propagate most at verification time on the verification domain .Consider perturbation δx_i of model state at t_i then to first order the induced perturbation at time t_v is

$$\delta x_v = M(x_i + \delta x_i) - M(x_i) \approx L(t_i, t_v)\delta x_i \tag{258}$$

$L(t_i, t_v)$ is the resolvent of TLM in time interval $t_v - t_i$.

On the tangent phase space consider the inner product

$$< \delta x, \delta y >_C = < \delta x, C\delta y > \tag{259}$$

where C is a symmetric positive definite matrix, $< \cdot, \cdot >_C$ is the inner product. The adjoint of L in $< \cdot, \cdot >_C$ defined as

$$< L^{*C}\delta x, \delta y >_C = < \delta x, L\delta y >_C \tag{260}$$

$L^{*C} = C^{-1}L^*C$ where L^* is the adjoint operator of L in $< \cdot, \cdot >$.

From last equation it follows

$$||\delta x(t_v)||_C = < \delta x(t_i), L^{*C}L\delta x(t_i) >_C \tag{261}$$

i.e. the directions characterized by maximum growth $||\delta x(t_v)||_C / ||\delta x(t_i)||_C$ are the singular vector $\nu_j(t_i)$

$$L^{*C}L\nu_j(t_i) = \sigma_j^2 \nu_j(t_i) \tag{262}$$

associated with the largest singular values σ_j. Singular vectors depend on the C-norm selection. Studies of Barkmeijer[13,14] performed in the 3-D VAR data assimilation framework show that if operator C at t_0 is taken to be Hessian of the cost functional then the computed Hessian singular vectors (HSV) are consistent with 3-D VAR estimates of analysis error statistics.

If P is the projection operator on the verification domain D_v the singular value problem

$$(C^{\frac{1}{2}}PLC^{\frac{-1}{2}})^*C^{\frac{1}{2}}PLC^{\frac{-1}{2}}\mu_j = \sigma^2\mu_j \tag{263}$$

where $\nu_j = C^{\frac{1}{2}}\mu_j$ has to be solved in the optimization interval $t_v - t_i$.

34.1 Target area definition using first N leading singular vectors

Consider the approach of Buizza and Montani[39]. If we consider the first N leading singular vectors with unit C-norm , i.e.

$$||\nu_j||_C = 1, \quad j = 1, 2, \cdots, N \tag{264}$$

and let us consider the value of the C norm taken to be the total energy norm at a mesh point (λ, θ) on the sphere (see Daescu and Navon[78]). We define a sensitivity function as

$$F_N^C(\lambda, \theta) = \sum_{j=1}^{N} (\frac{\sigma_j}{\sigma_1}) f_j^C(\lambda, \theta) \tag{265}$$

Taking additional observations at time t_i at locations where sensitivity field (265) is maximal are assumed to enhance forecast improvement.

A target area may be defined as

$$\mathcal{D}_i = \{(\lambda, \theta) | F_N^C(\lambda, \theta) \geq 0.5 F_{MAX}\} \tag{266}$$

where

$$F_{MAX} = \max_{(\lambda,\theta)} |F_N^C(\lambda, \theta)| \tag{267}$$

and adaptive observations at time t_i are selected at the first n_i locations (λ, θ) where sensitivity field $F_N^C(\lambda, \theta)$ attains the largest values.

Important targeted observing field programmes started with the Fronts and Atlantic Storm-Track Experiment in 1997 (FASTEX). Joly et al.[171] using ensemble transform Kalman filter (ETKF), Hessian Singular vectors and total energy singular vectors. These methods are surveyed in Langland[186].The North Pacific Experiment (NORPEX) Langland et al.[187] was based on total energy singular vectors (TESV). Winter Storm Reconnaissance Program

(WSRP) (see Toth et al.[363]) uses ETKF as targeting guidance to identify target areas.

34.2 Open issues in targeting

An important issue us the availability of objective methods to produce targeting guidance that are consistent with the data assimilation procedure and estimate relevant aspects of the analysis error covariance matrix (Langland[186]). Targeting has reduced short-range forecast errors. A promising new approach to objective targeting is observation-space targeting, which uses sensitivity information that can be provided by the adjoint of a data assimilation procedure or other method.

Another relevant issue is that of observation sample size in targeting , because the amount of targeted data collected in field programmes is generally small in comparison to that provided by regular observing systems.

35 Status of 3-D VAR data assimilation efforts

During development of variational data assimilation at operational centers, 3-D VAR which produces an "optimal" estimate of the true atmospheric state at analysis time through iterative solution of a prescribed 3-D (without the time dimension) cost function using the (Ide et al.[167]) notation

$$J(x) = J^b + J^o = \frac{1}{2}(x - x^b)^T B^{-1}(x - x^b) + \frac{1}{2}(y - y^o)^T (E + F)^{-1}(y - y^o) \quad (268)$$

where y^o are the observations.

The fit of the data points is weighted by estimates of their errors where (see description in Barker et al.[10,11]) B is the background error covariance matrix E and F are the observation and representativeness error covariance matrices respectively.

The representativity error is an estimate of inaccuracies introduced in observation operator H transforming grid analysis x to observation space $y = Hx$ for comparison with the observations.

The justification for the development of 3DVAR has been its use as a necessary step prior to attaining the goal of implementation of 4-D VAR or Kalman filter type of assimilation algorithms. This was due to lack of computing resources and its implications for cut-off time restrictions for operational centers

for applying the full 4-D VAR approach. Every operational center invested large team efforts to develop 3-DVAR methodology. Parrish and Derber[288] at NMC/NCEP, Rabier et al.[303] at ECMWF, Lorenc et al.[226] at UMO, Barker et al.[10,11] for the MM5 at UCAR/MMM. Daley and Barker[85] and Cohn et al.[64] used PSAS i.e. observation space 3DVAR.

3DVAR does not require the forecast adjoint model or for incremental 4-DVAR the corresponding linear model used to describe evolution of finite perturbations. It is obvious from the work of Lorenc and Rawlins[227] that even incremental 4-DVAR beats 3-DVAR due to its using time-evolved covariances and Lorenc and Rawlins[227] also think that 4-DVAR is improving the analysis of growing modes (see also Thepaut et al. [358]) which are more important for accurate forecasts. Nevertheless the usefulness of 3-DVAR is there to stay for a certain period despite the introduction of 4-D VAR due to its training and reference capabilities.

36 Conclusions

A condensed review of several aspects of 4-D VAR as it evolved in the last 20 or so years is presented. It aims to present both the history of 4-D VAR as well as its evolution by highlighting several topics of its application.

No attempt was made to cover the ensemble Kalman filter data assimilation due to space and time limitations. In the same vein this review is not exhaustive as it is not covering all the issues dealing with 4-D VAR applications.

It has become amply evident that in the last 15 years major improvements in NWP are due to large extent to development of sources of observations and that 4-D VAR can take advantage of them due to major research efforts at both research and operational centers.

For new opportunities for research see the article by McLaughlin et al. (2005)[241] that illuminates and outlines possibilities for enhanced collaboration within the data assimilation community .

It is certain that data assimilation concepts will become widely applied as more geoscience scientific disciplines gain access to larger amounts of data, from satellite remote sensing and from sensor networks, and as Earth system models increase in both accuracy and sophistication.

It is hoped that this review highlights several aspects of 4-D VAR data assimilation and serves to attract interest of both atmospheric science practitioners as well as variational optimization scientists.

References

[1] J.E. Aarnes, Iterative methods for data assimilation and an application to ocean state modeling. Technical Report, SINTEF Applied Mathematics Oslo, Norway, (2004).

[2] S. Akella, I.M. Navon, On forecast errors in variational data assimilation using high resolution advection schemes of the Lin-Rood finite volume shallow water model. Submitted for publication to Q.J.R.M.S., (2005).

[3] L. Amodei, Solution approachee pour un probleme d'assimilation de donnees meteorologiques avec prise en compte de l'erreur de modele. Comptes Rendus de l'Academie des Sciences 321, serie II a (1995) 1087-1094.

[4] E. Andersson, A. Hollingsworth, G. Kelly, P. L.Lonnberg, J. Pailleux, and Z. Zhang, Global observing system experiments on operational statistical retrievals of satellite sounding data. Mon. Wea. Rev. 199 (1991) 1851-1864.

[5] E. Andersson, J. Pailleux, J. N. Thepaut, J. R. Eyre, A. P. McNally, G. A. Kelly and P. Courtier, Use of radiance in 3D/4D variational data assimilation. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK (1992), 123-156.

[6] R.A. Anthes, Data assimilation and initialization of hurricane prediction models. J. Atmos. Sci. 31 (1974) 702-719.

[7] R. Aster, B. Borchers and C. Thurber, Parameter Estimation and Inverse Problems. Elsevier, (2005), pp 320.

[8] N.L. Baker, R. Daley, Observation and background adjoint sensitivity in the adaptive observation-targeting problem. Q.J.R. Meteorol. Soc., 126, (2000) 1431-1454.

[9] H.T. Banks, K. Kunisch, Estimation techniques for distributed parameter systems. Birkhauser, Boston (Systems and Control: Formulations and Applications) 11 (1989) pp 315.

[10] D. M. Barker, W. Huang, Y.-R. Guo, and A. Bourgeois, A three-dimensional variational (3DVAR) data assimilation system for use with MM5. NCAR Tech. Note.(2003) NCAR/TN-453 1 STR, 68 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO 80307.]

[11] D. M. Barker, W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N. Xiao,A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results. Monthly Weather Review, 132,(2004) 897-914.

[12] J. Barkmeijer, Local error growth in a barotropic model. Tellus 44A (1992) , 314-323.

[13] J. Barkmeijer, M. Van Gijzen, F. Bouttier , Singular vectors and estimates of the analysis-error covariance metric. Quart. J. Roy. Meteor. Soc. Part A (1998) 1695-1713 .

[14] J. Barkmeijer, R. Buizza, T.N. Palmer, 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. Quart. J. Roy. Meteor. Soc. Part B (1999) 2333-2351.

[15] J.W. Bao, T.T. Warner, Treatment of On/Off switches in the adjoint method: FDDA experiments with a simple model. Tellus, 45A (5) (1993) 525-538 .

[16] S.L. Barnes, A Technique for Maximizing Details in Numerical Weather Map Analysis. Journal Appl. Meteor. 3 (1964) 395-409.

[17] S.L. Barnes, Oklahoma Thunderstorms on 29C30 April 1970. Part I: Morphology of a Tornadic Storm. Mon Wea Rev. 106 (5) (1978) 673-684.

[18] R.E. Bellman, Dynamic Programming. Princeton University Press, Princeton, 1957.

[19] L. Bengtsson, M. Ghil and E. Källen (Eds.), Dynamic Meteorology: Data Assimilation Methods, Springer-Verlag (1981), pp 330.

[20] A. Bennett, B. Chua, L. Leslie, Generalized inversion of a global numerical weather prediction model. Meteorol. Atmos. Phys. 60 (1996) 165-178.

[21] A. Bennett, B. Chua, L. Leslie, Generalized inversion of a global numerical weather prediction model (II). Meteorol. Atmos. Phys. 62 (1997) 129-140.

[22] A. Bennett, L. Leslie, C. Hagelberg, P. Powers, Tropical cyclone prediction using a barotropic model initialized by a generalized inverse method. Mon. Weather Rev. 121 (1993) 1714-1729.

[23] A.F. Bennett ,Inverse Modeling of the Ocean and Atmosphere. Cambridge University Press. (2002), pp 256.

[24] T. Bergot, Influence of the assimilation scheme on the efficiency of adaptive observations. Q.J.R. Meteorol. Soc., 127, (2001) 635-660.

[25] T. Bergot, A. Doerenbecher, A study on the optimization of the deployment of targeted observations using adjoint-based methods. Q.J.R. Meteorol. Soc., 128, (2002) 1689-1712.

[26] P. Bergthorsson, B. Döös , Numerical Weather Map Analysis, Tellus 7(3) (1955) 329-340.

[27] L. M. Berliner, Lu, Z.Q., and C. Snyder, Statistical design for adaptive weather observations. J. Atmos. Sci., 56, (1999) 2536-2552.

[28] D.P. Bertsekas, Constrained optimization and Lagrange multiplier methods. Addison-Wesley, (1982), pp 491.

[29] D.P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods (Optimization and Neural Computation Series). Athena Scientific; 1st edition (January 1996), pp 410.

[30] M. Berz, C.H. Bischof, G. Corliss, and A. Griewank (eds.), Computational differentiation - techniques, applications, and tools, SIAM, Philadelphia, Penn., 1996.

[31] C.H. Bischof, G.F. Corliss, L.Green, A. Griewank, K. Haigler, Automatic differentiation of advanced CFD codes for multidisciplinary design, Journal on Computing Systems in Engineering 3 (1992), 625-638.

[32] C.H. Bischof, A. Carle, P. M. Khademi, and A. Mauer, The ADIFOR 2.0 System for the Automatic Differentiation of FORTRAN 77 Programs, IEEE Computational Science & Engineering, 3(3)(1996), 18-32.

[33] S.C. Bloom, S. Shubert, The influence of Monte-Carlo estimates of model error growth on the GLA OI assimilation system. Pp. 467-470 in International symposium on assimilation of observations in meteorology and oceanography, WMO, Geneva, Switzerland, 1990.

[34] G.J. Boer, A spectral analysis of predictability and error in an operational forecast system. Mon. Weather Rev. 112 (1984) 1183-1197

[35] F. Bouttier, P. Courtier, Assimilation concepts and methods, Meteorological Training Lecture Notes, ECMWF. Shinfield Park, Reading, (1999).

[36] F. Bouttier, Kelly, G. Observing-system experiments in the ECMWF 4D-Var data assimilation system. Q. J. R. Meteorol. Soc., 127, (2001) 1469-1488.

[37] A. Bratseth, Statistical interpolation by means of successive corrections , Tellus 38A (1986) pp 439-447.

[38] R. Buizza, Palmer, T. N., The singular-vector structure of the atmospheric general circulation. J. Atmos. Sci., 52, 9,(1995) 1434-1456.

[39] R. Buizza, Montani A. Targeted observations using singular vectors. J Atmos Sci 56, (1999) 2965-2985.

[40] R. Buizza, Chaos and weather prediction, Il Nuovo Cimento C, (2001) 24, 273-302.

[41] D.G. Cacuci, C.F. Weber, E.M. Oblow, J.H. Marable, Sensitivity Theory for General Systems of Nonlinear Equations, Nucl. Sci. Eng. 75(88) (1980).

[42] D.G. Cacuci, Sensitivity theory for nonlinear systems. I: Nonlinear functional analysis approach. J. Math. Phy. 12 (1981) 2794-2803.

[43] D.G. Cacuci, Sensitivity theory for nonlinear systems. II. Extensions to additional classes of responses. J. Math. Phy. 12 (1981) 2803-2812.

[44] D. G.,Cacuci, The Forward and the Adjoint Methods of Sensitivity Analysis, in Uncertainty Analysis, Y. Ronen, Ed., CRC Press, Inc., Boca Raton, Florida, Chap. 3, (1988).

[45] D. G.,Cacuci, Global Optimization and Sensitivity Analysis, Nucl. Sci. Eng., 104,(1990) 78-88.

[46] D.G. Cacuci, Sensitivity & Uncertainty Analysis: A Primer, Vol. 1.(2003) CRC Press,304pp

[47] D.G. Cacuci, On the Historical Evolution of the Use of Adjoint Operators in Sensitivity and Uncertainty Analysis. Sixth Workshop on Adjoint Applications in Dynamic Meteorology, 23 - 28 May (2004).Acquafredda di Maratea, Basilicata, Italy.

[48] D. G. Cacuci, M. Ionescu-Bujor, I. M. Navon, Sensitivity and Uncertainty Analysis, Volume II: Applications to Large-Scale Systems. (2005) CRC Press,368pp

[49] C. Cardinali, Pezzuli, S. and Andersson, E. Influence matrix diagnostic of a data assimilation system. Q. J. R. Meteorol. Soc., 130, (2004) 2767-2786.

[50] J. Carrera, S.P. Neuman, Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, Water Resources Research 2 (1986), 199-210.

[51] J. Carrera, S.P. Neuman, Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability and solution algorithms, Water Resources Research 2 (1986) 211-227.

[52] J. Carrera, S.P. Neuman, Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data, Water Resources Research 2 (1986) 228-242.

[53] W.C. Chao, L. Chang, Development of a 4-dimensional analysis system using the adjoint method at GLA. part 1: dynamics. Mon. Wea. Rev. 120 (1991) , 1661-1673.

[54] B. Chapnik, Desroziers, G., Rabier, F. and Talagrand, O. Diagnosis and tuning of observational error statistics in a quasi-operational data assimilation setting. Q. J. R. Meteorol. Soc., 132, (2006), 543-565.

[55] J.G. Charney, R. Fjörtoft, John von Neumann, Numerical Integration of the Barotropic Vorticity Equation, Tellus 2 (1950) 237-254.

[56] J.G. Charney, M. Halem, R. Jastrow, Use of incomplete historical data to infer the present state of the atmosphere. J. Atmos. Sci. 2 (1969) 1160-1163.

[57] G. Chavent, Identification of functional parameters in partial differential equations. In Identification of parameters in Distributed systems, ASME Conference on Automatic Control, Austin, Texas June17-21, (1974), Goodson R.E. and Polis M. Eds. 1974, 31-48.

[58] G. Chavent, P. Lemonnier, Identification de la non-linearite d'une equation parabolique quasi-lineaire. Applied Math. and Opt. 1 (1974) 121-162.

[59] G.A. Chepurin, J. Carton, D.P. Dee, Forecast model bias correction in ocean data assimilation. Mon. Wea. Rev. 133 (2005) 1328-1342.

[60] S.E. Cohn, Methods of sequential estimation for determining initial data in numerical weather prediction. Ph.D Thesis, New York University, 1982, pp 183.

[61] S.E. Cohn, D.F. Parrish, The behavior of forecast error covariances for a Kalman filter in two dimensions. Mon. Wea. Rev. 119(1991) 1757-1785.

[62] S.E. Cohn, Dynamics of short-term univariate forecast error covariances. Mon. Wea. Rev. 11 (1993) 3123-3149.

[63] S. E. Cohn, An introduction to estimation theory. J. Meteorol. Soc. Jpn. 75 (1997) 257-288

[64] S. E. Cohn, A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, Assessing the effects of data selection with the DAO Physical- Space Statistical Analysis System. Mon. Wea. Rev., 126,(1998) 2913- 2926.

[65] A.R. Conn, N. I. M. Could, and Ph. L. Toint, LANCELOT: a FORTRAN Package for Large-Scale Nonlinear Optimization (Release A), Springer Series in Computational Mathematics, Volume 17, Springer-Verlag, New York, New York, 1992.

[66] P. Courtier, Experiments in data assimilation using the adjoint model technique. Workshop on High-Resolution Analysis ECMWF (UK) June 1985.

[67] P. Courtier, Le modele adjoint, outil pour des experiences de sensibility. Note de travail no 166. EERM, Paris, France, (1986), 36pp. (available from Météo-France)

[68] P. Courtier, O.Talagrand, Variational assimilation of meteorological observations with the adjoint equations Part 2. Numerical results. Q. J. R. Meteorol. Soc. 113(1987) 1329-1347.

[69] P. Courtier, Application du controle optimal a la prevision numerique en meteorologie. These de Doctorat de Universite-Pierre-et-Marie, Paris, 1987.

[70] P. Courtier,O. Talagrand Variational assimilation of meteorological observations with direct and adjoint shallow-water equations. Tellus. 42A (1990) 531-549.

[71] P. Courtier, J. N. Thepaut, A strategy for operational implementation of 4D-VAR. In Proc. of ECMWF Workshop on variational assimilation, with special emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, (1992), 437-463.

[72] P. Courtier, P. Unden, W. Heckley, D. Vasijevic, E. Andersson, J.N. Thepaut, F. Rabier, R. Hoffman, Progress in variational data assimilation at ECMWF. Research activities in atmospheric and oceanic modeling, G. J. Boer (ed.), Report No. 18 1.42 (1993).

[73] P. Courtier, Introduction to Numerical Weather prediction Data Assimilation Methods. Proc. ECMWF Seminar on Developments in the use of Satellite Data In Numerical Weather Prediction. 6-10 September 1993.

[74] P. Courtier, Thepaut JN, A. Hollingsworth, A Strategy for Operational implementation of 4D-VAR, using an incremental approach. Quarterly Journal of the Royal Meteorological Society 120 (519, (1994) 1367-1387 Part B.

[75] P. Courtier, Dual formulation of four-dimensional variational assimilation . Quarterly Journal of the Royal meteorological Society 123 (544)(1997), 2449-2461 Part B

[76] G.P. Cressman, An operational objective analysis system. Mon. Wea Rev. 87 (1959) 367-374.

[77] D. N. Daescu, I.M. Navon, An Analysis of a Hybrid Optimization method for variational data assimilation, International Journal of Computational Fluid Dynamics, 17(4) (2003) 299-306.

[78] D.N. Daescu, I.M. Navon, Adaptive observations in the context of 4D-Var data assimilation. Meteorol. Atmos. Phys., 85, (2004) 205-226.

[79] A. Dalcher, E.Kalnay, Error growth and predictability in operational ECMWF forecasts. Tellus 39 A (1987) 474-491

[80] R. Daley, K. Puri, Four-Dimensional, Data Assimilation and the Slow Manifold. Mon. Wea. Rev. 108 (1980) 85-99.

[81] R. Daley, Atmospheric Data Analysis, Cambridge University Press, 1991.

[82] R. Daley, The effect of serially correlated observation and model error on atmospheric data assimilation. Mon. Weather Rev. 120 (1992) 164-177.

[83] R. Daley, Estimating model-error covariances for application to atmospheric data assimilation. Mon. Weather Rev. 120 (1992) 1735-1746

[84] R. Daley, R. Menard, Evaluation of assimilation systems. Notes for three lectures at the international summer school on assimilation of meteorological and oceanographic observations. La Garde, France,(1993) August 2-27.

[85] Daley, R., E. Barker, NAVDAS: Formulation and diagnostics. Mon. Wea. Rev., 129,(2001) 869-883.

[86] A. Da Silva, J. Pfaendtner, J. Guo, M. Sienkiewicz, S. Cohn, Assessing the Effects of Data Selection with DAO's Physical-space Statistical Analysis System. Proceedings of the second international Symposium on the assimilation of observations in meteorology and Oceanography. Tokyo, Japan, 1995.

[87] S.K. Das, R.W. Lardner, Variational parameter estimation for a two-dimensional numerical tidal model. International Journal for Numerical Methods in Fluids 15 (1992) 313-327.

[88] B. Das, H. Meirovitch, and I. M. Navon ,Performance of Enriched Methods for Large Scale Unconstrained Optimization as applied to Models of Proteins , Journal of Computational Chemistry, 24 (10) (2003) 1222-1231.

[89] E.R. Davidson, The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of a large real symmetric matrices. J. Comput. Phys. 17 (1975) 87-94.

[90] H.E. Davis, R.E. Turner, Updating prediction models by dynamic relaxation. An examination of the technique. Q. J. R. Meteorol. Soc. 103 (1977) 225-245

[91] H.C. Davis, R.E. Turner, Updating prediction models by dynamical relaxation: An examination of the technique. J. Appl. Meteor. 17 (1977) 998-1013.

[92] D. Dee, A. Da Silva, Data assimilation in the presence of forecast bias. Q. J. R. Meteorol. Soc. 124 (1998) 269-295

[93] D.P. Dee, R. Todling, Data assimilation in the presence of forecast bias: The GEOS moisture analysis. Mon. Wea. Rev. 128 (2000) 3268-3282.

[94] D.P. Dee, Variational bias correction of radiance data in the ECMWF system. Pp. 97-112 in Proceedings of the ECMWF workshop on assimilation of high spectral resolution sounders in NWP, 28 June-1 July 2004, Reading, UK.

[95] D.P. Dee, Bias and data assimilation - ECMWF/EUMETSAT NWP-SAF Workshop on Bias estimation in data assimilation, 8-11 November (2005) , ECMWF,Reading.pp 1-20

[96] T. DelSole, A.Y. Hou, Empirical correction of a dynamical model. Part I: fundamental issues. Mon. Weather Rev. 127 (1999) 2533-2545

[97] J.E. Dennis Jr., Robert B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations , Prentice Hall , Engelwood Cliffs, New Jersey 07632, 1983.

[98] J.C. Derber, The variational four dimensional assimilation of analysis using filtered models as constraints. Ph.D. Thesis, Univ. of Wisconsin-Madison, (1985), pp 141.

[99] J.C. Derber, Variational four dimensional analysis using the quasigeostrophic constraint. Mon. Wea. Rev. 115 (1987) 998-1008.

[100] J.C. Derber, A variational continuous assimilation technique. Mon. Wea. Rev. 117(1989) 2437-2446.

[101] J.C. Derber, D.F. Parrish, S.J. Lord, The new global operational analysis system at the National Meteorological Center. Weather and Forecast 6 (1991) 538-547.

[102] J.C. Derber, D.F. Parrish, J.G. Sela, The SSI analysis system and extensions to 4-D. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, (1992), 15-29.

[103] J.C. Derber, W.-S. Wu, The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. Mon. Wea. Rev. 126 (1998) 2287-2299.

[104] J. C. Derber, F. Bouttier, A reformulation of the background error covariance in the ECMWF global data assimilation system. Tellus. 51A (1999) 195-221

[105] G. Desroziers, Ivanov, S. Diagnosis and adaptive tuning of information error parameters in a variational assimilation. Q. J. R. Meteorol. Soc., 127, (2001) 1433-1452.

[106] G. Desroziers, Hello, G. and Thepaut, J.-N. A 4D-Var re-analysis of the FASTEX. Q. J. R. Meteorol. Soc., 129,(2003) 1301-1315.

[107] M. Ehrendorfer, Four-dimensional data assimilation: comparison of variational and sequential algorithms. Q. J. R. Meteorol. Soc. 118 (1992) 673-713.

[108] M. Ehrendorfer, F. Bouttier, An explicit low-resolution extended Kalman filter: Implementation and preliminary experimentation. Sept 1998. ECMWF Technical Memorandum No 259.

[109] A. Eliassen, Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. Report # 5, Institute of Weather and Climate Res., Academy of Science, Oslo, (1954), pp 11.

[110] Y.M. Ermolev, Methods for solving nonlinear extremal problems. Cybernetics 2(4) (1996) 1-14.

[111] R. M. Errico, and T. Vukicevic, Sensitivity analysis using an adjoint of the PSU-NCAR mesoscale model. Mon. Wea. Rev., 120, (1992) 1644-1660.

[112] R. M. Errico, What Is an Adjoint Model?, Bulletin of the American Meteorological Society, Vol 78, 11, (1997) 2577-2592.

[113] L. Euler, Methodus inveniendi lineas curvas maximi minimive propietate gandentes sive solutio problematis isoperimetrici latissimo sensu accepti, Lausanne, Opera S.I. 24 (1744) 174-189, ed. Constantin Caratheodory.

[114] L. Euler, Elementa calculi variationum novi comentarii academiae scietiarum petropolitane, 10 (1764) 51-93 Opera S.I., 25, Caratheodory, (1952), Basel, 141-176.

[115] M. Fisher, P. Courtier, Estimating the covariance matrices of analysis and forecast errors in variational data assimilation. ECMWF Tech Memorandum, 220 (1995) pp28.

[116] M. Fisher, Development of a simplified Kalman filter. ECMWF Tech Memorandum 260 (1998) pp 16.

[117] M. Fisher, Background Error Covariance Modelling M Fisher seminar ,Recent developments in data assimilation for atmosphere and ocean, 8-12 September 2003 European Centre for Medium-Range Weather Forecasts

[118] M. Fisher, Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. Technical Memo. 397, (2003) ECMWF, Reading, UK.

[119] M. Fortin, R. Glowinski, Methodes de Lagrangien augmente Dunod, Paris, 1982.

[120] R. Franke, W.J. Gordon, The structure of optimal interpolation functions. Technical Report, NPS-53-83-0004 , (1983) Naval Postgraduate School, Monterey, Ca. 1983.

[121] R. Franke, Statistical interpolation by iteration. Mon. Wea. Rev. 116 (1988) 961-963.

[122] L.S. Gandin, Objective analysis of meteorological fields. Leningrad: Gridromet(in Russian) (1963), English translation by the Israel program for scientific translations, 1965, pp 242.

[123] G. Gaspari, S.E. Cohn, Construction of correlation functions in two and three dimensions Quarterly Journal of the Royal Meteorological Society 125 (554) (1999), 723-757.

[124] G. Gaspari ,S.E. Cohn ,J. Guo , et al. Construction and application of covariance functions with variable length-fields Quarterly Journal of the Royal Meteorological Society 132 (619) (2006) , 1815-1838 .

[125] C.F. Gauss, Theoria motus corporum coelestium in sectionibus conicis solem ambientum Book 2, Section 2, (1809), pp. 205-224 English translation by Ch. Davis, reprinted 1963 by Dover, New York.142pp

[126] C.F Gauss, Theoria Combinationis Observationum Erronbus Minimus Obnoxiae, in two parts and a supplement. Commentatines Societatis Regiae Scietiarum Gottingemis Recentiores, (1823), 5, 33-62 and 63-90, 6, 57-93. Reprinted with facing English translation and notes by G.W. Stewart, (1995). Philadelphia: SIAM.

[127] P. Gauthier, P. Courtier, Data assimilation with an extended Kalman filter. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, (1992), 171-190.

[128] P. Gauthier, Chaos and quadri-dimensional data assimilation: A study based on the Lorentz model. Tellus 44A (1992) 2-17.

[129] P. Gauthier, Thepaut, J.-N. Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Meteo-France. Mon. Weather Rev., 129, (2001) 2089-2102.

[130] D.M. Gay, Subroutines for Unconstrained Minimization Using a Model Trust Region Approach, ACM Trans. Math. Soft. 9 (1983) 503-524.

[131] A. Gelb (Ed.), Applied Optimal Estimation, M.I.T. Press, Cambridge, MA, (1974), pp374.

[132] M. Ghil, B. Shkoller, V. Yangarber, A balanced diagnostic system compatible with a barotropic prognostic model. Mon. Wea. Rev. 105 (1977) 1223-1238.

[133] M. Ghil, M. Halem, R. Atlas, Time-continuous assimilation of remote-sounding data and its effect on weather forecasting, Mon. Wea. Rev. 107 (l979) l40-l7l.

[134] M. Ghil, S. Cohn, J. Tavantzis, K. Bube, and E. Issacson, Application of estimation theory to numerical weather prediction. L. Bengtsson, M. Ghil and E. Kallen (eds.), Dynamic Meteorology: Data assimilation methods, Springer-Verlag, New York, (1981), 139-244.

[135] M. Ghil, S. E. Cohn, A. Dalcher, Sequential estimation, data assimilation, and initialization. D. Williamson (Ed.), The interaction between objective analysis and initialization, Publ. Meteorol., 127 (Proc. 14th standard seminar), McGill University, Montreal, (1982), 83-97.

[136] M. Ghil, Sequential estimation and satellite data assimilation in meteorology and oceanography. Y. K. Sasaki, T. Gat-Chen, L. White, M. M. Zaman, C. Zieler, L. P. Chang and D. J. Rusk (ed.), Variational Methods in Geosciences, Elsevier, Amsterdam, (1986), 91-100.

[137] M. Ghil, P. Malonotte-Rizzoli, Data assimilation in meteorology and oceanography. Adv. Geophys. 33 (1991) 141-266.

[138] M. Ghil, K. Ide, A. F. Bennett, P. Courtier, M. Kimoto, N. Sato (Eds.), Data Assimilation in Meteorology and Oceanography: Theory and Practice, Meteorological Society of Japan and Universal Academy Press, Tokyo, (1997), 496pp.

[139] R. Giering and T. Kaminski, Recipes for adjoint code construction, Technical Report 212, Max-Planck-Institut fuer Meteorologie, 1997.

[140] J.C. Gilbert, Automatic differentiation and iterative processes. Optimization Methods and Software, (1992), 13-21.

[141] J.C. Gilbert , C. Lemarechal, Some Numerical Experiments with Variable Storage Quasi-Newton Algorithms, Mathematical Programming, 45,(1989) 407-436.

[142] B. Gilchrist, G.P. Cressman, An experiment in objective analysis. Tellus 6 (1954) 309-318.

[143] P.E. Gill, W. Murray, Practical optimization. Academic Press, (1981), pp 401.

[144] G. Golub , C. van Loan , Matrix computations, third edition, The Johns Hopkins University Press,(1996) London,728pp.

[145] A. Grammeltvedt, A survey of finite-difference schemes for the primitive equations for a barotropic fluid. Mon. Wea. Rev. 97 (1969) 387-404.

[146] A. Griewank, Ph. L. Toint, Partitioned variable metric updates for large structured optimization problems. Numer. Math. 39(1982) 119-137.

[147] A. Griewank, F.Corliss. George, Automatic differentiation of algorithms: theory, implementation, and application, SIAM, Philadelphia, (1991) pp 353.

[148] A.K. Griffith, N.K. Nichols, Accounting for model error in data assimilation using adjoint methods. Pp 195-205 in Computational Differentiation: Techniques, Applications, and Tools. Ed. M. Berz, C. Bischof, G. Corliss, and A. Griewank, SIAM, Philadelphia, 1996.

[149] A. Griewank, Evaluating Derivatives Principles and Techniques of Algorithmic Differentiation. Frontiers in Applied Mathematics 19 (2000) 369 pp

[150] A.K. Griffith, N.K. Nichols, Adjoint methods in data assimilation for estimating model error. Flow, Turbulence and Combustion 65 (2000) 469-488

[151] A.K. Griffith, Data Assimilation for Numerical Weather Prediction Using Control Theory, The University of Reading, Department of Mathematics, PhD Thesis, 1997.

[152] A.K. Griffith, M.J. Martin, N.K. Nichols, Techniques for treating systematic model error in 3D and 4D data assimilation, in Proceedings of the Third WMO Int. Symposium on Assimilation of Observations in Meteorology and Oceanography, World Meteorological Organization, WWRP Report Series No. 2, WMO/TD - 986 (2000) pp. 9-12.

[153] M.D. Gunzburger, Perspectives in Flow Control and Optimization. Advances in design and control. SIAM, Philadelphia, USA, 2003.

[154] M.C.G. Hall, D. G. Cacuci, M. E. Schlesinger, Sensitivity analysis of a Radiative-Convective Model by the Adjoint Method. J. of the Atmos. Sci. 39 (1982) 2038-2050.

[155] M.C.G. Hall, D. G. Cacuci, Physical interpretation of the adjoint functions for sensitivity analysis of atmospheric models. J. of Atmos. Sci. 40 (1983) 2537-2546.

[156] M.C.G. Hall, Application of adjoint sensitivity theory to an atmospheric general circulation model. J. Atmos. Sci. 43 (1986) 2644-2657.

[157] G.J. Haltiner, R. T. Williams, Numerical prediction and dynamic meteorology. John Wiley and Sons, New York, (1980), pp 477.

[158] P. C. Hansen. Rank-Deficient and Discrete Ill-Posed Problems. SIAM, Philadelphia, PA, (1998).

[159] L. Hascoet, V. Pascual. TAPENADE 2.1 user's guide, Technical report, INRIA,(2004), no 300.81pp.

[160] W.A. Heckley, P. Courtier, E. Andersson, The ECMWF variational analysis: general formulation and use of background information. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, (1992), 49-94.

[161] M.R. Hestenes, Multiplier and gradient methods. J. Opt. Theory Appl. 4 (1969) 303-320.

[162] R.N. Hoffmann, A four dimensional analysis exactly satisfying equations of motion. Mon. Wea. Rev. 114 (1986) 388-397.

[163] J.E. Hoke, R.A. Anthes, The initialization of numerical models by a dynamical initialization technique. Mon. Wea. Rev. 194 (1976) 1551-1556.

[164] A. Hollingsworth, P. Lonnberg, The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: the wind field. Tellus 38A (1986) 111-136.

[165] A. Hollingsworth, Uppala, S., Klinker, E., Onvlee, J., de Vries, J. W., Burridge, D., Vitart, F., de Roo, A. and Pfrang, C., The transformation of earth-system observations into information of socio-economic value in GEOSS. Q. J. R. Meteorol. Soc., 131,(2005), 3493-3512.

[166] C. Homescu and I. M. Navon, Optimal control of flow with discontinuities, Journal of Computational Physics 187 (2003) 660-682.

[167] K. Ide, P. Courtier, M. Ghil, A. Lorenc, Unified notation for data assimilation: Operational, sequential and variational. J. Meteor. Soc. Japan 75 (1997) 181-189.

[168] M. Janiskova, Thepaut, J.-N. and Geleyn, J.-F. , Simplified and regular physical parameterizations for incremental four-dimensional variational assimilation. Mon. Weather Rev., 127,(1999) 26-45.

[169] M. Janiskova, Mahfouf, J.-F., Morcrette, J.-J. and Chevallier, F. Linearized radiation and cloud schemes in the ECMWF model: Development and evaluation. Q. J. R. Meteorol. Soc., 128, (2002) 1505-1527.

[170] T. Janjic ,S.E. Cohn, Treatment of observation error due to unresolved scales in atmospheric data assimilation Monthly Weather Review 134 (10) (2006) , 2900-2915 .

[171] A. Joly, Browning, K. A., Bessemoulin, P., Cammas, J.-P., Caniaux, G., Chalon, J.-P., Clough, S. A., Dirks, R., Emanuel, K. A., Eymard, L., Gall, R., Hewson, T. D., Hildebrand, P. H., Jorgensen, D., Lalaurette, F., Langland, R. H., Lemaitre, Y., Mascart, P., Moore, J. A., Persson, P. O. G., Roux, F., Shapiro, M. A., Snyder, C., Toth, Z. and Wakimoto, R. M., Overview of the field phase of the Fronts and Atlantic storm-track Experiment (FASTEX). Q. J. R. Meteorol. Soc.,(1999) 125, 3131-3163.

[172] H. Kagiwada, R. Kalaba, N. Rasakhoo, and K. Spingarn, Numerical derivatives and nonlinear analysis, Mathematical Concepts and Methods in Science and Engineering, Plenum Press, New York - London, (1986).

[173] R. E. Kalman, A new approach to linear filtering and prediction problems. trans. ASME, J. Basic Eng. 82D (1960) 35-45.

[174] R. E. Kalman,R. S. Bucy, New results in linear filtering and prediction theory. Trans. ASME, J. Basic Eng. 82 (1961) 35-45.

[175] E. Kalnay, S.K. Park, Z-X. Pu and J Gao, Application of the quasi-inverse method to data assimilation. Mon. Wea. Rev. 128 (2000) 864-875.

[176] E. Kalnay, Atmospheric modeling, data assimilation and predictability. Cambridge Univ. Press,(2003), pp 341.

[177] R.E. Kistler, A study of data assimilation techniques in an autobarotropic primitive equation channel model. M.S. thesis. The Pennsylvania State University (1974) 84pp. .Available from the Dept. of Meteorology, The Pennsylvania State University, University Park, PA 16802.

[178] E. Klinker, Rabier, F., Kelly, G. and Mahfouf, J.-F., The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration. Q. J. R. Meteorol. Soc.,(2000), 126, 1191-1215.

[179] G. Kontarev, The adjoint equation technique applied to meteorological problem. ECMWF Tech. Rep., Reading, U. K. 21 (1980), pp 1-21.

[180] V.M. Krasnopolsky, Helmut Schillerb, Some neural network applications in environmental sciences. Part I: forward and inverse problems in geophysical remote measurements. Neural Networks 16 (2003) 321-334.

[181] V.M. Krasnopolsky , Michael S. Fox-Rabinovitz, Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction. Neural Networks 19 (2006) 122-134.

[182] D.G. Krige, A statistical approach to some mine valuation problems on the Witwatersrand. J. Chem. Metal. Min. Soc. South African 52 (1951) 119-139.

[183] T. N. Krishnamurti, H. Pan, C. B. Chang C. B., J. Ploshay, W. Oodally, Numerical weather prediction for GATE. Quart. J. Roy. Met. Soc. 108 (1979) 979-1010.

[184] J.F. Lacarra, O. Talagrand, Short-range evolution of small perturbations in a barotropic model. Tellus 40A (1988) 81-95.

[185] R.H. Langland, N.L. Baker, Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. Tellus, 56A, (2004) 189-203.

[186] R.H. Langland Issues in targeted observing. Q. J. R. Meteorol. Soc. (2005), 131, pp. 3409-3425.

[187] R. H. Langland, Toth, Z., Gelaro, R., Szunyogh, I., Shapiro, M. A., Majumdar, S. J.,Morss, R. E., Rohaly, G. D., Velden, C., Bond, N. and Bishop, C. H. The North Pacific Experiment (NORPEX-98): Targeted observations for improved North American weather forecasts. Bull. Am. Meteorol. Soc., 80,(1999) 1363-1384.

[188] J.L. Lagrange, 1760a, Essai d'une nouvelle methode pour determiner les maxima et les minima des formules indefinies, Miscellanea Taurinensia, 2 (1762) Oeuvres, 1,365-468.

[189] J.L. Lagrange, 1760b, Application de la methode expose dans le memoire precedent a la solution de differentes problemes de dynamique, Miscellanea Taurinensia, 2 (1762), Oeuvres, 1, 365-468.

[190] C. Lanczos, The Variational Principles of Mechanics. Toronto: University of Toronto Press,1970.

[191] R. W. Lardner, Optimal control of open boundary conditions for a numerical tidal model. Computer Methods in Applied Mechanics and Engineering 102 (3),(1993), 367-387.

[192] D.J. Lea, M.R. Allen, T.W.N. Haine, Sensitivity analysis of the climate of a chaotic system. Tellus 52A (2000), 523-532.

[193] D.J. Lea, T.W.N. Haine, Allen, M.R., Hansen, J.A., Sensitivity analysis of the climate of a chaotic ocean circulation model. Q.J.R. Meteorological Soc. 128 (586) (2002), 2587-2605.

[194] Richard B. Lehoucq, Danny C. Sorensen, and C. Yang, ARPACK User's Guide: Solution of Large-Scale Eigenvalue Problems With Implicitly Restarted Arnoldi Methods (Software, Environments, Tools) ,SIAM, Philadelphia, (1998), 142pp.

[195] M. Leutbecher, A reduced rank estimate of forecast error variance changes due to intermittent modifications of the observing network. J. Atmos. Sci., 60,(2003) 729-742.

[196] F.X. Le-Dimet, Une application des methodes de cˆontrole, optimal 'a l'analyse variationnelle," Rapport Scientifique LAMP, Universite Blaise-Pascal, 63170, Aubiere Cedex, (1981), 42 pages.

[197] F.X. Le-Dimet, A general formalism of variational analysis, CIMMS Report Le-Dimet, F. X., Y. K. Sasaki, and L. White, : Dynamic initialization with filtering of gravity waves, CIMMS Report No. 40 (1982), Norman, OK 73019, 12 pages.

[198] F.X. Le-Dimet, J. Segot, Variational analysis of wind field and geopotential at 500 mb, in Proc. Int. Symp. on Variation Meth. in Geosciences, Norman, OK, 73019, Oct. 15-17, (1985), 13-15.

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

[200] F.X. Le-Dimet, A. Nouallier, Assimilation of dynamic data in a limited-area model in variational methods in geosciences, (Y.K. Sasaki, editor) Elsevier, Amsterdam 181-198.(1986) No. 22, 34 pages, Norman, OK 73019.

[201] F.X. Le Dimet, H.E. Ngodock , B. Luong, J. Verron, Sensitivity analysis in variational data assimilation. J Meteorol Soc Japan, (1B) (1997), 245-255.

[202] F. X. Le-Dimet, I. M. Navon and Dacian N. Daescu , Second Order Information in Data Assimilation Monthly Weather Review 130 (3) (2002) 629-648.

[203] F. X. Le Dimet, I.M. Navon, Variational and optimization methods in meteorology: a review. Tech. Rep. FSU-SCRI-88-144, Florida State University, Tallahassee, Florida, (1988), pp 83.

[204] F. X. Le Dimet, I. M. Navon and D. N. Daescu , Second Order Information in Data Assimilation, Monthly Weather Review, 130,(2002) No. 3 , 629-648.

[205] C. Lemarechal, Nondifferentiable optimization, Subgradient and subgradient methods, in Optimization and Operations Research, Lecture Notes in Economics and Mathematical Systems 117, Springer-Verlag, (1976), pp191.

[206] Z. Li, I.M. Navon and M. Y. Hussaini, Analysis of the singular vectors of the full-physics Florida State University Global Spectral Model. , Tellus , Vol. 57A , (2005) 560-574 .

[207] P. Lopez, Moreau, E., A convection scheme for data assimilation: Description and initial tests. Q. J. R. Meteorol. Soc., 131,(2005) 409-436.

[208] J.-F. Louis and M. Zivković, Optimizing a Local Weather Forecast System, AER, Inc., Final Report, Cambridge, MA, NSF Grant Number ISI-9100720, (1994).

[209] A.N. Kolmogorov, Interpolation extrapolation of stationary random sequences. Bulletin of the Academy of Sciences, USSR series on Mathematics, 15 (1941). (Translation by RAND Corporation, memorandum RM-30090-PR, April (1962).

[210] D.M. Legler, I.M. Navon, J.J. O'Brien, Objective analysis and assimilation of meteorological observations: theoretical aspects. Mon. Wea. Rev. 38A (1989) 97-110.

[211] D.M. Legler, I.M. Navon, VARIATM:A Fortran code for objective analysis of pseudo-stress with large-scale conjugate-gradient minimization. Computers and Geosciences 17(1)(1991) 1-21.

[212] J.M. Lewis, J.C. Derber, The use of adjoint equations to solve a variational adjustment problem with advective constraints. Tellus 37A (1985) 309-322.

[213] J.M. Lewis, S. Lakshmivarahan, Sudarshan Dhall , Dynamic data assimilation, a least squares approach. Cambridge University Press. (2006)

[214] Y. Li, A note on the uniqueness problem of variational adjustment approach to four-dimensional data assimilation. J. Meteor. Soc. Japan 69 (1991) 581-585.

[215] Y. Li, I.M. Navon, P. Courtier, P. Gauthier, Variational data assimilation with a semi-implicit semi-Lagrangian global shallow-water equation model and its adjoint. Mon. Wea. Rev. 121 (1993) 159-169.

[216] Y. Li, K. K. Droegemeier, The influence of diffusion and associated errors on the adjoint data assimilation technique. Tellus 45A (1993) 1-14.

[217] Y. Li, I.M. Navon, W. Y. Yang, X. L. Zou, J. R. Bates, S. Moorthi, R. W. Higgins, 4-D assimilation experiments with a multilevel semi-Lagrangian semi-implicit GCM. Mon. Wea. Rev.122(5) (1994), 966-983.

[218] J.L. Lions, Controle Optimal des Systemes Gouvernes par des Equations aux Derives Partielles. Dunod et Gauthiers-Villars, Paris, (1968).

[219] J.L.
Lions, Optimal control of systems governed by partial differential equations. Translated by S. K. Mitter, Springer-Verlag, Berlin-Heidelberg,(1971), pp 404.

[220] D.C. Liu, Jorge Nocedal, On the limited memory BFGS method for large scale minimization. Math. Prog. 45 (1989) 503-528.

[221] A.C. Lorenc, A global three-dimensional multivariate statistical interpolation scheme. Mon. Wea. Rev. 109 (1981) 701-721.

[222] A.C. Lorenc, Analysis-methods for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society 112 (474) (1986) 1177-1194.

[223] A. C. Lorenc, Optimal nonlinear objective analysis. Q. J. R. Meteorol. Soc. 114 (1980) 205 -240.

[224] A. C. Lorenc, A practical approximation to optimal four dimensional objective analysis. Mon. Wea. Rev. 116 (1980) 730-745.

[225] A. C. Lorenc, R. S. Bell, B. Macpherson, The meteorological office analysis correction data assimilation scheme. Q. J. R. Meteorol. Soc. 117 (1991) 59-89.

[226] A.C. Lorenc, S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders, The Met. Office global three-dimensional variational data assimilation scheme. Quart. J. Roy. Meteor. Soc., 126,(2000), 2991-3012.

[227] A.C. Lorenc, Rawlins F. Why does 4D-Var beat 3D-Var? Quarterly Journal of the Royal Meteorological Society 131 (613), (2005) 3247-3257 Part C .

[228] E.N. Lorenz, A study of the predictability of a 28-variable atmospheric model. Tellus, 17,(1965) 321-333.

[229] E.N. Lorenz, K.A. Emanuel, Optimal sites for supplementary weather observations: simulation with a small model. J. Atmos. Sci., 55, (1998) 399-414.

[230] P. Lynch, X.Y. Huang, Initialization of the HIRLAM Model Using a Digital Filter. Mon Wea Rev. 120(6) (1992) 1019-1034.

[231] W. Lyne, R. Swinbank, N. Birch, A data assimilation experiment and the global circulation during the FGGE special observing periods. Quart. J. Roy. Meteor. Soc. 108 (1982) 575-594.

[232] P. Malanotte-Rizzoli, Modern Approaches to Data Assimilation in Ocean Modeling. Elsevier, Amsterdam, (1996), pp 455.

[233] G.I. Marchuk, Numerical Calculation of Nuclear Reactors. Moskwa, Atomizdat, (1958), 162 pp.

[234] G.I. Marchuk, Chislennye Metody v Prognoze Pogodi, (Numerical Methods in Weather Forecasting), Gidrometeoizdat, Leningrad,(1967), 356 pp.

[235] G.I. Marchuk, Numerical solution of the problem of dynamics of atmosphere and ocean (in Russian). Leningrad, Gidrometeoizdat,(1974), pp 303.

[236] G.I. Marchuk, Formulation of theory of perturbations for complicated models. Appl. Math. Optimization 2 (1975) 1-33.

[237] M.J. Martin, N.K. Nichols, M.J. Bell, treatment of Systematic Errors in Sequential Data Assimilation, The Met Office, Ocean Applications Division, Tech. Note No. 21, (1999).

[238] M.J. Martin, M.J. Bell, N.K. Nichols, Estimation of systematic error in an equatorial ocean model using data assimilation, in M.J. Baines (ed.), Numerical Methods for Fluid Dynamics VII, ICFD, Oxford,(2001), pp. 423-430.

[239] M.J. Martin, M. Bell, N. Nichols, Estimation of systematic error in an equatorial ocean model using data assimilation. International Journal for Numerical Methods in Fluids 40 (2002) 435-444.

[240] T. Matsuno, Numerical integrations of the primitive equations by simulated backward difference scheme. J. Meteor. Soc. Japan 44 (1966) 76-84.

[241] D. McLaughlin, O'Neill A, Derber J, M.Kamachi , Opportunities for enhanced collaboration within the data assimilation community Quarterly Journal of the Royal Meteorological Society 131 (2005) (613), 3683-3693, Part C.

[242] R.D. McPherson, Progress, problems, and prospects in meteorological data assimilation. Bulletin of the American Meteorological Society 56 (11)(1975) 1154-1166.

[243] R.D. McPherson, Bergman KH, Kistler RE, et al., NMC operational global data assimilation system . Mon. Wea. Rev. 107 (11) (1979) 1445-1461.

[244] D. Menemenlis, M. Chechelnitsky, Error estimates for an ocean general circulation model from altimeter and acoustic tomography data. Mon. Weather Rev. 128 (2000) 763-778.

[245] R.N. Miller, M Ghil, Gauthiez F Advanced data assimilation in strongly nonlinear dynamical-systems. Journal of the Atmospheric Sciences 51 (8)(1994) 1037-1056.

[246] H.L. Mitchell, R. Daley, Discretization error and signal/error correlation in atmospheric data assimilation. I: All scales resolved. Tellus 49A (1997) 32-53

[247] K. Miyakoda, L. Umscheid, D. H. Lee, J. Sirutis, R. Lusen, F. Pratte, The near-real-time, global, four-dimensional analysis experiment during the GATE period, Part I. Journal of the Atmospheric Sciences, 33 (4) (1976) 561-591.

[248] F. Molteni, Buizza, R., Palmer, T. N., Petroliagis, T., The new ECMWF ensemble prediction system: methodology and validation. Q. J. R. Meteorol. Soc., 122,(1996) 73-119.

[249] A.M Moore, B.F. Farrell, Rapid perturbation growth on spatially and temporally varying oceanic flows determined using adjoint method: application to the gulf stream. J. Phys. Oceanogr., (1993), 1682-1702.

[250] J.L. Morales, J. Nocedal, Algorithm 809: PREQN: Fortran 77 subroutines for preconditioning the conjugate gradient method, ACM Trans. Math. Soft. 27(1) (2001) 83-91.

[251] J.L. Morales, J. Nocedal, Enriched methods for large-scale unconstrained optimization, Comput. Optim. Appl. 21 (2002) 143-154.

[252] R. E. Morss, Emanuel, K.A., and C. Snyder, Idealized adaptive observation strategies for improving numerical weather prediction. J. Atmos. Sci., 58(2), (2001) 210-234.

[253] P. Morel, O. Talagrand, Dynamic approach to meteorological data assimilation, Tellus 26 (3) (1974) 334-344

[254] S.G. Nash, Truncated-Newton methods. Ph.D. Thesis, Computer Science Department, Stanford University, CA, (1982).

[255] S.G. Nash, Newton-type minimization via the Lanczos method. SIAM J. Numer. Anal. 21(4) (1984) 770-788.

[256] S.G. Nash, Truncated-Newton methods for large-scale function minimization. Applications of Nonlinear Programming to Optimization and Control, H.E. Rauch (ed.), Pergamon Press, Oxford, (1984), 91-100.

[257] S.G. Nash, User's guide for TN/TNBC: Fortran routines for nonlinear optimization. Tech. Rep. 397 (1984) pp17.

[258] S.G. Nash, Solving nonlinear programming problems using truncated-Newton techniques. Numerical optimization, P. T. Boggs, R. H. Byrd and R. B. Schnabel (ed.), SIAM, Philadelphia, (1984), 119-136.

[259] S.G. Nash, 1985: Preconditioning of truncated-Newton methods. SIAM J. Sci. Stat. Comput. 6(3) (1985) 599-616.

[260] S.G. Nash, A. Sofer, Block truncated-Newton methods for parallel optimization. Math. Prog.,45 (3), (1989), 529-546.

[261] S.G. Nash, A. Sofer, Linear and Nonlinear Programming. McGraw-Hill, (1996), pp 692.

[262] I.M. Navon, H.A. Riphagen, The Impact of GARP Global Data Sets DST-5 and DST-6 on Predictability Degradation in the Southern Hemisphere in: Numerical Experimentation Programme, 1, eds. Jan Rutherford, (1979) 2.1-2.5

[263] I.M. Navon, The Effect of Global Data Sets on Predictability Degradation in the Southern Hemisphere, (1981).Int. Report: 308 CSIR, Pretoria, South-Africa.

[264] I.M. Navon, Implementation of A Posteriori Methods for Enforcing Conservation of Potential Enstrophy and Mass in Discretized Shallow-Water Equations Models. Mon. Wea. Rev. 109 (1981) 946-958.

[265] I.M. Navon, R. de Villiers, Combined penalty multiplier optimization methods to enforce integral invariants conservation. Mon. Wea. Rev. 111 (1983) 1228-1243.

[266] I.M. Navon, A review of variational optimization methods in meteorology. Proc. of the International Symposium on Variational Methods in Geosciences, Y. K. Sasaki (ed.), University of Oklahoma, Norman, OK , (1985), October 15-17.

[267] I.M. Navon, A review of variational and optimization methods in meteorology. Festive Volume of the International Symposium on Variational Methods in Geosciences (Y.K. Sasaki, Ed.), Elsevier Science Pub. Co. Developments in Geo-mathematics 5 (1986) 29-34.

[268] I.M. Navon, D.M. Legler, Conjugate gradient methods for large scale minimization in meteorology. Mon. Wea. Rev. 115 (1987) 1479-1502.

[269] I.M. Navon, The Bayliss-Isaacson algorithm and the constraint restoration method are equivalent. Meteorology and Atmospheric Physics 13 (1987) 143-152.

[270] I.M. Navon, X. Zou, Application of the adjoint model in meteorology. In Proceedings of the international conference on automatic differentiation of algorithms: Theory, Implementation and Application, SIAM Publications, A. Griewank and G. Corliss (Eds.), SIAM, Philadelphia, PA, (1991).

[271] I.M. Navon, X. L. Zou, J. Derber and J. Sela, Variational data assimilation with an adiabatic version of the NMC Spectral Model. Mon. Wea. Rev. 122(1992) 1433-1446.

[272] I.M. Navon, X.Zou, M. Berger, P.K.H. Phua, T.Schlick, F.X.Le Dimet, Numerical experience with Limited Memory Quasi-Newton and truncated Newton Methods, Optimization Techniques and Applications, Vol 1, Edited by K.H. Phua et al. World Scientific Publishing Co., (1992), 1232pp, 33-48.

[273] I.M. Navon, X.Zou, M. Berger, P.K.H. Phua, T.Schlick, F.X.Le Dimet, Testing for Reliability and Robustness of Optimization Codes for Large Scale Optimization Problems, Optimization Techniques and Applications, Vol 1, Edited by K.H. Phua et al. World Scientific Publishing Co., (1992), 1232pp, 445-480.

[274] I. M. Navon, Practical and Theoretical Aspects of Adjoint Parameter Estimation and Identifiability in Meteorology and Oceanography, Dynamics of Atmospheres and Oceans. Special Issue in honor of Richard Pfeffer , 27, Nos.1-4 (1998) 55-79.

[275] I.M. Navon , Dacian N. Daescu, Z. Liu , The Impact of Background Error on Incomplete Observations for 4D-Var Data Assimilation with the FSU GSM. , V.S. Sunderam et al.(Eds.):Computational Science-ICCS 2005, LNCS 3515,(2005), pp 837-844. Springer Verlag, Heidelberg.

[276] H.E. Ngodock, Data Assimilation and sensitivity Analysis. Ph. D. Thesis, LMC-IMAG Laboratory, University Joseph Fourier, Grenoble, France, (1996), pp 213.

[277] N.K. Nichols, 2003, Treating Model Error in 3-D and 4-D Data Assimilation. Proceedings of the NATO Advanced Study Institute on Data Assimilation for the Earth System, Maratea, Italy, June (2002).

[278] J. Nocedal, Updating quasi-Newton matrices with limited storage. Mathematics of Computation, (1980), 773-782.

[279] J. Nocedal, Theory of algorithms for unconstrained optimization, Acta Numerica, 1, 199–242, (1991).

[280] J. Nocedal, S.J. Wright, Numerical optimization, Springer, (1999).

[281] E.M. Oblow, An automated procedure for sensitivity analysis using computer calculus, Tech. Report ORNL/TM-8776, ORNL, (1983).

[282] J. Pailleux, W. Heckly, D. Vasiljevic, J. N. Thepaut, F. Rabier, C. Cardinali, E. Andersson, Development of a variational assimilation system. ECMWF Research Department Technical Memorandum 179, Reading, U.K., (1991).

[283] J. Pailleux, Organisation of 3D variational analysis within the "IFS/ARPEGE" project future plans at meteo France. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, (1992), 37-46.

[284] T.N. Palmer, Gelaro, R., Barkmeijer, J., and R. Buizza, Singular vectors, metrics, and adaptive observations. J. Atmos. Sci., 55, (1998) 633- 653.

[285] Panel on Model-Assimilated Data Sets for Atmospheric and Oceanic Research, Four- Dimensional Model Assimilation of Data: A Strategy for the Earth System Sciences . National Academy Press, Washington, D.C., (1991), pp78.

[286] H.A. Panofsky, Objective weather-map analysis. J. Appl. Meteor. 6 (1949) 386-392.

[287] S.K. Park, E. Kalnay, Application of the quasi-inverse method to storm-scale data assimilation. Abstracts, 5th SIAM Conference on Mathematical and Computational Issues in the Geosciences, March 24-27, (1999), San Antonio, Texas, SIAM, 104.

[288] D.F. Parrish, J.C. Derber, The National-Meteorological-Centers' spectral statistical-interpolation analysis system. Mon. Wea. Rev. 120 (8) (1992), 1747-1763.

[289] V. Penenko, N.N. Obratsov, A variational initialization method for the fields of the meteorological elements. Meteorol. Gidrol. 11 (1976), 1-11. (English Translation)

[290] F.G. Pin, E. M. Oblow, J. E. Horwedel, and J. L. Lucius, ADGEN: An Automated Adjoin Code Generator for Large-Scale Sensitivity Analysis, Trans. Am. Nucl. Soc. 55, 311 (1987).

[291] B.T. Polyak , Tsypkin YZ Pseudogradient adaptation and training algorithms. Automation and Remote Control 34 (3) (1973) 377-397.

[292] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, E.F. Mishchenko, The Mathematical Theory of Optimal Processes. John Wiley & Sons, Inc. New York, (1962).

[293] M.J.D. Powell, A method for nonlinear constraints in minimization problems. Optimization, R. Fletcher, Ed., Academic Press, Chap. 19 (1969) pp 283-298.

[294] M.J.D. Powell, Restart procedures for the conjugate-gradient method. Math. Program. 12 (1977) 241-254.

[295] Z.X. Pu, E. Kalnay , J. Sela, Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method Mon. Wea. Rev 10 (1997) 2479-2503.

[296] Z.X. Pu, E. Kalnay, Targeting observations with the quasi-inverse linear and adjoint NCEP global models: Performance during FASTEX.Quart. J. Roy. Meteor. Soc. 125 (561), (1999) pp 3329-3337.

[297] F.P. Rabier, Courtier, O. Talagrand, An application of adjoint models to sensitivity analysis. Beitr. Phys. Atmos. (3) (1992) 177-192.

[298] F. Rabier, P. Courtier, Four-dimensional assimilation in the presence of baroclinic instability. Q. J. R. Meteorol. Soc. 118 (1992) 649-672.

[299] F. Rabier, P. Courtier, J. Pailleux, O. Talagrand, J. N. Thepaut, D. Vasiljevic, Comparison of four-dimensional variational assimilation with simplified sequential assimilation. In Proc. of ECMWF Workshop on variational assimilation with emphasis on three-dimensional aspects, Shinfield Park, Reading RG2 9AX, UK, (1992), 271-325.

[300] F. Rabier, P. Courtier, J. Pailleux, O. Talagrand, D. Vasiljevic, A comparison between four-dimensional variational assimilation and simplified sequential assimilation relying on three-dimensional variational analysis. Q. J. R. Meteorol. Soc. (1993) 845-880.

[301] F. Rabier, Klinker E, Courtier P, Hollingsworth A., Sensitivity of forecast errors to initial conditions. Quart JR Meteorol Soc 122, (1996) 121-150.

[302] F. Rabier, J-F. Mahfouf, M. Fisher, H. Jarvinen, A. Simmons, E. Andersson, F. Bouttier, P. Courtier, M. Hamrud, J. Haseler, A. Hollingsworth, L. Isaksen, E. Klinker, S. Saarinen, C.Temperton, J-N. Thepaut, P. Unden, and D. Vasiljevic, Recent experimentation on 4D-Var and first results from a Simplified Kalman Filter. October (1997).ECMWF Technical Memorandum No. 240.

[303] F. Rabier, Jaarvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A. 2000, The ECMWF operational implementation of 4D variational assimilation Part I: experimental results with simplified physics. Q. J. R. Meteorol. Soc., 126, (2000) 1143-1170.

[304] F. Rabier, Overview of global data assimilation developments in numerical weather-prediction centres. Quarterly Journal of the Royal Meteorological Society 131 (613) (2005) 3215-3233.

[305] L.B. Rall, Automatic differentiation: Techniques and applications, Lecture Notes in Computer Science, vol. 120, Springer-Verlag, Berlin, (1981).

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

[307] M.K. Ramamurthy, I.M. Navon, The conjugate-gradient variational analysis and initialization method: An application to MONEX SOP-2 Data. Mon. Wea. Rev. 120(10) (1992) 2360-2377.

[308] M.K. Ramamurthy, T.Y. Xu, Continuous data assimilation experiments with the NMC Eta model: A GALE IOP 1 Case study. Mon. Wea. Rev. 121 (11)(1993) 3082-3105.

[309] Richard Courant and David Hilbert, Partial Differential equations. Methods of Mathematical Physics. Wiley-Interscience, New York, (1962).

[310] L.F. Richardson, Weather prediction by numerical processes. Cambridge University Press. (1922) Reprinted by Dover (1965, New York). With a New Introduction by Sydney Chapman, pp. xvi+236.

[311] C.D. Rodgers, Inverse methods for atmospheres: theories and practice.(2000) World Scientific Publ., Singapore.

[312] N. Rostaing-Schmidt, Differentiation automatique: Application a un probleme d'optimisation. Ph.D. Thesis. Universite de Nice Sophia-Antipolis. (1993).

[313] I.D. Rutherford, Data assimilation by statistical interpolation of forecast error fields. J. Atmos. Sci. 29 (1972) 809-815.

[314] S. Saha, Response of the NMC MRF model to systematic error correction within the integration. Mon. Weather Rev. 120 (1992) 345-360.

[315] Y.K. Sasaki, A fundamental study of the numerical prediction based on the variational principle, J. Meteor. Soc. Japan 33 (1955) 262-275.

[316] Y.K. Sasaki, An objective analysis based on the variational method, J. Meteor. Soc. Japan 36 (1958) 77-88.

[317] Y.K. Sasaki, Proposed inclusion of time-variation terms, observational and theoretical in numerical variational objective analysis, J. Meteor. Soc. Japan 47 (1969) 115- 203.

[318] Y.K. Sasaki, Some basic formalisms in numerical variational analysis, Mon. Wea. Rev. 98 (1970a) 857-883.

[319] Y.K. Sasaki, Numerical variational analysis formulated under the constraints as determined by long-wave equations as a low-pass filter, Mon. Wea. Rev. 98 (1970b) 884-898.

[320] Y.K. Sasaki, Numerical variational analysis with weak constraint and application to the surface analysis of severe storm gust, Mon. Wea. Rev. 98 (1970c) 899-910.

[321] Y.K. Sasaki, A theoretical interpretation of anisotropically weighted smoothing on the basis of numerical variational analysis, Mon. Wea. Rev. 99 (1970d) 698-707.

[322] Y.K. Sasaki, Numerical variational analysis formulated under the constraints as determined by long-wave equations as a low-pass filter. Mon. Wea. Rev. 98 (1970) 884-898.

[323] Y.K. Sasaki, Variational design of finite-difference scheme for initial-value problem of conservative system. Dept. Meteorology University of Oklahoma, (1975), 31.

[324] Y.K. Sasaki, Variational design of finite-difference schemes for initial value-problems with an integral invariant Journal of Computational Physics 21 (3) (1976) 270-278.

[325] Y.K. Sasaki, Variational design of finite-difference schemes for initial value problems with a global divergent barotropic model. Contrib. Atmos. Phys. 50 (1977) 184-289.

[326] T. Schlick, A. Fogelson, TNPACK: Truncated Newton minimization Package for large-scale problems: I. Algorithm and usage. ACM TOMS 18(1) (1992) 46-70.

[327] T. Schlick, A. Fogelson, TNPACK: Truncated Newton minimization package for large-scale problems: II. Implementation examples. ACM TOMS 18(1) (1992) 71-111.

[328] R.S. Seaman, Some real data tests of the interpolation accuracy of Bratseth's successive corrections method. Tellus 40A (1988) 173-176.

[329] J.H. Seinfeld, C. Kravaris, Distributed parameter identification in geophysics -petroleum reservoirs and aquifers. in Distributed parameter control systems, Ed. S.G. Tzafestas, (1982), pp 367-390, Pergamon Press, Oxford.

[330] D.F. Shanno, K. H. Phua, Remark on algorithm 500 C a variable method subroutine for unconstrained nonlinear minimization. ACM TOMS (1980), 618-622.

[331] N.Z. Shor., Minimization Methods for Nondifferentiable Functions. Springer-Verlag, (1985), 162 pp. (Translated from Russian by K. C. Kiwiel and A. Ruszczynski.)

[332] A. J. Simmons, Hollingsworth, A. Some aspects of the improvement in skill of numerical weather prediction. Q. J. R. Meteorol. Soc., 128,(2002) 647-677.

[333] G.L.G. Sleijpen, H.A. Van der Vorst, A Jacobi-Davidson iteration method for linear eigenvalue problems. SIAM J Matrix Anal A (2) (1996) 401-425 .

[334] O.M. Smedstad, J.J. O'Brien, Variational data assimilation and parameter estimation in an equatorial Pacific Ocean model. Prog. Oceanogr. 26 (1991) 1250-1277.

[335] D.A. Stainforth, T. Aina, C. Christensen, M. Collins, N. Faull,D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R. A. Spicer, A. J. Thorpe and M. R. Allen, Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433 (2005), 403-406.

[336] D.R. Stauffer, J. Bao, Optimal Determination of nudging coefficients using the adjoint equations, Tellus A (1993) 358-369.

[337] D.R. Staufer, M.L. Seaman, Use of four-dimensional data assimilation on a limited area mesoscale model. Part I: Experiments with synoptic-scale data. Mon. Wea. Rev. 26 (1990) 1250-1277.

[338] D.J. Stensrud, J.W. Bao, Behaviors of variational and nudging assimilation techniques with a chaotic low-order model. Mon. Wea. Rev. 120 (1992) 3016-3028.

[339] J.J. Stephens, A variational approach to numerical weather analysis and prediction. Ph.D. Thesis, Texas A and M University, College-Station, TX, 77863, Rep. 3 (1966) pp243.

[340] T.F. Stocker, Models change their tune. Nature 430 (2004) 737-738.

[341] N.-Z. Sun, Inverse Problems in Groundwater Modeling (Theory and Applications of Transport in Porous Media). Springer, (1994), 352pp.

[342] N-Z Sun, Structure reduction and robust experimental design for distributed parameter identification , Inverse Problems 21 (2) (2005) 739-758.

[343] N.-Z. Sun, and A. Y. Sun, Inverse Methods for Parameter Estimations. Encyclopaedia of Hydrological Sciences. Edited by M. Anderson.2005 John Wiley and Sons, Ltd, (2006).

[344] W.C. Thacker, The Role of Hessian Matrix in Fitting models to Measurements. J. Geophys. Res., (1989), 6177-6196.

[345] I.G. Tadjbakhsh , Utilization of Time-Dependent Data in Running Solution of Initial Value Problems. Journal of Applied Meteorology 8(3) (1969) 389-391.

[346] L L. Takacs, Effects of Using A Posteriori Methods for the Conservation of Integral Invariants. Mon Wea Rev. 116 (1988) 525-545

[347] O. Talagrand, K. Miyakoda, Assimilation of past data in dynamical analysis .2. Tellus 23(4-5) (1971) 318-327.

[348] O. Talagrand, On the mathematics of data assimilation Tellus 33 (4) (1981) 321-339 .

[349] O. Talagrand, P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation-Part 1. Theory. Q. J. R. Meteorol. Soc. 113 (1987) 1311-1328.

[350] O. Talagrand, Assimilation of observations, an introduction . Journal of the Meteorological Society of Japan 75 (1B) (1997) 191-209.

[351] A. Tarantola, Inverse problem theory. methods for data fitting and model parameter estimation. Elsevier, Amsterdam,(1987), pp613.

[352] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, (2004), pp 352.

[353] W.C. Thacker, R.B. Long, Fitting dynamics to data. J. Geophys. Res. 93 (1986) 1227-1240.

[354] W.C. Thacker, Oceanographic inverse problems. Physica D, (1992), 16-37.

[355] J.-N. Thepaut, P. Moll, Variational inversion of simulated TOVS radiances using the adjoint technique. Quart. J. Roy. Meteorol. Soc., (1990), 1425-1448.

[356] J.N. Thepaut, P. Courtier, Four-dimensional variational assimilation using the adjoint of a multilevel primitive-equation model. Q. J. R. Meteorol. Soc. 117 (1991) 1225-1254.

[357] J.N. Thepaut, D. Vasiljevic, P. Courtier, J. Pailleux, Variational assimilation of meteorological observations with a multilevel primitive-equation model. Q. J. R. Meteorol. Soc. 119 (1993) 153-186

[358] J. -N. Thepaut, Courtier, P., Belaud, G. and Lemaitre, G. Dynamical structure functions in a four-dimensional variational assimilation: A case study. Q. J. R. Meteorol. Soc., 122, (1996) 535-561.

[359] H.J. Thiébaux, L.L. Morone, Short-term systematic errors in global forecasts: their estimation and removal. Tellus 42A (1990) 209-229

[360] P.D. Thompson, Reduction of Analysis Error Through Constraints of Dynamical Consistency. Journal of Applied Meteorology 8 (1969) 738-742.

[361] A. N. Tichonov, Regularization of ill-posed problems. Dokd. Akad. Nauk, USSR, (1963), 49-52.

[362] A.N. Tichonov, V.Y. Arsenin, Solutions of ill-posed problems. Halsted Press, New-York, (1977), pp 258.

[363] Z. Toth, Szunyogh, I., Bishop, C., Majumdar, S., Morss, R., Moskaitis, J., Reynolds, D., Weinbrenner, D., Michaud, D., Surgi, N., Ralph, M., Parrish, J., Talbot, J., Pavone, J. and Lord, S. ,Adaptive observations at NCEP: Past, present, and future. In Proceedings of the symposium on observations, data assimilation, and probabilistic prediction, 13-17 January (2002), Orlando, Fla. American Meteorological Society, Boston, USA.

[364] Y. Trémolet, Model error in variational data assimilation, Pp. 361-367 in Proceedings of the ECMWF annual seminar on recent developments in data assimilation for atmosphere and ocean, 8-12 September (2003), Reading, UK.

[365] Y. Trémolet, Model bias in data assimilation - 2006 ECMWF/EUMETSAT NWP-SAF Workshop on Bias estimation in data assimilation, 8-11 November (2005) ,ECMWF,Reading. pp 177-184.

[366] Y. Trémolet, Accounting for an imperfect model in 4D-Var NCEP talk .February (2006).

[367] M. D. Tsyrulnikov, Stochastic modelling of model errors: A simulation study. Q. J. R. Meteorol. Soc. 131 (2005) pp. 3345-3371

[368] F. Uboldi, M. Kamachi, Time-space weak-constraint data assimilation for nonlinear models. Tellus 52 A (2000) 412-421

[369] S. Uryasev, New Variable metric algorithms for non-differentiable optimization problems. Journal of Optimization Theory and Applications 71(2) (1991) 359-388.

[370] F. Veerse, Thepaut, J.-N. Multiple-truncation incremental approach for four-dimensional variational data assimilation. Q. J. R. Meteorol. Soc., 124,(1998) 1889-1908.

[371] F. Veerse, Variable-storage quasi-Newton operators as inverse forecast/analysis error covariance matrices in data assimilation. INRIA Technical Report 3685 (1999) Theme 4, pp 28.

[372] P.A. Vidard, E. Blayo, F.-X. Le Dimet, A. Piacentini, 4-D variational data analysis with imperfect model. Flow, Turbulence and Combustion 65 (2000) 489-504

[373] T. Vukicevic, Nonlinear and linear evolution of initial forecast errors. Mon. Wea. Rev. 119 (1991) 1602-1611.

[374] T. Vukicevic, R.M. Errico, Linearization and adjoint of parametrized moist diabatic processes. NCAR 1030/92-1,(1992).

[375] Z. Wang, I. M. Navon, F. X. Le Dimet, X. Zou, The second order adjoint analysis: Theory and application. Meteorol. and Atmos. Phy. 50 (1992) 3-20.

[376] Z. Wang, I. M. Navon and X. Zou, The adjoint truncated Newton algorithm for large-scale unconstrained optimization. Tech. Rep. FSU-SCRI-92-170, Florida State University, Tallahassee, Florida, (1993), pp 44.

[377] Z. Wang, Variational Data Assimilation with 2-D Shallow Water Equations and 3-D FSU Global Spectral Models, Ph. D. Dissertation, Department of Mathematics, College of Arts and Sciences, The Florida State University, (1993), 235 pp.

[378] Z. Wang, I. M. Navon, X. Zou, and F.X. Le Dimet, A Truncated Newton Optimization Algorithm in Meteorology Applications with Analytic Hessian Vector Products. Computational Optimization and Applications, 4, 241-262 (1995)

[379] Z. Wang, K.K. Droegemeier, L. White, I. M. Navon, Application of a New Adjoint Newton Algorithm to the 3-D ARPS Storm Scale Model Using Simulated Data. Monthly Weather Review 125 (10), (1997) 2460-2478.

[380] Z. Wang, K. Kelvin, Droegemeier, L. White, The Adjoint Newton algorithm for large-scale unconstrained optimization in meteorology applications. Computational Optimization and Applications 10 (1998) 283-320.

[381] A. Weaver, P. Courtier, Correlation modelling on the sphere using a generalized diffusion equation Q.J.R.M.S. 127 (2001), 1815-1846.

[382] C.F. Weber, D.G. Cacuci, E.M. Oblow, Differential sensitivity theory for nonlinear equations with nonlinear responses. Trans. Am. Nucl. Soc. ; Vol/Issue: 33; American Nuclear Society meeting; 12 Nov (1979); San Francisco, CA, USA.

[383] W. Wergen, Effect of model errors in variational assimilation Tellus 44 A (1992) 297-313.

[384] N. Wiener, Extrapolation, interpolation and smoothing of stationary time series with engineering applications. M.I.T. Press, Cambridge, MA, (1949), pp 163.

[385] E.P. Wigner, , Effects of small perturbations on pile period, Chicago Rep., CP-G 3048.(1945)

[386] C. Wunsch, The Ocean Circulation Inverse Problem. Cambridge Univ. Press., (1996), pp442.

[387] L. Xu, T. Rosmond, R. Daley, Development of NAVDAS-AR: formulation and initial tests of the linear problem. Tellus 57A (2005) 546-559

[388] W.W.G. Yeh, Review of parameter estimation procedures in groundwater hydrology. Water Res. Rev. 22 (1986) 95-108.

[389] A. Yaglom, An Introduction to the Theory of Stationary Random Functions, Dover, NY, (1962).

[390] W. Yang, I. M. Navon, P. Courtier,A New Hessian Preconditioning Method Applied to Variational Data Assimilation Experiments Using Adiabatic Version of NASA/GEOS-1 GCM. Monthly Weather Review 124 (5) (1996) 1000-1017.

[391] S. Zhang, X Zou, J Ahlquist, I.M. Navon and J.G. Sela, Use of Differentiable and Non-Differentiable Optimization Algorithms for Variational Data Assimilation with Discontinuous Cost Functions, Monthly Weather Review , 128, No. 12 ,(2000) 4031-4044.

[392] K. Zhu, I.M. Navon, X. Zou, Variational Data Assimilation with a Variable Resolution Finite-Element Shallow-Water Equations Model. Monthly Weather Review 122 (5) (1994) 946-965.

[393] J. Zhu, M. Kamachi, An adaptive variational method for data assimilation with imperfect models. Tellus 52 A (2000) 265-279.

[394] Y.Q. Zhu, I.M. Navon ,Impact of parameter estimation on the performance of the FSU Global Spectral Model using its full physics adjoint, Monthly Weather Review 127 (7) (1999) 1497-1517.

[395] X. Zou, I.M. Navon, F. X. Le Dimet, A. Nouailler and T. Schlick, A comparison of efficient large-scale minimization algorithms for optimal control applications in meteorology. Tech. Rep. FSU-SCRI-90-167, Florida State University, Tallahassee, Florida, (1990), pp44.

[396] X. Zou, I.M. Navon, On the effect of gravity oscillation, boundary control and incomplete observations in a shallow water equations model with application to variational data assimilation. Conference on Numerical Weather Prediction, AMS, October 14-18, 1991. Denver, Colorado, B.4, 459-462.

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

[398] X. Zou, I.M. Navon, F.X. Le Dimet, An optimal nudging data assimilation scheme using parameter estimation. Q. J. R. Meteorol. Soc. 118 (1992) 1163-1186.

[399] X. Zou, I.M. Navon, M. Berger, Paul K. H. Phua, T. Schlick, F. X. Le Dimet, Numerical experience with limited-memory quasi-Newton methods and truncated Newton methods. SIAM Jour. on Numerical Optimization, 3 (1993), 582-608.

[400] X. Zou,I. M. Navon, J. Sela, Control of gravity oscillations in variational data assimilation. Mon. Wea. Rev. 121 (1993) 272-289.

[401] X. Zou, X., I.M. Navon, J. Sela, Variational data assimilation with threshold processes using the NMC multilevel primitive-equation model. Tellus 45A (1993) 370-387.

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

[403] X. Zou, I.M. Navon, Variational data assimilation: Some aspects of theory and application. Environmental Modeling, Vol II, P. Zannetti (Editor), Failure Analysis Association Inc, California, USA, Computational Mechanics Publications, Southampton and Boston, (1994) 277-325, 350pp

[404] M. Zupanski, Regional four-dimensional variational data assimilation in a quasi-operational forecasting Environment. Mon. Wea. Rev. 121 (1993) 2396-2408.

[405] D. Zupanski, The effects of discontinuities in the Belts-Miller cumulus convection scheme on four-dimensional variational data assimilation. Tellus 45(5) (1993) pp 511-524.

[406] D. Zupanski, A general weak constraint applicable to operational 4DVAR data assimilation systems. Mon. Weather Rev. 125 (1997) 2274-2292.

[407] M. Zupanski, E. Kalnay, Principles of Data Assimilation. Global Energy and Water Cycles, Cambridge Univ. Press. Ed. K.A. Browning and R.J. Gurney, (1999), 48-54.

[408] D. Zupanski, M. Zupanski, Fine-resolution 4D-VAR data assimilation for the great plains tornado outbreak of 3 May 1999. Weather and Forecasting 121 (2002) 506-525

[409] M. Zupanski, D. Zupanski, T. Vukicevic, K. Eis, T. Haar, CIRA/CSU Four-Dimensional variational data assimilation system. Mon. Weather Rev. 123 (2005) 829-843