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 , USA

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.
Four-Dimensional Variational Data Assimilation (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 addresses 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 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. [87]

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

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 work by Richardson [160] and Charney [31] were based on hand interpolations [92]. This was done 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 [145], Gilchrist and Cressman [73], Cressman [39], Barnes [9] . Use of prior information to supplement rather insufficient data was pioneered by Bergthorsson and Doos [12], Cressman [39] followed by the comprehensive work of Lev Gandin [65].

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[178]) were provided by Le Dimet and Navon[104], an in-depth survey of Ghil and Malanotte-Rizzoli[70] as well as by the outstanding book of Daley " Atmospheric Data Analysis"[43].

A collection of papers by Ghil et al. (1997) in "Data Assimilation in Meteorology and Oceanography: Theory and Practice [71] summarizes state of the art of data assimilation for that period. See also a short survey by Zupanski and Kalnay [399] along with the excellent book of Kalnay [92] "Atmospheric Modeling, Data Assimilation and Predictability". An early effort linking Optimal Interpolation (O.I.) with the variational method was done by Sasaki [163] [164] and in more final form by Sasaki [165,166,167,168,169] which can be viewed as a 3-D- variational (3D-Var) data assimilation approach. It was Lorenc [115] that showed that optimal interpolation (OI) and 3-D VAR were equivalent provided the cost functional assumes the form:

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

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.

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

Theoretical material related to the set-up that led to modern data assimilation may be found in the " Inverse Problem Theory" of Tarantola[179], the optimal control book of Lions[113], the " Perspectives in Flow Control and Optimization" by Max Gunzburger[79] along with "Inverse Modeling of the Ocean and Atmosphere" by Andrew Bennett [11] and "Dynamic Data Assimilation: A Least Squares Approach" by John Lewis et al.[108] ,Cacuci [28] and Cacuci et al [29]).

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, and the incremental method which alleviated the computational burden of 4-D VAR and made it operationally viable at large operational numerical weather prediction (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 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 variational 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, Charney, Halem and Jastrow [32].

Insertion of observations directly into primitive equations models excited spurious inertia-gravity oscillations in the model and required the use of damping schemes [122] for damping the high-frequency components. A full-account of these techniques and the history of continuous data assimilation are provided in the seminal book of Daley [43]. 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 [92], Daley [43], Talagrand [178], Zupanski and Kalnay [209], Ghil et al. (Eds) [71], works of the present author and his collaborators, the review of Ghil and Malanotte-Rizzoli [70] and an early review that remained an unpublished technical report [104].

Panofsky [145] is credited for pioneering the first objective analysis based on 2-D polynomial interpolation. It was followed by Gilchrist and Cressman [73] 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,

$$\min_{a_{ij}} E = \min_{a_{ij}}\{\sum_{k=1}^{K_v} p_v \cdot (E_0^v - E(x_v, y_v))^2 \tag{4}$$

$$+\sum_{k=1}^{K_v} q_v \cdot \{[u_v^0 - u_g(x_v, y_v)]^2 + [v_v^0 - v_g(x_v, y_v)]^2\}\}$$

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[12]. Usually either climatology or a combination of it with forecast was used in the analysis cycle. See also the influential work of Gandin[65], 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[39]. 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"

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

f_i^n - n-th iteration estimate at i th 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[39] proposed the following weights in the SCM (Successive Corrections Method).

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

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

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 [43], Kalnay [92].

Cressman[39]took the coefficient ε^2 to be zero. For noisy data with errors it may lead to erroneous analysis. Taking $\varepsilon^2 > 0$ i.e. assuming observations with errors, allows some impact to the background field. Barnes[9] defined the weights to follow a Gaussian or normal distribution

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

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[21] 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.[108] quote also similar independent work done by Franke and Gordon[63], Franke[64] and Seaman[170].

3. The variational calculus approach

It was introduced in meteorology by Yoshi Sasaki in his Ph.D Thesis[163] and later extended by him to include dynamic model laws [165,166,167,168]. 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 was formulated as constraints in the functionals for the first two formalisms. This approach is now generically referred to as three dimensional variational analysis (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 [114], [115] that the OI and the 3-D VAR approaches are equivalent provided the cost function is defined as

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

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

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

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 and

Lagrange 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[97] and Courant and Hilbert [159].

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 [142].

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

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

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

As an example of a typical application of variational methods, consider work of Sasaki[166, 167,168]. Navon [129] 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 plane.

5. First interlude

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

Charney, Halem and Jastrow[32] 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.[32] experiment suggested continuous data assimilation I.G. Tadjbakhsh (1969) [174]. Problems of real data insertion soon emerged in the form of an inertia-gravity wave shock [42] (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[127] (Miyakoda et al.).

Other approaches were via geostrophic wind correction outside the tropics or nudging also referred to as Newtonian relaxation (Hoke and Anthes)[76], Davis and Turner[49]. See also work of Talagrand[176] [177]. Ghil, Halem and Atlas[69],

Mcpherson[123]. McPherson[235] viewed data assimilation as "a process by something is absorbed into something else". During 1974 Marchuk [119] proposed application of adjoint method in meteorology (Russian article of 1967 [118]) and in 1976 Penenko and Obratsov [148]-used these methods to study adjoint sensitivity.

In 1969, Thompson[184] 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[117] and collaborators used adjoint methods for linear sensitivity analysis problems . Atmospheric issues were also addressed in the same fashion (see Marchuk[119]). Adjoint operators have been introduced by Lagrange[96] and have been used in modern times since Wigner[198] and by many others in different domains.

The advent of optimal control theory of partial differential equations is attributed to Bellman starting in the late 50's Bellman[10] (the Hamilton-Jacobi-Bellman equation) and to Pontryagin (Pontryagin's minimum principle)[150].

The major impetus in this area came from the monograph of Lions[112] 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.[23], D.G. Cacuci[24],[25]. Historically one can trace back linear adjoint sensitivity to work of Wiener (1940-1942). See Cacuci 2004 lecture, [26]. Wiener [197] was the first to interpret physically the adjoint functions (see also Lewins, 1962) as importance functions. As mentioned above Cacuci (1979-1981) [195] presented a complete rigorous theory for adjoint sensitivity of general nonlinear systems of equations.

LeDimet[98] 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), [101]

6. Emergence of early data assimilation works

LeDimet[99], Lewis and Derber[107], Courtier [34], Le Dimet and Talagrand(1986) [101] were the first to work on adjoint data assimilation. Cacuci(1981a,1981b) [24,25]

extended adjoint sensitivity analysis to the fully nonlinear case. Lagrange multiplier methods were presented in detail by Bertsekas[14], while Navon and De Villiers[130] exhibited the method in detail applied to enforcing conservation of integral invariants.

7. Optimal interpolation (OI) methods

Lev Gandin[65] coined the term (OI) but the technique of statistical interpolation can be traced back to Kolmogorov[106] and Wiener[197] and the terminology of optimal interpolation was apparently due to Wiener[197].

A review of the work of these two mathematician is provided in the Yaglom[200] book on stochastic processes (see Lewis et al.[108]). In atmospheric sciences use of statistical interpolation goes back to Eliassen[58] while Krige[94], used it in the mining industry.

Use of least-squares 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[66]. 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[19]).

One can show that the best linear unbiased estimator [178](Talagrand(1997), Bouttier and Courtier[19]) may be obtained as the solution of the following variational optimization problem.

$$\begin{aligned} \min J &= (X - X_b)^T B^{-1}(X - X_b) + (y - H(X))^T R^{-1}(y - H(X)) \\ &= J_b(X) + J_o(X) \end{aligned} \tag{33}$$

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 $R \to R$ 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[114] and Lorenc[115]. The most important advantage of using statistical interpolation schemes such as OI and 3-D VAR instead of empirical schemes such as SCM [39] 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)[92], Navon et al.[138]. 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 [61]. The first approach by Hollingsworth and

Lönnenberg[85] 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" (National Meteorological Center), now National Centers for Environmental Prediction(NCEP) method having been introduced by Parrish and Derber[147]. 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 [61] who presents also modern approaches to background error covariance matrix construction.

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[101], Rabier[157]).

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

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

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

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

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

If

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

(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[101] 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 $\|,\|$ 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[14]) 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[142], Nash and Sofer[132]) For shortcomings of penalty and duality algorithms see Bertsekas[14] and Navon and De Villiers[130].

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

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

This algorithm was initially proposed by Hestenes [82] and independently by Powell [151]. Here $\rho>0$ is the quadratic penalty coefficient.

11. Optimal control view point

In optimal control of partial differential equations developed by Lions[112,113] 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[79] 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

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

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

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

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

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

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[98] technical report were by Lewis and Derber[107] and LeDimet and Talagrand[101] as well as Courtier[34]. These research efforts started the meteorological optimal control application called “adjoint operator” approach.

Work of Navon and De Villiers[130] on augmented Lagrangian methods is related to the same topic and is referred to in the early work of LeDimet and Talagrand [101]

John Lewis and John Derber[107] were the first to present application of adjoint method, having read the report of Francois le Dimet (1982) [99] and inspired by earlier work of Thompson[184] . Lorenc[115] presented a detailed account of state of theory in data assimilation for that period.

It became soon apparent that size and complexity of atmospheric equations is such that enormous computational resources were required-limiting applications of 4-D VAR to undergo drastic approximations for actual operation forecast circumstances.

Penenko and Obratsov[148] used adjoint data assimilation to perform simple experiments on a linear model (see Talagrand and Courtier[177]) , while Derber[52] used it in his Ph.D thesis to adjust analysis to a multi-level quasi-geostrophic model.

Hoffmann[83] was the next to use 4-D VAR (even though he used a simplified primitive equation model and in order to estimate the gradient he perturbed in turn all the components of the initial state.)

Talagrand and Courtier[177] 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[35].

13. OI, 3-D VAR and Physical Space Analysis Scheme (PSAS).

Lorenc [115] 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

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

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[115], Kalnay[92] and Courtier[38]. See also research work of Da Silva et al.[47] 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[92], Lorenc[115] OI, 3-D VAR, 4-D VAR and PSAS are mathematically equivalent but 3-D VAR and related PSAS have the advantage with respect to 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[92].

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[115], Kalnay [92] and using suggestion of Jim Purser (see Kalnay [92])

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

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[38] 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 representers technique (Bennett[11]), which is a practical algorithm for decoupling the Euler-Lagrange equations associated with the variational problem with weak constraint. (see Amodei[3])

14. 4-D VAR developments in early 1990's

A comprehensive list of adjoint applications to meteorological problem is provided by Courtier[36]. 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[181], Navon et al.[135] and Zupanski[207]. Thepaut et al.[182] used real observations while Rabier and Courtier[154] studied the performance of 4-D VAR in the presence of baroclinic instability. Courtier et al.[37] 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[53] 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[196] and Miller et al.[125] 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.[205]), while at the suggestion and advice of Francois Le Dimet, Zhi Wang completed a doctoral thesis on second order adjoint methods (Wang[191]), as well as a first paper on second order adjoint data assimilation.(Wang et al.)[192] 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 National Aeronautics and Space Administration (NASA) multilevel model.(see Li et al.[109,[110],[111]) Basic work on optimization methods suitable for 4-D VAR was carried out by Zou et al.[206] based on Navon et al.[136],[137]. Application of 4-D VAR to a finite-element model of the shallow-water equations was carried out by Zhu et al.[202] while a novel Hessian preconditioning method based on an idea of Courtier et al.[37] was written by W. Yang et al.[201] Aspects of 4-D VAR dealing with boundary conditions as control variables were dealt amongst others in the work of Zou et al. [204].

15. 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:

$$J(\mathbf{x}) = [H(\mathbf{x}) - y]^T R^{-1}[H(\mathbf{x}) - y] + \tag{109}$$
$$(\mathbf{x}_0 - \mathbf{x}_b)^T B^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \Phi(\mathbf{x})^T C^{-1}\Phi(\mathbf{x})$$

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[116]). One can see that model M verified exactly although it is not perfect.

15.1. 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:

$$\Phi_i(\mathbf{x}) = \mathbf{x}_i - \mathbf{M}_{i-1} \tag{110}$$

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

The cost function becomes:

$$J(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{n}(H(\mathbf{x}_i) - y_i)^T R_i^{-1}(H(\mathbf{x}_i) - y_i) + \tag{111}$$
$$\frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T B^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\sum_{i=1}^{n}\eta_i^T Q_i^{-1}\eta$$

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[18];Dalcher and Kalnay[41]; Bloom and Shubert[17] and Zupanski[207].

For early methods on estimating modeling errors in operational NWP models see Thiébaux and Morone[183] and Saha[162]. 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 $Q = E(\eta(t)\eta^T(t')) = 0, \forall\ t \text{ and } t'$ 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, $\eta \equiv 0$. 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[45] [46]; Dee and Da Silva[50] and Zhu and Kamachi[203]) a fact which led researchers to suggest a variety of assumptions to approximate and parameterize the ME.

Early efforts to model the systematic component of ME were pioneered by Derber[53]. 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[196]; Zupanski[208]) it was shown that inclusion of ME allowed significant reduction in forecast root mean square error (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[75],[77]). Various simple forms of evolution of ME in time were considered by Griffith and Nichols[77],[140], At any time step, t_k, the evolution of ME is

$$\eta_k = T_k(\mathbf{e}_k) + \mathbf{q}_k \tag{112}$$

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, $\mathbf{x}_k$,

$$\mathbf{e}_{k+1} = g_k(\mathbf{x}_k, \mathbf{e}_k). \tag{113}$$

15.2. 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

$$\eta_k = T_k(\mathbf{e}_k) + \mathbf{q}_k \tag{114}$$

$$\mathbf{e}_{k+1} = \mathbf{g}_k(\mathbf{x}_k, \mathbf{e}_k) \tag{115}$$

where T_k describes the distribution of systematic errors in the NWP model equations, and q_k (stochastic component) is an unbiased, serially uncorrelated, normally distributed random vector, with known covariance.

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 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 [77] were provided by neglecting $\mathbf{q}_k$, the stochastic component of ME and using the constant and evolving forms of the systematic component, see [77] for additional details. Similar approaches for modeling the systematic component of ME was considered by Martin et al. (2001) [121] 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. [189]

Other choices can be prescribed, including piecewise constant error and linearly growing error (see Griffith[78], Martin et al.[121]and Griffith et al.[77]. These techniques have been applied successfully in practice to estimate systematic errors in an equatorial ocean model (Martin et al.[121]) Zupanski et al. (2005)[211] provided results obtained used the CSU-RAMS (Colorado State University Regional Atmospheric Modeling System) model in weak constraint VDA framework. Akella and Navon [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.

When the number of observations is considerably smaller, the method of representers (Bennett 2002)[11] provides a computationally efficient (in storage/ space requirements) formulation of VDA. Incorporation of ME in such framework was shown by Uboldi and Kamachi [186].

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) [45] suggested use of a first order (in time) linear model for MEs. That approach was implemented by Zupanski (1997) [208] 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)[51] considered the state-dependent part of ME and proposed a respective estimator.

Mitchell and Daley (1997)[126] considered the discretization part of ME and its effect on data assimilation. Menemenlis and Chechelnitsky [124] 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) [185] structure in meteorology and oceanography are available such as ensemble forecasting, weak-constraint four-dimensional variational assimilation (4D-Var, e.g. Zupanski 1997 [208]; Xu et al. 2005) [199], and Kalman filtering (e.g. Cohn 1997) [33]. 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.

16. 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 can be obtained.

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 to achieve the same goal are available 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 [143]) , later endowed with the adjoint variant ADGEN for reverse AD (1987) by Pin et al.[149]. Later ADIFOR (Bischof et al.[16]) was developed at Argonne National Laboratory , Odyssee was developed at INRIA and

TAMC by Giering and Kaminski [72]. In France the TAPENADE code is used (see Hascoet and Pascual [81]. There are many more languages. Earlier books on AD are by Rall[158] and Kagiwada et al.[90] .

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 [74]). The most comprehensive book and work is that of Andreas Griewank [76] . See also Berz et al.[15] and Griewank and Corliss[74] .

17. 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[173] followed by work of Wang et al.[190,191] stimulated by advice and expertise of F.X. Le Dimet, Wang[368]. Wang et al.[191] and Wang et al.[194] considered use of

second order information for optimization purposes namely to obtain truncated - Newton and Adjoint Newton algorithms using exact Hessian/vector products. Application of these ideas was presented in Wang et al.[193]. Kalnay et al.[169] introduced an elegant and novel pseudo-inverse approach and showed its connection to the adjoint Newton algorithm of Wang et al.[193]. (See Kalnay et al.[91], Pu and Kalnay[153], Park and Kalnay[146], Pu et al.[152]). Ngodock[139] 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.[102] 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. [103] considering all aspects of second order adjoint methods.

18. Computing the second order information

In what follows we follow closely the presentation in Le Dimet et al. (2002)[103].

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

$$F(\mathbf{X}, \mathbf{U}) = 0 \tag{116}$$

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:

$$J(\mathbf{X}, \mathbf{U}) \tag{117}$$

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.

18.1. First order necessary conditions

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

$$\nabla J(\mathbf{U}) = 0 \tag{118}$$

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

$$\frac{\partial F}{\partial \mathbf{X}} \times \hat{\mathbf{X}} + \frac{\partial F}{\partial \mathbf{U}} \times \mathbf{u} = 0 \tag{119}$$

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:

$$\hat{Z}(\mathbf{U},\mathbf{u}) = \lim_{\alpha \to 0} \frac{Z(\mathbf{U}+\alpha\mathbf{u}) - Z(\mathbf{U})}{\alpha} \tag{120}$$

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

$$\hat{Z}(\mathbf{U},\mathbf{u}) = < \nabla Z(\mathbf{U}), \mathbf{u} > \tag{121}$$

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

$$\hat{J}(\mathbf{X},\mathbf{U},\mathbf{u}) = < \frac{\partial J}{\partial \mathbf{X}}, \hat{\mathbf{X}} > + < \frac{\partial J}{\partial \mathbf{U}}, \mathbf{u} > \tag{122}$$

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

$$< \frac{\partial F}{\partial \mathbf{X}} \times \hat{\mathbf{X}}, P > + < \frac{\partial F}{\partial \mathbf{U}} \times \mathbf{u}, P >= 0 \tag{123}$$

$$< (\frac{\partial F}{\partial \mathbf{X}})^T \times P, \hat{\mathbf{X}} > + < (\frac{\partial F}{\partial \mathbf{U}})^T \times P, \mathbf{u} >= 0 \tag{124}$$

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

$$(\frac{\partial F}{\partial \mathbf{X}})^T \times P = \frac{\partial J}{\partial \mathbf{X}} \tag{125}$$

then we obtain

$$\nabla J(\mathbf{U}) = (\frac{\partial F}{\partial \mathbf{U}})^T \times P + \frac{\partial J}{\partial \mathbf{U}} \tag{126}$$

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

18.2. 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

$$(\frac{\partial^2 F}{\partial \mathbf{X}^2} \times \hat{\mathbf{X}} + \frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{U}} \times \mathbf{u})^T \times P + (\frac{\partial F}{\partial \mathbf{X}})^T \times \hat{P} = \tag{127}$$

$$\frac{\partial^2 J}{\partial \mathbf{X}^2} \times \hat{\mathbf{X}} + \frac{\partial^2 J}{\partial \mathbf{X} \partial \mathbf{U}} \times \mathbf{u}$$

and

$$\widehat{\nabla J(\mathbf{U})} = G(\mathbf{U}) \times \mathbf{u} = -(\frac{\partial^2 F}{\partial \mathbf{U}^2} \times \mathbf{u} + \frac{\partial^2 F}{\partial \mathbf{U} \partial \mathbf{X}} \times \hat{\mathbf{X}})^T \times P - \tag{128}$$

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

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

$$< \hat{\mathbf{X}}, (\frac{\partial F}{\partial \mathbf{X}})^T \times Q > + < \mathbf{u}, (\frac{\partial F}{\partial \mathbf{U}})^T \times Q > + < P, (\frac{\partial^2 F}{\partial \mathbf{X}^2}) \times \hat{\mathbf{X}} \times R > \tag{129}$$

$$+ < P, (\frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{U}}) \times \mathbf{u} \times R > + < \hat{P}, (\frac{\partial F}{\partial \mathbf{X}}) \times R >$$

$$=< \hat{\mathbf{X}}, (\frac{\partial^2 J}{\partial \mathbf{X}^2})^T \times R > + < \mathbf{u}, (\frac{\partial^2 J}{\partial \mathbf{X} \partial \mathbf{U}})^T \times R >$$

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

$$< G(\mathbf{U}) \times \mathbf{u}, \mathbf{u} >=< -(\frac{\partial^2 F}{\partial \mathbf{U}^2} \times \mathbf{u} + \frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{U}} \times \hat{\mathbf{X}})^T \times P, \mathbf{u} > + \tag{130}$$

$$< \hat{P}, (-\frac{\partial F}{\partial \mathbf{U}}) \times \mathbf{u} >< \frac{\partial^2 J}{\partial \mathbf{U}^2} \times \mathbf{u}, \mathbf{u} > + < \hat{\mathbf{X}}, \frac{\partial^2 J}{\partial \mathbf{X} \partial \mathbf{U}})^T \times \mathbf{u} >$$

From (173) we get

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

$$< \mathbf{u}, -(\frac{\partial F}{\partial \mathbf{U}})^T \times Q - (\frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{U}} \times P)^T \times R + \frac{\partial^2 J}{\partial \mathbf{X} \partial \mathbf{U}} \times R > \tag{131}$$

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

$$(\frac{\partial F}{\partial \mathbf{X}})^T < \mathbf{u}, -(\frac{\partial F}{\partial \mathbf{U}})^T \times Q + (\frac{\partial^2 F}{\partial \mathbf{X}^2} < \mathbf{u}, -(\frac{\partial F}{\partial \mathbf{U}})^T \times P) \times R - \tag{132}$$

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

$$(\frac{\partial F}{\partial \mathbf{X}}) \times R = -\frac{\partial F}{\partial \mathbf{U}} \times \mathbf{u} \tag{133}$$

then we obtain:

$$G(\mathbf{U})\times\mathbf{u} = -(\frac{\partial^2 F}{\partial \mathbf{U}^2}\times\mathbf{u})\times P + \frac{\partial^2 J}{\partial \mathbf{U}^2}\times\mathbf{u} - (\frac{\partial F}{\partial \mathbf{U}})^T \times Q - \tag{134}$$

$$(\frac{\partial^2 F}{\partial \mathbf{X}\partial \mathbf{U}}\times P)\times R + \frac{\partial^2 J}{\partial \mathbf{X}\partial \mathbf{U}}\times R$$

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

$$\frac{\partial^2 F}{\partial \mathbf{X}^2} = (\frac{\partial^2 F}{\partial \mathbf{X}^2})^T \tag{135}$$

leading to some simplifications. The system (131-134) will be called the second order adjoint. Therefore we can obtain the product of the Hessian by a vector **u** by (i) solving the system (131-134). (ii) applying formula (134).

18.3. Remarks

a) The system (131-134) which has to be solved to obtain the Hessian/vector product can be derived from the Gateaux derivative (131) which is the same as (134). In the literature, the system (131-134) 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 131 differs from the adjoint model by the forcing terms which will depend on **u** and R.

c) The system (131-134) will yield the exact value of the Hessian/vector product. An approximation could be obtained by using the standard finite differences, i.e.,

$$G(\mathbf{U})\times\mathbf{u} \approx \frac{1}{\alpha}[\nabla J(\mathbf{U}+\alpha\mathbf{u}) - \nabla J(\mathbf{U})] \tag{136}$$

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 et al. [192] and references therein).

18.4. 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 $\mathbf{X} \in H[C(0,T)]^n$ between 0 and T is governed by the differential system,

$$\frac{\partial \mathbf{X}}{\partial t} = F(\mathbf{X}) + \mathbf{B} \times \mathbf{V} \tag{137}$$

The input variable is often the initial condition,

$$\mathbf{X}(0) = \mathbf{U} \in R^n \tag{138}$$

In this system F is a nonlinear operator which describes the dynamics of the model, $\mathbf{V} \in V[C(0,T)]^m$ 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 (137)-(138) and observations

$$J(\mathbf{U},\mathbf{V}) = \frac{1}{2}\int_0^T || \mathbf{HX} - \mathbf{X}_{obs} ||^2 \, dt \tag{139}$$

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

$$\frac{d\hat{\mathbf{X}}}{dt} = \frac{\partial F}{\partial \mathbf{X}} \times \hat{\mathbf{X}} + \mathbf{B} \times \mathbf{V} \tag{140}$$

$$\hat{\mathbf{X}}(0) = \mathbf{u} \tag{141}$$

$$\hat{J}(\mathbf{U},\mathbf{V},\mathbf{u},\mathbf{v}) = \frac{1}{2}\int_0^T < \mathbf{HX} - \mathbf{X}_{obs}, \mathbf{H}\hat{\mathbf{X}} > dt \tag{142}$$

Let us introduce P the adjoint variable, we take the product of (140) 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 (142), we conclude that of P is defined as the solution of

$$\frac{dP}{dt} = \frac{\partial F}{\partial \mathbf{X}}^T \times P + \mathbf{H}^T \mathbf{H}(\mathbf{X} - \mathbf{X}_{obs}) \tag{143}$$

$$P(T) = 0 \tag{144}$$

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

$$\nabla J_{\mathbf{U}} = -P(0) \tag{145}$$

$$\nabla J_{\mathbf{V}} = -\mathbf{B}^T P \tag{146}$$

$\mathbf{V}$ 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 $\mathbf{V}$

$$h = \begin{pmatrix} h_U \\ h_V \end{pmatrix} \tag{147}$$

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

$$\frac{d\hat{\mathbf{X}}}{dt} = \frac{\partial F}{\partial \mathbf{X}} \hat{\mathbf{X}} + \mathbf{B} h_V \tag{148}$$

$$\hat{\mathbf{X}}(0) = h_U \tag{149}$$

$$\frac{d\hat{\mathbf{P}}}{dt} + (\frac{\partial^2 F}{\mathbf{X}^2} \times \hat{\mathbf{X}})^T \times P + (\frac{\partial F}{\partial \mathbf{X}})^T \times P = \mathbf{H}^T \mathbf{H} \hat{\mathbf{X}} \tag{150}$$

$$\nabla J_{\mathbf{U}} = -\hat{P}(0) \tag{151}$$

$$\nabla J_{\mathbf{V}} = -\mathbf{B}^T \hat{P} \tag{152}$$

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

Taking the inner product of (148) with Q and of (150) with R, integrating from 0 to T, then adding the resulting equations, we may write:

$$\int_0^T [< \frac{d\hat{\mathbf{X}}}{dt}, Q > - < \frac{\partial F}{\partial \mathbf{X}} \times \hat{\mathbf{X}}, Q > - < \mathbf{B} h_V, Q > + < \frac{d\hat{P}}{dt}, R > + \tag{153}$$

$$< [\frac{\partial^2 F}{\partial \mathbf{X}^2} \times \hat{\mathbf{X}}]^T \times P, R > + < [\frac{\partial F}{\partial \mathbf{X}}]^T \times P, R > - < \mathbf{H}^T \mathbf{H} \hat{\mathbf{X}}, R >] dt = 0$$

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

$$\int_0^T < \hat{\mathbf{X}}, -\frac{dQ}{dt} - [\frac{\partial F}{\partial \mathbf{X}}]^T \times Q + [\frac{\partial^2 F}{\partial \mathbf{X}^2} \times P]^T \times R - \mathbf{H}^T \mathbf{H} R > dt + \tag{154}$$

$$\int_0^T < \hat{P}, -\frac{dR}{dt} + (\frac{\partial F}{\partial \mathbf{X}}) \times R > dt - \int_0^T < h_V, \mathbf{B}^T \times Q > dt +$$

$$< \hat{\mathbf{X}}(T), Q(T) > - < \hat{\mathbf{X}}(0), Q(0) > + < \hat{P}(T), R(T) > -$$

$$< \hat{P}(0), R(0) >= 0$$

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

$$\mathbf{G} = \begin{pmatrix} G_{UU} & G_{UV} \\ G_{VU} & G_{VV} \end{pmatrix} \tag{155}$$

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

$$\frac{dQ}{dt} + [\frac{\partial F}{\partial \mathbf{X}}]^T \times Q = [\frac{\partial^2 F}{\partial \mathbf{X}^2} P]^T \times R - \mathbf{H}^T \mathbf{H} R \tag{156}$$

$$\frac{dR}{dt} = [\frac{\partial F}{\partial \mathbf{X}}] \times R \tag{157}$$

and

$$Q(T) = 0 \tag{158}$$

$$R(0) = h_U \tag{159}$$

then we finally obtain

$$< -h_U, Q(0) > = < \hat{P}(0), R(0) > \tag{160}$$

$$\hat{P}(0) = -Q(0) \tag{161}$$

We would like to point out that Eq. (161) follows directly from Eq. (160) by using Eq. (159). The product of the Hessian by a vector r is obtained exactly by a direct integration of (157) and (159) followed by a backward integration in time of (156) and (158).

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

$$\frac{dQ}{dt} + [\frac{\partial F}{\partial \mathbf{X}}]^T \times Q = [\frac{\partial^2 F}{\partial \mathbf{X}^2} \times P]^T \times R - \mathbf{H}^T \mathbf{H} R \tag{162}$$

$$\frac{dR}{dt} = [\frac{\partial F}{\partial \mathbf{X}}] R \tag{163}$$

with the conditions

$$Q(T) = 0 \tag{164}$$

$$R(0) = \mathbf{e}_i \tag{165}$$

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

$$\mathbf{G}_{UU} \mathbf{e}_i = Q(0) \tag{166}$$

$$\mathbf{G}_{UV}\mathbf{e}_i = \mathbf{B}^T \times Q \tag{167}$$

One then integrates m times the differential system

$$\frac{dQ}{dt} + [\frac{\partial F}{\partial \mathbf{X}}]^T \times Q = [\frac{\partial^2 F}{\partial \mathbf{X}^2} \times P]^T \times R - \mathbf{H}^T \mathbf{H} R \tag{168}$$

$$\frac{dR}{dt} - [\frac{\partial F}{\partial \mathbf{X}}] \times R = \mathbf{f}_j \tag{169}$$

with initial and terminal conditions

$$Q(T) = 0 \tag{170}$$

$$R(0) = 0 \tag{171}$$

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

$$\mathbf{G}_{VV} \times \mathbf{f}_j = \mathbf{B}^T \times Q \tag{172}$$

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.

18.5. 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[173], Rabier and Courtier[154], Yang et al.[201], Le Dimet et al.[102]).

Following Courtier et al.[37] 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 and Courtier [67] we consider the observations as random variables and

we look at variational analysis as attempting to solve the minimization problem

$$\min J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(\mathbf{Hx}-\mathbf{y})^T \mathbf{O}^{-1}(\mathbf{Hx}-\mathbf{y}) \tag{173}$$

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

$$J'' = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{O}^{-1} \mathbf{H} \tag{174}$$

obtained by differentiating (173) twice.

Moreover the analysis error covariance matrix is the inverse of the Hessian as shown in Appendix B of Rabier and Courtier[154]. Calling $\mathbf{x_a}$ the result of the minimization (i.e. the analysis) and $\mathbf{x_t}$ the truth, one has

$$\mathbf{E}[(x_a - x_t)(x_a - x_t)^T] = (J'')^{-1} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{O}^{-1} \mathbf{H})^{-1} \tag{175}$$

A requirement is that the background error and the observation error are uncorrelated (Rabier and Courtier[154], Fisher and Courtier[59]). See also work of Thepaut and Moll[180] pointing out that the diagonal of the Hessian is optimal among all diagonal preconditioners.

19. 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.[7,8]. 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 [36] and tested by Rabier et al.[155] for the development of a simplified Kalman filter fully described by Fisher[60] and compared with a low resolution explicit extended Kalman filter by Ehrendorfer and Bouttier[57].

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

$$\frac{<\mathbf{P}x(t), \mathbf{EP}x(t)>}{<x(t_0), \mathbf{C}x(t_0)>} = \lambda^2 \tag{176}$$

under an Euclidean norm are solution of generalized eigenvalue problem.

$$\mathbf{M}^*\mathbf{P}^*\mathbf{EPMv} = \lambda^2\mathbf{Cv} \tag{177}$$

E is the total energy metric, $\mathbf{v}$ are the Hessian singular vectors and λ^2 the Hessian eigenvalues.

In HSV, the operator $\mathbf{C}$ is equal to the Hessian of the 3-D Var cost function. As suggested by Barkmeijer et al.[7], one can solve (177) by using the generalized eigenvalue algorithm (Davidson [48]). See also Sleijpen and Van der Vorst[172]. Using

$$\mathbf{C} \equiv \nabla^2 J = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{O}^{-1}\mathbf{H} \tag{178}$$

and carrying out a coordinate transformation

$$\mathbf{x} = \mathbf{L}^{-1}\mathbf{x}, \mathbf{L}^{-1}\mathbf{L} = \mathbf{B} \tag{179}$$

Then we obtain a transformed operator

$$(\mathbf{L}^{-1})^T\mathbf{CL} \tag{180}$$

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 [59]).

Veerse[188] 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 Broyden-Fletcher-Goldfarb-Shanno (BFGS) minimization algorithm.

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

$$\mathbf{s} = \mathbf{x}^{n+1} - \mathbf{x}^n \tag{181}$$

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

$$\mathbf{y} = \mathbf{g}^{n+1} - \mathbf{g}^{n} \tag{182}$$

is the corresponding gradient increment. One has the formula

$$\mathbf{H}^{+} = \mathbf{U}(\mathbf{H}, \mathbf{y}, \mathbf{s}) = (\mathbf{I} - \frac{\mathbf{s} \otimes \mathbf{y}}{< \mathbf{y}, \mathbf{s} >}) \frac{\mathbf{s} \otimes \mathbf{s}}{< \mathbf{y}, \mathbf{s} >} \tag{183}$$

where $<,>$ is a scalar product with respect to which the gradient is defined and $\otimes$ 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[141] algorithm. These methods are useful when the second order adjoint method is not available due to either memory or CPU limitations.

20. 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.[156), Klinker et al. [93], while a similar system was operational at Meteo-France in 2000 (Janiskova et al.[88], Gauthier and Thépaut[68], Desroziers et al.[56]). More recently 4-D VAR was implemented at UK Met office, Japan and Canada.

Park and Zupanski [144] survey the status and progress of the four-dimensional variational data assimilation with emphasis on application to prediction of mesoscale/storm-scale atmospheric phenomena.

The impact of adopting 4-D VAR was qualified as a substantial, resulting in an improvement in NWP quality and accuracy (see Rabier[157] in a special issue of Quarterly Journal of the Royal Meteorological Society (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[171]).

Hollingsworth et al. [86] shows how observing system improvements led to improvements of forecast scores while Bouttier and Kelly[20] 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.

21. Recent algorithmic developments of note for 4-D VAR

Following an idea of Derber, Courtier et al. [37] 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[157])

$$J(\delta w_0) = \frac{1}{2}\delta w_0{}^T B^{-1}\delta w_0 + \frac{1}{2}\sum_{I=1}^{N}(H_i\delta_{X_i} - d_i)^T R_i^{-1}(H_i\delta_{X_i} - d_i) \tag{184}$$

Where $\delta w_0 = \mathbf{S}(X_0 - X_b)$ is the simplified increment at initial time t_0 and

$$d_i = y_i^o - H_i(X_i) \tag{185}$$

is the observation increment at time t_i. **S** is an operator going from high to low resolution. The solution resulting from minimization of the cost function is added to the background X_b to obtain analysis at t_0 i.e

$$X_0^a = X^b - S^{-I}\delta_{w_0}^a \tag{186}$$

where S^{-I} is the generalized inverse of operator S transforming the field from low to high resolution (i.e. S projects from high to low 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[187].

Physical parameterizations that have been modified to allow use in the linear models used in the incremental procedure were implemented by Janiskova et al. [89], Lopez and Moreau[105].

22. 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[161], Fisher[62] and Cardinali et al.[27].

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

$$x_a = x_b + (B^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} (y^o - H(x_b)) \tag{187}$$

which can be written compactly as

$$x_a = x_b + Kd \tag{188}$$

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

The DFS is defined as

$$DFS = Tr(HK) \tag{189}$$

where the trace of the matrix HK measures the gain in information due to the observations of how an assimilation system extracts information signal from the background. One way to calculate DFS is the use of estimation the Hessian of the cost function provided. Fisher[62] and Cardinali et al.[27] used estimation of Hessian of the cost function provided by the minimization algorithm. Chapnik et al.[30] use evaluation of trace of the KH matrix, using a method put forward by Desroziers and Ivanov[55] 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 which is a topic of increased research efforts in 4-D VAR data assimilation (Berliner et al [13]), Baker and Daley[4]) Daescu and Navon[40], Langland and Baker [95] .

23 Conclusions

A condensed review of several aspects of 4-D VAR as it evolved in the last 30 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 here the vast ensemble Kalman filter data assimilation and its various flavors 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 and sequential data assimilation can take advantage of them due to major research efforts at universities, federal laboratories and operational centers.
For new opportunities for research see the article by McLaughlin et al. [212] that illuminates and outlines possibilities for enhanced collaboration within the data assimilation community .
It is certain that data assimilation concepts will become widely applied in all the geosciences 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 real time PDE constrained optimization research scientists.

References

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

[2] S. Akella, I.M. Navon,(2007) Different approaches to model error formulation in 4D-Var:a study with high resolution advection schemes. In revision process with Tellus.

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

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

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

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

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

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

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

[10] R.E. Bellman, (1957.) Dynamic Programming. Princeton University Press,

Princeton.
[11] A.F. Bennett (2002) ,Inverse Modeling of the Ocean and Atmosphere. Cambridge University Press. pp 256.
[12] P. Bergthorsson, B. Döös , (1955) Numerical Weather Map Analysis, Tellus 7(3) 329-340.
[13] Berliner, L.M., Lu, Z.Q., and C. Snyder, (1999), Statistical design for adaptive weather observations. J. Atmos. Sci., 56,2536-2552.
[14] D.P. Bertsekas,(1982) Constrained optimization and Lagrange multiplier methods. Addison-Wesley, pp 491.
[15] M. Berz, C.H. Bischof, G. Corliss, and A. Griewank (eds.), (1996) Computational differentiation - techniques, applications, and tools, SIAM, Philadelphia, Penn.,
[16] C.H. Bischof, G.F. Corliss, L.Green, A. Griewank, K. Haigler, (1992) Automatic differentiation of advanced CFD codes for multidisciplinary design, Journal on Computing Systems in Engineering 3, 625-638.
[17] S.C. Bloom, S. Shubert, (1990)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,
[18] G.J. Boer, (1984) A spectral analysis of predictability and error in an operational forecast system. Mon. Weather Rev. 112 1183-1197
[19] F. Bouttier, P. Courtier, (1999) Assimilation concepts and methods, Meteorological Training Lecture Notes, ECMWF. Shinfield Park, Reading,
[20] F. Bouttier, Kelly, G. (2001) Observing-system experiments in the ECMWF 4D-Var data assimilation system. Q. J. R. Meteorol. Soc., 127, 1469-1488.
[21] A. Bratseth, (1986) Statistical interpolation by means of successive corrections , Tellus 38A (1986) pp 439-447.
[22] Buizza, R., Palmer, T. N., ,(1995) The singular-vector structure of the atmospheric general circulation. J. Atmos. Sci., 52, 91434-1456.
[23] D.G. Cacuci, C.F. Weber, E.M. Oblow, J.H. Marable, (1980). Sensitivity Theory for General Systems of Nonlinear Equations, Nucl. Sci. Eng. 75(88)
[24] D.G. Cacuci, (1981a) Sensitivity theory for nonlinear systems. I: Nonlinear functional analysis approach. J. Math. Phy. 12 ,2794-2803.
[25] D.G. Cacuci, (1981b) Sensitivity theory for nonlinear systems. II. Extensions to additional classes of responses. J. Math. Phy. 12, 2803-2812.
[26] D.G. Cacuci, (2004) 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.
[27] Cardinali, C., Pezzuli, S. and Andersson, E. , (2004) Influence matrix diagnostic of a data assimilation system. Q. J. R. Meteorol. Soc., 130, 2767-2786.
[28] D.G. Cacuci (2003), Sensitivity & Uncertainty Analysis: A Primer, Vol. 1CRC Press,304pp
[29] D. G. Cacuci, M. Ionescu-Bujor, I. M. Navon, .(2005) Sensitivity and Uncertainty Analysis, Volume II: Applications to Large-Scale Systems. CRC Press,368pp
[30] Chapnik, B., Desroziers, G., Rabier, F. and Talagrand, O. (2006), Diagnosis and tuning of observational error statistics in a quasi-operational data assimilation setting. Q. J. R. Meteorol. Soc., 132, 543-565.
[31] J.G. Charney, R. Fjörtoft, John von Neumann, (1950) Numerical Integration of the Barotropic Vorticity Equation, Tellus 2 237-254.

[32] J.G. Charney, M. Halem, R. Jastrow, (1969) Use of incomplete historical data to infer the present state of the atmosphere. J. Atmos. Sci. 2 1160-1163.
[33] S.E. Cohn, (1997) An introduction to estimation theory. J. Meteorol. Soc. Jpn. 75 257-288
[34] P. Courtier, (1985.) Experiments in data assimilation using the adjoint model technique. Workshop on High-Resolution Analysis ECMWF (UK) June
[35] P.Courtier, O.Talagrand, (1987) Variational assimilation of meteorological observations with the adjoint equations Part 2. Numerical results. Q. J. R. Meteorol. Soc. 113, 1329-1347.
[36] P. Courtier, P. Unden, W. Heckley, D. Vasijevic, E. Andersson, J.N. Thepaut, F. Rabier, R. Hoffman, (1993). Progress in variational data assimilation at ECMWF. Research activities in atmospheric and oceanic modeling, G. J. Boer (ed.), Report No. 18 1.42
[37] P. Courtier, Thepaut JN, A. Hollingsworth, (1994) A Strategy for Operational implementation of 4D-VAR, using an incremental approach. Quarterly Journal of the Royal Meteorological Society 120 (519): 1367-1387 Part B JUL
[38] P. Courtier, (1997) Dual formulation of four-dimensional variational assimilation . Quarterly Journal of the Royal meteorological Society 123 (544), 2449-2461 , Part B .
[39] G.P. Cressman, (1959) An operational objective analysis system. Mon. Wea Rev. 87 367-374.
[40] Daescu, D.N., and I.M. Navon, (2004) Adaptive observations in the context of 4D-Var data assimilation. Meteorol. Atmos. Phys., 85, 205-226.
[41] A. Dalcher, E.Kalnay, (1987) Error growth and predictability in operational ECMWF forecasts. Tellus 39 A 474-491 .
[42] R. Daley, K. Puri, (1980) Four-Dimensional, Data Assimilation and the Slow Manifold. Mon. Wea. Rev. 108 85-99.
[43] R. Daley, (1991)Atmospheric Data Analysis, Cambridge University Press,
[44] R. Daley, (1992) The effect of serially correlated observation and model error on atmospheric data assimilation. Mon. Weather Rev. 120 ,164-177.
[45] R. Daley, (1992) The effect of serially correlated observation and model error on atmospheric data assimilation. Mon. Weather Rev. 120, 164-177.
[46] R. Daley, (1992) Estimating model-error covariances for application to atmospheric data assimilation. Mon. Weather Rev. 120 1735-1746
[47] A. Da Silva, J. Pfaendtner, J. Guo, M. Sienkiewicz, S. Cohn,(1995) 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,
[48] E.R. Davidson, (1975) The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of a large real symmetric matrices. J. Comput. Phys. 17 87-94.
[49] H.E. Davis, R.E. Turner, (1977) Updating prediction models by dynamic relaxation. An examination of the technique. Q. J. R. Meteorol. Soc. 103 225-245
[50] D. Dee, A. Da Silva, (1998) Data assimilation in the presence of forecast bias. Q. J. R. Meteorol. Soc. 124 ,269-295
[51] T. DelSole, A.Y. Hou, (1999) Empirical correction of a dynamical model. Part I: fundamental issues. Mon. Weather Rev. 127, 2533-2545 .
[52] J.C. Derber,(1985) The variational four dimensional assimilation of analysis using filtered models as constraints. Ph.D. Thesis, Univ. of Wisconsin-Madison, pp 141.

[53] J.C. Derber, (1989) A variational continuos assimilation technique. Mon. Wea. Rev. 117, 2437-2446.
[54] J.C. Derber, W.-S. Wu, (1998) The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. Mon. Wea. Rev. 126, 2287-2299.
[55] Desroziers, G. and Ivanov, S. , (2001) Diagnosis and adaptive tuning of information error parameters in a variational assimilation. Q. J. R. Meteorol. Soc., 127, 1433-1452.
[56] Desroziers, G., Hello, G. and Thepaut, J.-N. ,(2003) A 4D-Var re-analysis of the FASTEX. Q. J. R. Meteorol. Soc., 129,1301-1315.
[57] M. Ehrendorfer, F. Bouttier, (1998) An explicit low-resolution extended Kalman filter: Implementation and preliminary experimentation. Sept. ECMWF Technical Memorandum No 259.
[58] A. Eliassen,(1954) 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, pp 11.
[59] M. Fisher, P. Courtier, (1995) Estimating the covariance matrices of analysis and forecast errors in variational data assimilation. ECMWF Tech Memorandum, 220 pp28.
[60] M. Fisher, (1998) Development of a simplified Kalman filter. ECMWF Tech Memorandum 260 pp 16.
[61] M. Fisher, (2003)Background Error Covariance Modelling M Fisher seminar ,Recent developments in data assimilation for atmosphere and ocean, 8-12 September European Centre for Medium-Range Weather Forecasts .
[62] Fisher, M. (2003) Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. Technical Memo. 397, ECMWF, Reading, UK.
[63] R. Franke, W.J. Gordon,(1983) The structure of optimal interpolation functions. Technical Report, NPS-53-83-0004 , Naval Postgraduate School, Monterey, Ca.
[64] R. Franke, (1988) Statistical interpolation by iteration. Mon. Wea. Rev. 116 961-963.
[65] L.S. Gandin,(1965) Objective analysis of meteorological fields. Leningrad: Gridromet(in Russian) (1963), English translation by the Israel program for scientific translations, pp 242.
[66] C.F. Gauss, (1809), Theoria motus corporum coelestium in sectionibus conicis solem ambientum Book 2, Section 2, pp. 205-224 English translation by Ch. Davis, reprinted 1963 by Dover, New York.142pp.
[67] P. Gauthier, P. Courtier,(1992) 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, , 171-190.
[68] Gauthier, P. and Thepaut, J.-N. , (2001) Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Meteo-France. Mon. Weather Rev., 129, 2089-2102.
[69] M. Ghil, M. Halem, R. Atlas, (l979) Time-continuous assimilation of remote-sounding data and its effect on weather forecasting, Mon. Wea. Rev. 107, l40-l7l.
[70] M. Ghil, P. Malanotte-Rizzoli, (1991) Data assimilation in meteorology and oceanography. Adv. Geophys. 33 141-266.
[71] M. Ghil, K. Ide, A. F. Bennett, P. Courtier, M. Kimoto, N. Sato (Eds.), (1997) Data Assimilation in Meteorology and Oceanography: Theory and Practice, Meteorological Society of Japan and Universal Academy Press, Tokyo, 496.

[72] R. Giering and T. Kaminski,(1997) Recipes for adjoint code construction, Technical Report 212, Max-Planck-Institut für Meteorologie.
[73] B. Gilchrist, G.P. Cressman, (1954) An experiment in objective analysis. Tellus 6, 309-318.
[74] A. Griewank, F.Corliss. George, (1991) Automatic differentiation of algorithms: theory, implementation, and application, SIAM, Philadelphia, pp 353.
[75] A.K. Griffith, N.K. Nichols, (1996) 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.
[76] A. Griewank, (2000) Evaluating Derivatives Principles and Techniques of Algorithmic Differentiation. Frontiers in Applied Mathematics 19, 369 pp
[77] A.K. Griffith, N.K. Nichols, (2000) Adjoint methods in data assimilation for estimating model error. Flow, Turbulence and Combustion 65 469-488
[78] A.K. Griffith, (1997) Data Assimilation for Numerical Weather Prediction Using Control Theory, The University of Reading, Department of Mathematics, PhD Thesis,.
[79] M.D. Gunzburger, (2003) Perspectives in Flow Control and Optimization. Advances in design and control. SIAM, Philadelphia, USA,.
[80] G.J. Haltiner, R. T. Williams, (1980)Numerical prediction and dynamic meteorology. John Wiley and Sons, New York, , pp 477.
[81] L. Hascoet, V. Pascual. (2004) TAPENADE 2.1 user's guide, Technical report, INRIA, , no 300.81pp. .
[82] M.R. Hestenes, (1969) Multiplier and gradient methods. J. Opt. Theory Appl. 4 303-320.
[83] R.N. Hoffmann, (1986) A four dimensional analysis exactly satisfying equations of motion. Mon. Wea. Rev. 114, 388-397.
[84] J.E. Hoke, R.A. Anthes, (1976) The initialization of numerical models by a dynamical initialization technique. Mon. Wea. Rev. 194 1551-1556.
[85] A. Hollingsworth, P. Lonnberg, (1986) The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: the wind field. Tellus 38A 111-136.
[86] Hollingsworth, A., Uppala, S., Klinker, E., Onvlee, J., de Vries, J. W., Burridge, D., Vitart, F., de Roo, A. and Pfrang, C. (2005) The transformation of earth-system observations into information of socio-economic value in GEOSS. Q. J. R. Meteorol. Soc., 131, 3493-3512.
[87] K. Ide, P. Courtier, M. Ghil, A. Lorenc, (1997) Unified notation for data assimilation: Operational, sequential and variational. J. Meteor. Soc. Japan 75 181-189.
[88] Janiskova, M., Thepaut, J.-N. and Geleyn, J.-F. ,(1999) Simplified and regular physical parameterizations for incremental four-dimensional variational assimilation. Mon. Weather Rev., 127, 26-45.
[89] Janiskova, M., Mahfouf, J.-F., Morcrette, J.-J. and Chevallier, F. (2002) Linearized radiation and cloud schemes in the ECMWF model: Development and evaluation. Q. J. R. Meteorol. Soc., 128, 1505-1527.
[90] H. Kagiwada, R. Kalaba, N. Rasakhoo, and K. Spingarn, (1986) Numerical derivatives and nonlinear analysis, Mathematical Concepts and Methods in Science and Engineering, Plenum Press, New York – London.
[91] E. Kalnay, S.K. Park, Z-X. Pu and J Gao, (2000) Application of the quasi-inverse method to data assimilation. Mon. Wea. Rev. 128, 864-875.

[92] E. Kalnay, (2003) Atmospheric modeling, data assimilation and predictability. Cambridge Univ. Press, pp 341.
[93] Klinker, E., Rabier, F., Kelly, G. and Mahfouf, J.-F., .,(2000) The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration. Q. J. R. Meteorol. Soc, 126, 1191-1215.
[94] D.G. Krige, (1951) A statistical approach to some mine valuation problems on the Witwatersrand. J. Chem. Metal. Min. Soc. South African 52 119-139.
[95] Langland, R.H., and N.L. Baker, (2004) Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. Tellus, 56A, 189-203.
[96] J.L. Lagrange, (1760a), Essai d'une nouvelle methode pour determiner les maxima et les minima des formules inde_nies, Miscellanea Taurinensia, 2 (1762) Oeuvres, 1,365-468.
[97] C. Lanczos, (1970) The Variational Principles of Mechanics. Toronto: University of Toronto Press.
[98] F.X. Le-Dimet, (1981) Une application des méthodes de contrôle, optimal ‘a l’analyse variationnelle," Rapport Scientifique LAMP, Université Blaise-Pascal, 63170, Aubière Cedex, 42 pages.
[99] Le Dimet, F. X., (1982) A general formalism of variational analysis. CIMMS Report, Norman, OK 73091, 22, 1–34 pp.
[100] Le-Dimet, F. X., Y. K. Sasaki, and L. White, (1982): Dynamic initialization with filtering of gravity waves, CIMMS Report No. 40 (1982), Norman, OK 73019, 12 pages. .
[101] F.X. Le-Dimet, O. Talagrand, (1986) Variational algorithms for analysis and assimilation of meteorological observations. Tellus 38A, 97-110.
[102] F.X. Le Dimet, H.E. Ngodock , B. Luong, J. Verron, (1997), Sensitivity analysis in variational data assimilation. J Meteorol Soc Japan, (1B) 245-255.
[103] Francois X. Le-Dimet, I. M. Navon and Dacian N. Daescu , (2002) Second Order Information in Data Assimilation Monthly Weather Review 130 (3) 629-648.
[104] F.X. Le Dimet, I.M. Navon, (1988)Variational and optimization methods in meteorology: a review. Tech. Rep. FSU-SCRI-88-144, Florida State University, Tallahassee, Florida, pp 83.
[105] Lopez, P. and Moreau, E., (2005) A convection scheme for data assimilation: Description and initial tests. Q. J. R. Meteorol. Soc., 131,409-436.
[106] A.N. Kolmogorov, (1941). Interpolation extrapolation of stationary random sequences. Bulletin of the Academy of Sciences, USSR series on Mathematics, 15 (Translation by RAND Corporation, memorandum RM-30090-PR, April (1962).
[107] J.M. Lewis, J.C. Derber, (1985) The use of adjoint equations to solve a variational adjustment problem with advective constraints. Tellus 37A , 309-322.
[108] J.M. Lewis, S. Lakshmivarahan, Sudarshan Dhall , (2006) Dynamic data assimilation, a least squares approach. Cambridge University Press .
[109] Y. Li, I.M. Navon, P. Courtier, P. Gauthier, (1993) Variational data assimilation with a semi-implicit semi-Lagrangian global shallow-water equation model and its adjoint. Mon. Wea. Rev. 121 , 159-169.
[110] Y. Li, K. K. Droegemeier, (1993)The influence of diffusion and associated errors on the adjoint data assimilation technique. Tellus 45A , 1-14.
[111] Y. Li, I.M. Navon, W. Y. Yang, X. L. Zou, J. R. Bates, S. Moorthi, R. W. Higgins, (1994) 4-D assimilation experiments with a multilevel semi-Lagrangian semi-implicit GCM. Mon. Wea. Rev.122(5), 966-983.

[112] J.L. Lions, (1968) Contrôle Optimal des Systèmes Gouvernes par des Equations aux Dérivés Partielles. Dunod et Gauthiers-Villars, Paris,
[113] J.L. Lions, (1971) Optimal control of systems governed by partial differential equations. Translated by S. K. Mitter, Springer-Verlag, Berlin-Heidelberg, , pp 404.
[114] A.C. Lorenc, (1981) A global three-dimensional multivariate statistical interpolation scheme. Mon. Wea. Rev. 109 701-721.
[115] A.C. Lorenc, (1986) Analysis-methods for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society 112 (474) 1177-1194.
[116] P. Lynch, X.Y. Huang, (1992) Initialization of the HIRLAM Model Using a Digital Filter. Mon Wea Rev. 120(6) 1019-1034.
[117] G.I. Marchuk, (1958) Numerical Calculation of Nuclear Reactors. Moskwa, Atomizdat, , 162 pp.
[118] G.I. Marchuk, (1967) Chislennye Metody v Prognoze Pogodi, (Numerical Methods in Weather Forecasting), Gidrometeoizdat, Leningrad, , 356 pp.
[119] G.I. Marchuk, (1974) Numerical solution of the problem of dynamics of atmosphere and ocean (in Russian). Leningrad, Gidrometeoizdat, , pp 303.
[120] M.J. Martin, N.K. Nichols, M.J. Bell,(1999) treatment of Systematic Errors in Sequential Data Assimilation, The Met Office, Ocean Applications Division, Tech. Note No. 21,
[121] M.J. Martin, M.J. Bell, N.K. Nichols,(2001) 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, , pp. 423-430.
[122] T. Matsuno, (1966) Numerical integrations of the primitive equations by simulated backward difference scheme. J. Meteor. Soc. Japan 44 76-84.
[123] R.D. McPherson, (1975) Progress, problems, and prospects in meteorological data assimilation. Bulletin of the American Meteorological Society 56 (11) 1154-1166.
[124] D. Menemenlis, M. Chechelnitsky, (2000) Error estimates for an ocean general circulation model from altimeter and acoustic tomography data. Mon. Weather Rev. 128 763-778.
[125] R.N. Miller, M Ghil, Gauthiez F (1994)Advanced data assimilation in strongly nonlinear dynamical-systems. Journal of the Atmospheric Sciences 51 (8)(1994) 1037-1056.
[126] H.L. Mitchell, R. Daley, (1997) Discretization error and signal/error correlation in atmospheric data assimilation. I: All scales resolved. Tellus 49A 32-53 .
[127] K. Miyakoda, L. Umscheid, D. H. Lee, J. Sirutis, R. Lusen, F. Pratte, (1976) The near-real-time, global, four-dimensional analysis experiment during the GATE period, Part I. Journal of the Atmospheric Sciences, 33 (4), 561-591.
[128] D. McLaughlin, O'Neill A, Derber J, M.Kamachi , (2005) Opportunities for enhanced collaboration within the data assimilation community Quarterly Journal of the Royal Meteorological Society 131 (613), 3683-3693, Part C.
[129] I.M. Navon, (1981) Implementation of A Posteriori Methods for Enforcing Conservation of Potential Enstrophy and Mass in Discretized Shallow-Water Equations Models. Mon. Wea. Rev. 109 946-958.
[130] I.M. Navon, R. de Villiers, (1983) Combined penalty multiplier optimization methods to enforce integral invariants conservation. Mon. Wea. Rev. 111, 1228-1243.
[131] I.M. Navon, (1986) 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 29-34.

[132] S.G. Nash, A. Sofer, (1996)Linear and Nonlinear Programming. McGraw-Hill, , pp 692.
[133] I.M. Navon, (1987) The Bayliss-Isaacson algorithm and the constraint restoration method are equivalent. Meteorology and Atmospheric Physics 13 143-152.
[134] I.M. Navon, X. Zou, (1991)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.
[135] I.M. Navon, X. L. Zou, J. Derber and J. Sela, (1992) Variational data assimilation with an adiabatic version of the NMC Spectral Model. Mon. Wea. Rev. 122 1433-1446.
[136] I.M. Navon, X.Zou, M. Berger, P.K.H. Phua, T.Schlick, F.X.Le Dimet, (1992) 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., , 1232pp, 33-48.
[137] I.M. Navon, X.Zou, M. Berger, P.K.H. Phua, T.Schlick, F.X.Le Dimet, (1992) 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., 1232pp, 445-480.
[138] I.M. Navon , Dacian N. Daescu, Z. Liu ,(2005) 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.
[139] H.E. Ngodock, (1986) Data Assimilation and sensitivity Analysis. Ph.D. Thesis, LMC-IMAG Laboratory, University Joseph Fourier, Grenoble, France, , pp 213.
[140] 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.
[141] J. Nocedal,(1980) Updating quasi-Newton matrices with limited storage. Mathematics of Computation, , 773-782.
[142] J. Nocedal, S.J. Wright, (1999) Numerical optimization, Springer,
[143] E.M. Oblow, (1983) An automated procedure for sensitivity analysis using computer calculus, Tech. Report ORNL/TM-8776, ORNL, .
[144] Park, S. K., and D. Zupanski, (2003): Four-dimensional variational data assimilation for mesoscale and storm-scale applications. Meteor. Atmos. Phys.,82, 173-208.
[145] H.A. Panofsky, (1949) Objective weather-map analysis. J. Appl. Meteor. 6 386-392.
[146] S.K. Park, E. Kalnay, (1999)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, San Antonio, Texas, SIAM, 104.
[147] D.F. Parrish, J.C. Derber, (1992) The National-Meteorological-Centers' spectral statistical-interpolation analysis system. Mon. Wea. Rev. 120 (8), 1747-1763.
[148] V. Penenko, N.N. Obratsov, (1976)A variational initialization method for the fields of the meteorological elements. Meteorol. Gidrol. 11, 1-11. (English Translation)

[149] Pin, F. G., E. M. Oblow, J. E. Horwedel, and J. L. Lucius, (1987) ADGEN: An Automated Adjoin Code Generator for Large-Scale Sensitivity Analysis, Trans. Am. Nucl. Soc. 55, 311.
[150] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, E.F. Mishchenko, (1962) The Mathematical Theory of Optimal Processes. John Wiley & Sons, Inc. New York.
[151] M.J.D. Powell, (1969)A method for nonlinear constraints in minimization problems. Optimization, R. Fletcher, Ed., Academic Press, Chap. 19 pp 283-298.
[152] Z.X. Pu, E. Kalnay , J. Sela, (1997) Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method . Mon. Wea. Rev Vol 125(10), 2479-2503.
[153] Z.X. Pu, E. Kalnay, (1999) Targeting observations with the quasi-inverse linear and adjoint NCEP global models: Performance during FASTEX. Quart. J. Roy. Meteor. Soc. 125 (561) ,3329-3337.
[154] F. Rabier, P. Courtier, (1992)Four-dimensional assimilation in the presence of baroclinic instability. Q. J. R. Meteorol. Soc. 118 649-672.
[155] 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,(1997): Recent experimentation on 4D-Var and first results from a Simplified Kalman Filter. October 1997.ECMWF Technical Memorandum No. 240.
[156] Rabier, F., 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, 1143-1170.
[157] F. Rabier, (2005) Overview of global data assimilation developments in numerical weather-prediction centres. Quarterly Journal of the Royal Meteorological Society 131 (613) 3215-3233.
[158] L.B. Rall,(1981) Automatic differentiation: Techniques and applications, Lecture Notes in Computer Science, vol. 120, Springer-Verlag, Berlin,
[159] Richard Courant and David Hilbert, (1962) Partial Differential equations. Methods of Mathematical Physics. Wiley-Interscience, New York,.
[160] L.F. Richardson, (1922) Weather prediction by numerical processes. Cambridge University Press. Reprinted by Dover (1965, New York). With a New Introduction by Sydney Chapman, xvi+236.
[161] Rodgers, C.D., (2000) Inverse methods for atmospheres: theories and practice. World Scientific Publ., Singapore.
[162] S. Saha, (1992) Response of the NMC MRF model to systematic error correction within the integration. Mon. Weather Rev. 120, 345-360.
[163] Y.K. Sasaki, (1955) A fundamental study of the numerical prediction based on the variational principle, J. Meteor. Soc. Japan 33 262-275.
[164] Y.K. Sasaki, (1958) An objective analysis based on the variational method, J. Meteor. Soc. Japan 36 77-88.
[165] Y.K. Sasaki, (1969) Proposed inclusion of time-variation terms, observational and theoretical in numerical variational objective analysis, J. Meteor. Soc. Japan 47 115- 203.
[166] Y.K. Sasaki, (1970a) Some basic formalisms in numerical variational analysis, Mon. Wea. Rev. 98 , 857-883.
[167] Y.K. Sasaki, (1970b) Numerical variational analysis formulated under the constraints as determined by long-wave equations as a low-pass filter, Mon. Wea. Rev. 98 , 884-898.
[168] Y.K. Sasaki, (1970c) Numerical variational analysis with weak constraint and

application to the surface analysis of severe storm gust, Mon. Wea. Rev. 98 , 899-910.
[169] Y.K. Sasaki, (1970d) A theoretical interpretation of anisotropically weighted smoothing on the basis of numerical variational analysis, Mon. Wea. Rev. 99 698-707.
[170] R.S. Seaman, (1988) Some real data tests of the interpolation accuracy of Bratseth's successive corrections method. Tellus 40A , 173-176.
[171] Simmons, A. J. and Hollingsworth, A. (2002) Some aspects of the improvement in skill of numerical weather prediction. Q. J. R. Meteorol. Soc., 128, 647-677.
[172] G.L.G. Sleijpen, H.A. Van der Vorst, (1996) A Jacobi-Davidson iteration method for linear eigenvalue problems. SIAM J Matrix Anal Vol 17, Issue 2 401-425 .
[173] W.C. Thacker, (1989)The Role of Hessian Matrix in Fitting models to Measurements. J. Geophys. Res., **94 (C5)**, 6177-6196.
[174] I.G. Tadjbakhsh , (1969) Utilization of Time-Dependent Data in Running Solution of Initial Value Problems. Journal of Applied Meteorology 8(3) 389-391.
[175] O. Talagrand, K. Miyakoda, (1971) Assimilation of past data in dynamical analysis.2. Tellus 23(4-5) ,318-327.
[176] O. Talagrand, (1981) On the mathematics of data assimilation Tellus 33 (4) ,321-339 .
[177] O. Talagrand, P. Courtier, (1987) Variational assimilation of meteorological observations with the adjoint vorticity equation-Part 1. Theory. Q. J. R. Meteorol. Soc. 113 1311-1328.
[178] O. Talagrand, (1997) Assimilation of observations, an introduction . Journal of the Meteorological Society of Japan 75 (1B) (1997) 191-209.
[179] A. Tarantola, (1987) Inverse problem theory. Methods for data fitting and model parameter estimation. Elsevier, Amsterdam, pp613.
[180] J.-N. Thepaut, P. Moll, (1990) Variational inversion of simulated TOVS radiances using the adjoint technique. Quart. J. Roy. Meteorol. Soc., 1425-1448.
[181] J.N. Thepaut, P. Courtier, (1991) Four-dimensional variational assimilation using the adjoint of a multilevel primitive-equation model. Q. J. R. Meteorol. Soc. 117 1225-1254.
[182] J.N. Thepaut, D. Vasiljevic, P. Courtier, J. Pailleux, (1993) Variational assimilation of meteorological observations with a multilevel primitive-equation model. Q. J. R. Meteorol. Soc. 119 153-186
[183] H.J. Thiébaux, L.L. Morone, (1990) Short-term systematic errors in global forecasts: their estimation and removal. Tellus 42A 209-229
[184] P.D. Thompson, (1969) Reduction of Analysis Error Through Constraints of Dynamical Consistency. Journal of Applied Meteorology 8 738-742.
[185] M. D. Tsyrulnikov, (2005) Stochastic modelling of model errors: A simulation study. Q. J. R. Meteorol. Soc. 131 pp. 3345-3371 .
[186] F. Uboldi, M. Kamachi, (2000) Time-space weak-constraint data assimilation for nonlinear models. Tellus 52 A 412-421 .
[187] Veerse, F. and Thepaut, J.-N. (1998) Multiple-truncation incremental approach for four-dimensional variational data assimilation. Q. J. R. Meteorol. Soc., 124, 1889-1908.
[188] F. Veerse, (1999) Variable-storage quasi-Newton operators as inverse forecast/analysis error covariance matrices in data assimilation. INRIA Technical Report 3685 Theme 4, pp 28.
[189] P.A. Vidard, E. Blayo, F.-X. Le Dimet, A. Piacentini, (2000)

4-D variational data analysis with imperfect model. Flow, Turbulence and Combustion 65 ,489-504
[190] Z. Wang, I. M. Navon, F. X. Le Dimet, X. Zou, (1992) The second order adjoint analysis: Theory and application. Meteorol. and Atmos. Phy. 50 3-20.
[191] Z. Wang, I. M. Navon and X. Zou, (1993) The adjoint truncated Newton algorithm for large-scale unconstrained optimization. Tech. Rep. FSU-SCRI-92-170, Florida State University, Tallahassee, Florida, , pp 44.
[192] Z. Wang, I. M. Navon, X. Zou, and F.X. Le Dimet, (1995) A Truncated Newton Optimization Algorithm in Meteorology Applications with Analytic Hessian Vector Products. Computational Optimization and Applications, 4, 241-262 .
[193] Z. Wang, K.K. Droegemeier, L. White, I. M. Navon, (1997) Application of a New Adjoint Newton Algorithm to the 3-D ARPS Storm Scale Model Using Simulated Data. Monthly Weather Review 125 (10) 2460-2478 .
[194] Z. Wang, K. Kelvin, Droegemeier, L. White, (1998) The Adjoint Newton algorithm for large-scale unconstrained optimization in meteorology applications. Computational Optimization and Applications 10 283-320.
[195] C.F. Weber, D.G. Cacuci, E.M. Oblow, (1979) Differential sensitivity theory for nonlinear equations with nonlinear responses. Trans. Am. Nucl. Soc. ; Vol/Issue: 33; American Nuclear Society meeting; 12 Nov San Francisco, CA, USA.
[196] W. Wergen, (1992) Effect of model errors in variational assimilation Tellus 44 A 297-313.
[197] N. Wiener, (1949), Extrapolation, interpolation and smoothing of stationary time series with engineering applications. M.I.T. Press, Cambridge, MA, pp 163.
[198] E.P. Wigner, (1945). Effects of small perturbations on pile period, Chicago Rep., CP-G 3048.
[199] L. Xu, T. Rosmond, R. Daley, (2005) Development of NAVDAS-AR: formulation and initial tests of the linear problem. Tellus 57A, 546-559
[200] A. Yaglom, (1962) An Introduction to the Theory of Stationary Random Functions, Dover, NY,.
[201] W. Yang, I. M. Navon, P. Courtier, (1996) A New Hessian Preconditioning Method Applied to Variational Data Assimilation Experiments Using Adiabatic Version of NASA/GEOS-1 GCM. Monthly Weather Review 124 (5), 1000-1017.
[202] K. Zhu, I.M. Navon, X. Zou, (1994) Variational Data Assimilation with a Variable Resolution Finite-Element Shallow-Water Equations Model. Monthly Weather Review 122 (5) 946-965.
[203] J. Zhu, M. Kamachi, (2000) An adaptive variational method for data assimilation with imperfect models. Tellus 52 A 265-279.
[204] J. Zou, W. H. Hsieh, and I. M. Navon (1995) Sequential Open-Boundary Control by Data Assimilation in a Limited Area Model, , *Monthly Weather Review*, **123, No. 9**, 2899-2909.
[205] X. Zou, I.M. Navon, F. X. Le Dimet, (1992) Incomplete observations and control of gravity waves in variational data assimilation. Tellus 44A 273-296.
[206] X. Zou, I.M. Navon, M. Berger, Paul K. H. Phua, T. Schlick, F. X. Le Dimet, (1993) Numerical experience with limited-memory quasi-Newton methods and truncated Newton methods. SIAM Jour. on Numerical Optimization, 3, 582-608.
[207] M. Zupanski, (1993) Regional four-dimensional variational data assimilation in a quasi-operational forecasting Environment. Mon. Wea. Rev. 121 , 2396-2408.
[208] D. Zupanski, (1997)A general weak constraint applicable to operational 4DVAR data assimilation systems. Mon. Weather Rev. 125 , 2274-2292.
[209] M. Zupanski, E. Kalnay, (1999) Principles of Data Assimilation. Global

Energy and Water Cycles, Cambridge Univ. Press. Ed. K.A. Browning and R.J. Gurney, 48-54.
[210] D. Zupanski, M. Zupanski, (2002) Fine-resolution 4D-VAR data assimilation for the great plains tornado outbreak of 3 May 1999. Weather and Forecasting 121, 506-525.
[211] M. Zupanski, D. Zupanski, T. Vukicevic, K. Eis, T. Haar, (2005) CIRA/CSU Four- Dimensional variational data assimilation system. Mon. Weather Rev. 123, 829-843.
[212] D. McLaughlin, O'Neill A, Derber J, M.Kamachi , (2005) Opportunities for enhanced collaboration within the data assimilation community . Quarterly Journal of the Royal Meteorological Society 131 (613), 3683-3693, Part C.