4-D Var Data Assimilation

Ionel M. Navon, R. Ştefănescu

Department of Scientific Computing
Florida State University
Tallahassee, Florida

November 26, 2012

Contents

Contents

Contents

Contents

Contents

Contents

Introduction of optimal control in Numerical Weather Prediction

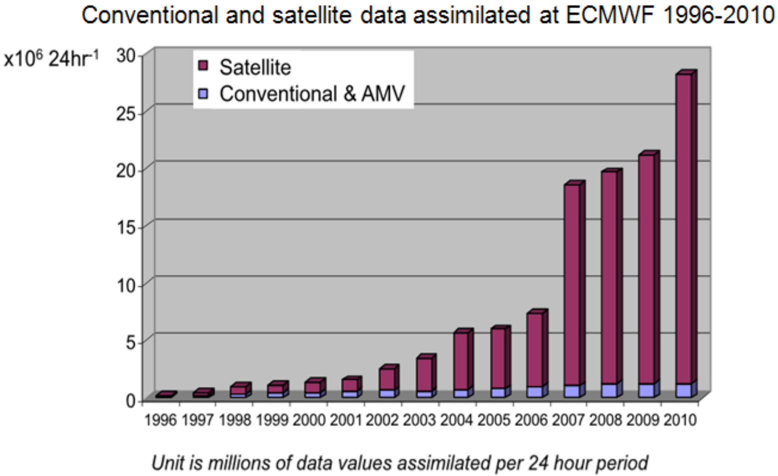
- Variational methods (Sasaki, 1958, 1970), Marchuck (1967 - onwards), Penenko and Obratsov (1976) - linear adjoint sensitivity
- Adjoint operators introduced in modern times by Wigner (1945).
- Optimal control of PDE's - Bellman (1957), Pontryagin et al. (1962)
- Major impetus is the monograph of Lions (1968) on optimal control of PDE's.
- Adjoint sensitivity analysis (Cacuci 1981a, 1981b) presented in a rigorous manner for general nonlinear systems of equations.
- Le Dimet F.X. (1981,1982) introduced in a technical report optimal control in meteorology.

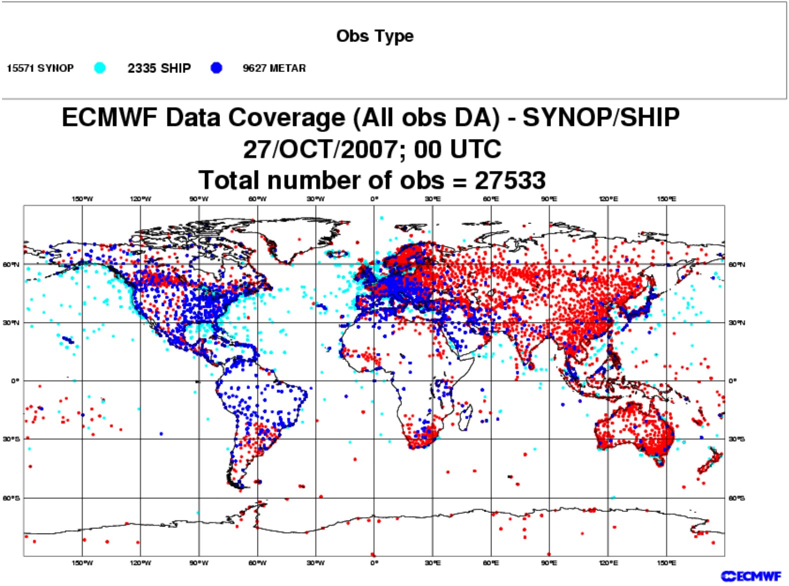
Early 4-D Var works

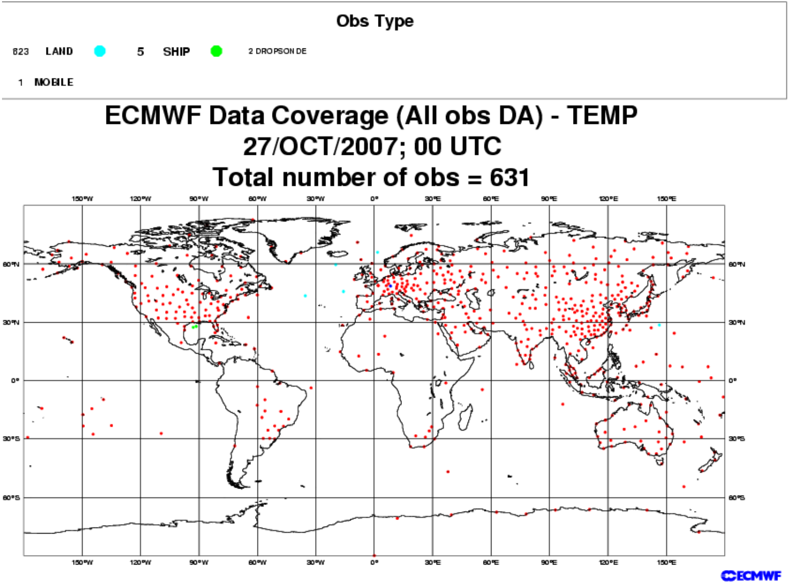
- Le Dimet and Talagrand (1986), Talagrand and Courtier (1987) tested 4-D Var on highly simplified models including large-scale atmospheric dynamics.
- Navon et al. (1992) considered the National Meteorological Center Spectral model and its adjoint
- Despite encouraging results, 4-D Var data assimilation remained impractical for operational use due to the prohibitively high computational cost.

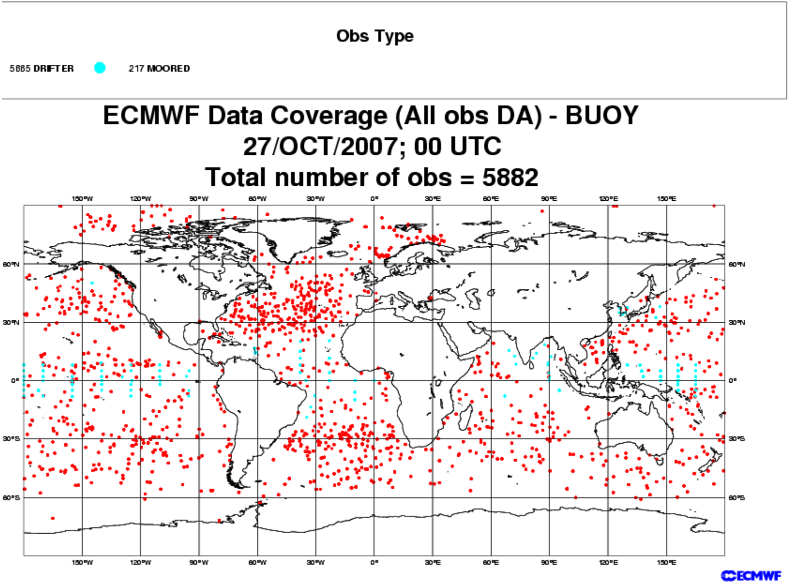
Data

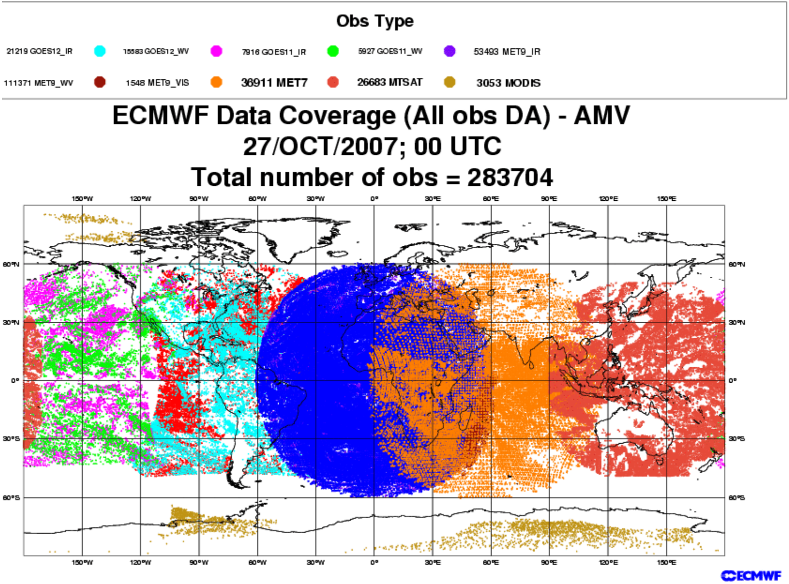
- More than 300 million
- Pilot balloons (wind)
- Radiosondes
- Drifting buoys (temperatures of air and ocean, salinity and wind)
- Satellites
- Geostationary (10) around $300,000$ observations assimilated
- Polar orbiting satellites - Vertical temperature profiles (radiances are measured then temperatures are estimated as solution of an inverse problem)
- Oceanography (scarcer data) altimetric satellites, Lagrangian floates.
- 800 million of variables of ECMWF operational model in 2007 assimilate data is 18 million. Retrieving state of atmosphere from observations is clearly an ill-posed problem.


Conventional and satellite data assimilated at ECMWF 1996-2010
x10^6 24hr^-1
30
25
20
15
10
5
0
Satellite
Conventional & AMV
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
Unit is millions of data values assimilated per 24 hour period

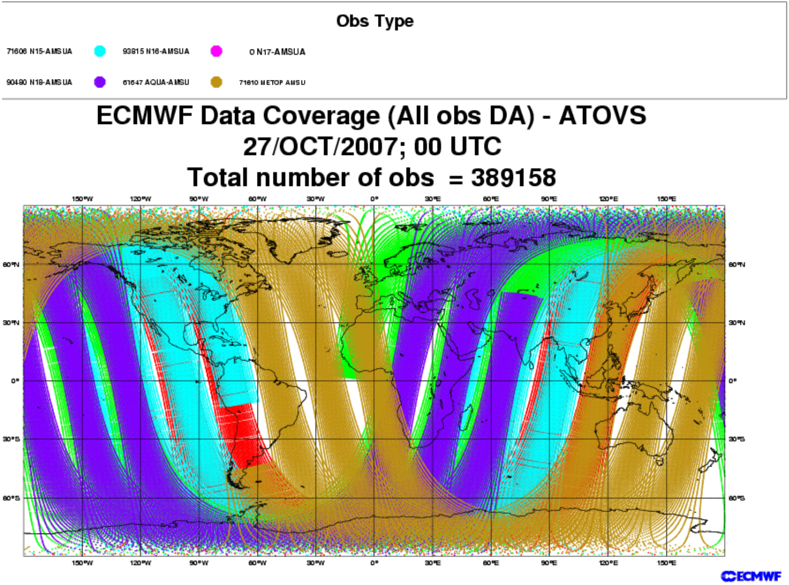

Obs Type
15571 SYNOP
2335 SHIP
9627 METAR
ECMWF Data Coverage (All obs DA) - SYNOP/SHIP
27/OCT/2007; 00 UTC
Total number of obs = 27533
150°W
120°W
90°W
60°W
30°W
0°
30°E
60°E
90°E
120°E
150°E
60°N
30°N
0°
30°S
60°S
ECMWF

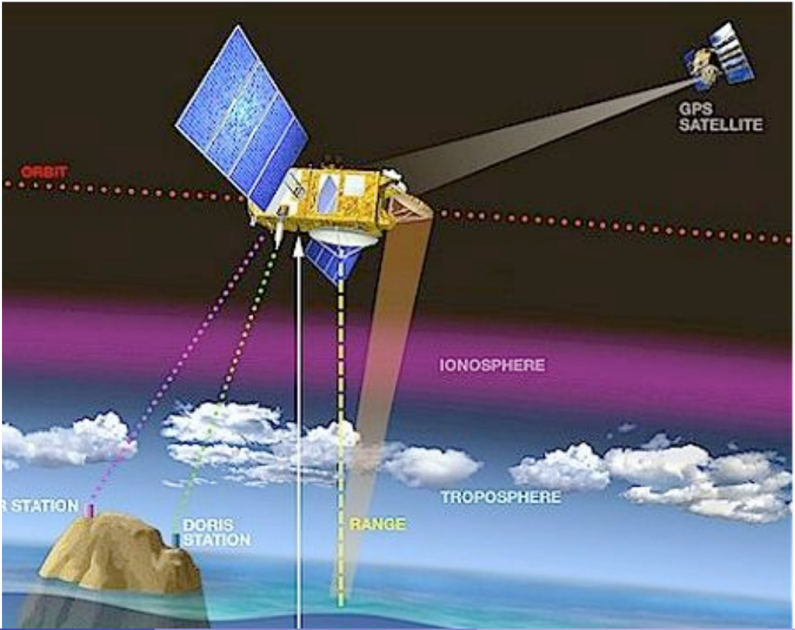

Obs Type
623 LAND
5 SHIP
2 DROPSONDE
1 MOBILE
ECMWF Data Coverage (All obs DA) - TEMP
27/OCT/2007; 00 UTC
Total number of obs = 631
150°W
120°W
90°W
60°W
30°W
0°
30°E
60°E
90°E
120°E
150°E
60°N
30°N
0°
30°S
60°S
ECMWF


Obs Type
5665 DRIFTER
217 MOORED
ECMWF Data Coverage (All obs DA) - BUOY
27/OCT/2007; 00 UTC
Total number of obs = 5882
150°W
120°W
90°W
60°W
30°W
0°
30°E
60°E
90°E
120°E
150°E
60°N
30°N
0°
30°S
60°S
ECMWF


Obs Type
21219 GOES12_IR
15583 GOES12_WV
7916 GOES11_IR
5927 GOES11_WV
53493 MET9_IR
111371 MET9_WV
1548 MET9_VIS
36911 MET7
26683 MTSAT
3053 MODIS
ECMWF Data Coverage (All obs DA) - AMV
27/OCT/2007; 00 UTC
Total number of obs = 283704
150°W
120°W
90°W
60°W
30°W
0°
30°E
60°E
90°E
120°E
150°E
60°N
30°N
0°
30°S
60°S
ECMWF

I.M. Navon, R. Ştefănescu




GPS
SATELLITE
ORBIT
IONOSPHERE
TROPOSPHERE
STATION
DORIS
STATION
RANGE

- A state variable $X \in \mathcal{X}$ describing evolution of the model at the grid points
- X depends on time and is of large dimension for operational models (10^8or more for European Centre for Medium-Range Weather Forecasts (ECMWF) model)
- A model describing evolution of the fluid. It is a system of nonlinear PDE's written as

$$\frac{dX}{dt} = F(X, U)$$

$$X(0) = V$$

- A control variable $(U, V) \in P$ space of control.
- In our case the control is the initial conditions and/or parameters or boundary conditions.

- For sake of simplicity we will not consider case of constraints on the state variable. Humidity and salinity cannot be negative. Hence the set of controls does not necessarily have the structure of a vector space
- Observations $X^{obs} \in O_{obs}$ are discrete and depend on space and time, and are not from either physical or geographical view point in same space as the state variable.
- We need to introduce an observation operator C mapping space of state into state of observations. The operator C can be as simple as interpolation operator or be very complex (imagine lightning observations which are not directly related to state control variables)

The cost functional

- A cost functional J (also called 'objective') measuring discrepency between model solution of the model associated to (U, V) and the observations

$$J(U, V) = \frac{1}{2} \int_0^T ||C(X(U, V)) - X^{obs}||^2 dt$$

- The choice of norm is quite important as it allows introducing 'a-priori' information - the statistics of the fields via the covariance matrix which is positive definite.
- In practice, an additional term the so-called background term, a quadratic difference between initial optimal variable and the last prediction is introduced
- It acts like a Tikhonov regularization term.

The cost functional

- Typically in meteorology

$$J(X) = \frac{1}{2}(X - X_b)^T B^{-1}(X - X_b) + \frac{1}{2}[X^{obs} - C(X)]^T R^{-1}[X^{obs} - C(X)]$$

in discrete form

$$J(X) = \frac{1}{2}(X - X_b)B^{-1}(X - X_b) +$$

$$\frac{1}{2}\sum_{k=0}^{N}[X_k^{obs} - H_k(M_k(X))]^T R_k^{-1}[X_k^{obs} - H_k(M_k(X))],$$

for the time interval $[t_0, t_N]$.

- M -forward model, H - observation operator
- It can also be written

$$J(X) = \frac{1}{2}||X - X_b||^2_{B^{-1}} + \frac{1}{2}\sum_{k=0}^{N}||H_k(X_k) - X_k^{obs}||^2_{R_k^{-1}},$$

Model error in 4D Var

- The three-dimensional model error at time t_k

$$\eta_k = X_k - M_k(X_{k-1}),$$

$$J(x_0, \eta_0, .., \eta_{N-1}) = \frac{1}{2}(X - X_b)B^{-1}(X - X_b) +$$

$$\frac{1}{2}\sum_{k=0}^{N}[H_k(X_k) - X_k^{obs}]^T R_k^{-1}[H_k(X_k) - X_k^{obs}] + \frac{1}{2}\sum_{k=0}^{N-1}\eta_k^T Q_k^{-1}\eta_k,$$

  where Q_k is the covariance matrix associated with the model errors η_k.
- Assumptions - errors in the observations and observation operators are assumed to be unbiased, Gaussian and uncorrelated with other sources of error.
- They are characterized by the model error covariance matrices Q_k.
- The prior estimate of the state of the model usually consists in a forecast from the most recent analysis without resorting to the current observations X^{obs}.

Model error in 4D Var

- This defines a weak constrained 4-D Var problem, in which the model equations do not have to be exactly satisfied over the assimilation window $[t_0, t_N]$.
- The prior estimate of the mean of the state is represented by X_b and called the "background".
- We assume that the background error is unbiased and uncorrelated with the other errors. It is characterized by the background error covariance matrix B
- Errors in $\eta(X)$ are assumed also to be unbiased, Gaussian and uncorrelated with other sources of error. They are characterized by their covariance matrix Q.
- Another way that is mathematically equivalent to the previous

$$J(X) = \frac{1}{2}(X-X_b)B^{-1}(X-X_b)+\frac{1}{2}\sum_{i=1}^{n}[H(x_i)-X_i^{obs}]^T R_i^{-1}[H(x_i)-X_i^{obs}]$$

$$+\frac{1}{2}\sum_{i=0}^{N-1}[X_{i+1}-M_i(X_i)]^T Q_i^{-1}[X_{i+1}-M_i(X_i)].$$

I.M. Navon, R. Ştefănescu

Model error in 4D Var

- What is the information content of data?
- One can conduct observing system experiments (OSEs)
- One diagnostic is

$$X^a = X^b + (B^{-1} + \mathbf{H}^T R^{-1} \mathbf{H})^{-1} \mathbf{H}^T R^{-1} (X^{obs} - H(x^b))$$

$$X^a = X^b + Kd$$

- B represents background error covariance matrix
- R observation error covariance
- $\mathbf{H}$ is the linearized observation operator of H that allows the computation of the model equivalents in the observation space

Model error in 4D Var

- K the Kalman gain matrix

$$K = (B^{-1} + \mathbf{H}^T R^{-1} H)^{-1} \mathbf{H}^T R^{-1}$$

- d innovation vector $= (X^{obs} - H(x^b))$
- To measure impact of observations on the analysis the Degrees of Freedom for Signal has been used.
- An information measure referred to as DOF for signal (d_s) is often used in information theory

$$d_s = \mathrm{tr}[I_{\mathrm{state}} - P_a P_f^{-1}],$$

  where P_a and P_f are the analysis and forecast error covariance matrices respectively.
- tr - denotes trace and I_{state} is the identity matrix of dimension $N_{\mathrm{state}} \times N_{\mathrm{state}}$.

Model error in 4D Var

- d_s- measures the forecast error reduction due to new information from the observations
- Another equivalent notation is

$$d_s = \mathrm{tr}[R^{-\frac{1}{2}} H P_a H^T R^{-\frac{1}{2}}] = tr[A]$$

- This concept can be linked to Kalman Filter and 3-D Var methods along with Shannon information content.

Optimality system

- The problem of variational data assimilation can be set as

$$\begin{cases} \text{Find} \quad U^*, V^* \in \mathcal{P} \ \text{such that :} \\ J(U^*, V^*) = \underset{(U,V)\in\mathcal{P}}{\mathrm{Inf}} J(U, V) \end{cases}$$

- It is an unconstrained optimization problem with respect to (U, V).

- The problem will have a unique solution if J is strictly convex, lower semi continuous and if:

$$\lim_{||(U,V)||\to+\infty} J(U, V) \to +\infty$$

Optimality system

- When J is differentiable, a necessary condition for (U^*, V^*) to be a solution is given by the Euler-Lagrange equation:

$$\nabla J(U^*, V^*) = 0$$

  where ∇J is the gradient of J with respect to (U, V).
- Determination of ∇J permits one to implement optimization methods of gradient type.

Optimality system

- Let $(u, v) \in \mathcal{P}$, $\widehat{X}$ be the Gâteaux-derivative (directional derivative) of X in the direction (u, v) that is the solution of:

$$\begin{cases} \dfrac{d\widehat{X}}{dt} = \left[\dfrac{\partial F}{\partial X}\right].\widehat{X} + \left[\dfrac{\partial F}{\partial U}\right].u, \\ \widehat{X}(0) = v \end{cases}$$

where $\left[\dfrac{\partial F}{\partial X}\right]$ is the Jacobian of the model with respect to the state variable. This equation is know as the *linear tangent model*.

Optimality system

- By same token we get the directional derivative of J:

$$\widehat{J}(U, V, u, v) = \int_0^T \left(C.X - X_{obs}, C.\widehat{X} \right) dt.$$

- The gradient is obtained by exhibiting the linear dependence of $\widehat{J}$ with respect to (u, v).

Optimality system

- We introduce $P \in \mathcal{X}$ the so called *adjoint variable*.
- The adjoint system is calculated by taking the inner product of tangent linear system with P and then integrate between 0 and T.
- An integration by part shows that P is defined as the solution of:

$$\begin{cases} \dfrac{dP}{dt} + \left[\dfrac{\partial F}{\partial X}\right]^T .P = C^t\left(C.X - X_{obs}\right) \\ P\left(T\right) = 0. \end{cases}$$

- Therefore the gradient is obtained by a backward integration of the adjoint model.

$$\nabla J = \begin{pmatrix} \nabla_U J \\ \nabla_V J \end{pmatrix} = \begin{pmatrix} -\left[\dfrac{\partial F}{\partial U}\right]^t .P \\ -P\left(0\right) \end{pmatrix}$$

Optimization

- The determination of (U^*, V^*) is carried out by performing a descent-type unconstrained optimization method.
- Given a first guess (U_0, V_0), we define a sequence by :

$$\begin{pmatrix} U_n \\ V_n \end{pmatrix} = \begin{pmatrix} U_{n-1} \\ V_{n-1} \end{pmatrix} + \rho_n D_n$$

  where D_n is the direction of descent.
- Usually conjugate gradient or Newton type methods are used. ρ_n is the step size defined by:

$$\rho_n = \mathrm{ArgMin}\ J\left(\begin{pmatrix} U_{n-1} \\ V_{n-1} \end{pmatrix} + \rho D_n\right).$$

Optimization

- For a non linear problem the minimization entails a high computational cost since several integrations of the model are required for the evaluation of J.
- Optimization libraries, e.g. MODULOPT are widely used and are efficient.
- For a comprehensive test of powerful large-scale unconstrained minimization methods applied to variational data assimilation see Zou, Navon, Berger, Phua, Schlick and Le Dimet (1993).

Differentiate then discretize vs discretize-then-differentiate

- In differentiate then discretize one obtains adjoint equations at the PDE level and then taking adjoint of the results. One then discretizes the continuous adjoint system.
- In the discretize then differentiate one discretizes the continuous flow equations. Then one obtains adjoint equations of the differentiate discrete approximate flow equations.
- Differentiation and discretization steps do not commute.

Implementation

- A major difficulty encountered in the implementation of this method is the derivation of the adjoint model.
- A bad solution would be to derive the adjoint model from the continuous direct model, then to discretize it.
- The convergence of the optimization algorithm requires having the gradient of the cost function with a precision of the order of the computer's round-off error.

Implementation

- Two steps are carried out for the derivation of the adjoint :
  1. **Differentiation of the direct model.** This step serves to determine the linear tangent model. This task is easily carried out by differentiating the direct code line by line.
  2. **Transposition of the linear tangent model.** Transposition with respect to time is simply the backward integration. To carry out the transposition one should start from the last statement of the linear tangent code and transpose each statement. The difficulty stems from the hidden dependencies.

Implementation

- If some rules in the direct code are adhered to then the derivation of the adjoint model can be made simpler, otherwise it is a long and painful task.
- We can use automatic differentiation codes such as ODYSSEE, TAPENADE, TAMC, FASTOPT (see Andreas Griewank book (2008)).

Remarks

- If the model is non linear then the cost function is not necessarily convex and the optimization algorithm may converge toward a local minimum.
- In this case one can expect convergence toward a global optimum only if the first guess is in the vicinity of the solution.
- This may occur in meteorology where the former forecast is supposed to be close to the actual state of the atmosphere. In practice a so-called background term is added to the cost function, measuring the quadratic discrepancy with the prediction.
- In terms of control this term could be considered as a regularization in the sense of Tykhonov.

Remarks

- The optimization algorithm could converge to a correct mathematical solution but would be physically incorrect (e.g. negative humidity).
- The solution may be far away from the attractor, the regularization term will force the model to verify some additional constraints, e.g. for the solution to remain in the vicinity of the geostrophic equilibrium.
- Regularization terms permit to take into account the statistical information on the error by an adequate choice of the quadratic norm including the error covariance matrix.
- If the control variable U is time dependent, which is the case if boundary conditions are controlled, then we may get problems with a huge dimension. In this case it will be important to choose an appropriate discretization of the control variable in order to reduce its dimension.

Measure of skill - the anomaly correlation coefficient

- Another way to measure the quality of a forecast system is to calculate the correlation between forecasts and observations. However, correlating forecasts directly with observations or analysis may give misleadingly high values because of the seasonal variations.
- It is therefore established practice to subtract the climate average from both the forecast and the verification and to verify the forecast and observed anomalies according to the anomaly correlation coefficient (ACC), which in its most simple form can be written:

$$ACC = \frac{\overline{(f-c)(a-c)}}{\sqrt{\overline{(f-c)^2}\ \overline{(a-c)^2}}}$$

Measure of skill - the anomaly correlation coefficient

- The World Meteorological Organization (WMO) definition also takes any mean error into account:

$$ACC = \frac{\overline{[(f-c)-\overline{(f-c)}][(a-c)-\overline{(a-c)}]}}{\sqrt{\overline{((f-c)-\overline{(f-c)})^2}\;\overline{((a-c)-\overline{(a-c)})^2}}}$$

- The ACC can be regarded as a skill score relative to the climate. It is positively orientated, with increasing numerical values indicating increasing "success".
- It has been found empirically that ACC=60% corresponds to the range up to which there is synoptic skill for the largest scale weather patterns.
- ACC=50% corresponds to forecasts for which the error is the same as for a forecast based on a climatological average, i.e. RMSE = A_a. An ACC of about 80% would correspond to a range where there is still some skill in large-scale synoptic patterns.

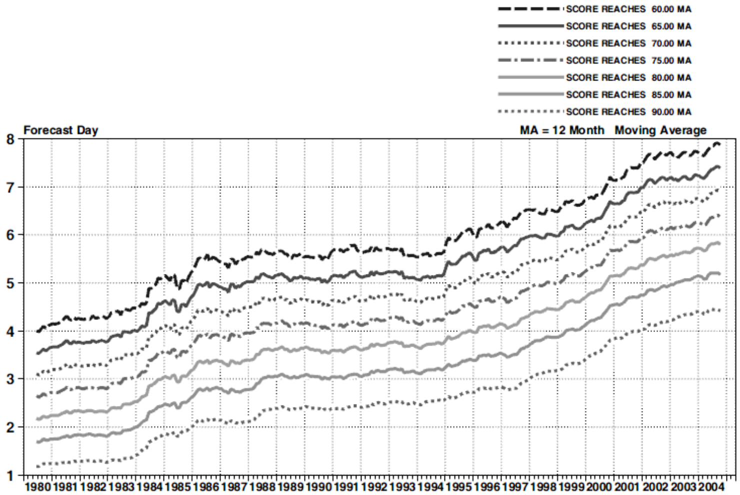


Fig.1 Time series of anomaly correlation reaching 60 **to** 90% **for geopotential errors at** 500 **hPa for ECMWF forecasts over the southern hemisphere (**$20 - 90^o$ **S). Running mean values are computed over periods of one year**

Example: The 2-D Shallow water equations model

- The shallow water equations simulate an incompressible fluid for which the depth is small with respect to the horizontal dimensions.
- General equations of geophysical fluid dynamics are vertically integrated using the hydrostatic hypothesis, therefore vertical acceleration is neglected.
- In Cartesian coordinates they are :

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

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

$$\frac{\partial \phi}{\partial t} + \frac{\partial u\phi}{\partial x} + \frac{\partial v\phi}{\partial y} = 0$$

Shallow water equations model

- In this system $X = (u, v, \phi)^T$ is the state variable, u and v are the components of the horizontal velocity; ϕ is the geopotential (proportional to the height of the free surface)and f the Coriolis' parameter.
- The following hypotheses are used:
  1. The error of the model is neglected. Only the initial condition will be considered as a control variable.
  2. Lateral boundary conditions are periodic. This is verified in global models.
  3. Observations are supposed to be continuous with respect to time. Of course this is not the case in practice. $\mathbf{C} \equiv \mathbf{I}$, where $\mathbf{I}$ is the identity operator.

I.M. Navon, R. Ştefănescu

Shallow water equations model

- If $U_0 = (u_0, v_0, \phi_0)^T$, is the initial condition and the cost function given by

$$J(U_0) = \frac{1}{2}\int_0^T [(u - u_{obs})^2 + (v - v_{obs})^2 + \gamma(\phi - \phi_{obs})^2]dt.$$

where γ is a weight function, then the directional derivatives $\bar{X} = (\bar{u}, \bar{v}, \bar{\phi})^T$ in the direction $h = (h_u, h_v, h_\phi)^T$ (in the control space) will be solutions of the linear tangent model:

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

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

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

Shallow water equations model

- The adjoint model is obtained by transposition of the linear tangent model.
- Let $P = (\tilde{u}, \tilde{v}, \tilde{\phi})^T$ be the adjoint variable and after some integrations by parts both in time and space we see that the adjoint model is defined as being the solution of:

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

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

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

with final conditions equal to 0.

I.M. Navon, R. Ştefănescu

Shallow water equations model

- The gradient of J is given by:

$$\nabla J(U_0) = -P(0) = -\begin{pmatrix} \tilde{u}(0) \\ \tilde{v}(0) \\ \tilde{\phi}(0) \end{pmatrix}.$$

Numerical Results

- SWE Initial conditions were derived from the initial height condition No. 1 of Grammeltvedt (1969) i.e.

$$h(x,y) = H_0 + H_1 + \tanh\left(9\frac{D/2-y}{2D}\right) + H_2 \operatorname{sech}^2\left(9\frac{D/2-y}{2D}\right)\sin\left(\frac{2\pi x}{L}\right)$$

$$0 \le x \le L,\ 0 \le y \le D.$$

- The initial velocity fields were derived from the initial height field using the geostrophic relationship

$$u = \left(\frac{-g}{f}\right)\frac{\partial h}{\partial y},\ v = \left(\frac{g}{f}\right)\frac{\partial h}{\partial x}.$$

- The constants used were $L = 6000km,\ D = 4400km,\ \hat{f} = 10^{-4}s^{-1}\ ,\ g = 10ms^{-2},$ $H_0 = 2000m,\ H_1 = -220m,\ H_2 = 133m.$

Numerical Results

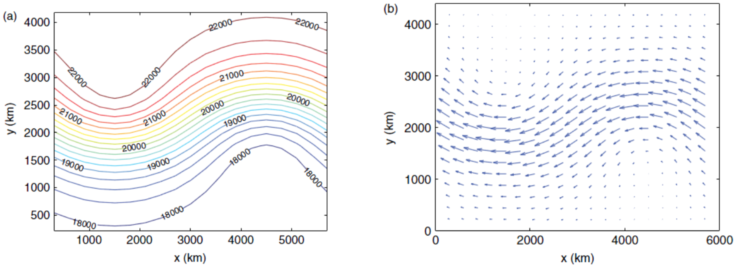


Fig.2 Initial conditions (u_0, v_0, ϕ_0): (a) ϕ_0 from Grammeltvedt (1969) and (b) wind field

I.M. Navon, R. Ştefănescu

Essence of data assimilation(History)

4-D Var - Theory of VDA

The incremental method in 4-D Var and its formulation

3D-Var Incremental Data Assimilation of Lightning using WRF model

Nudging Method

4-D Var issues

Numerical Results

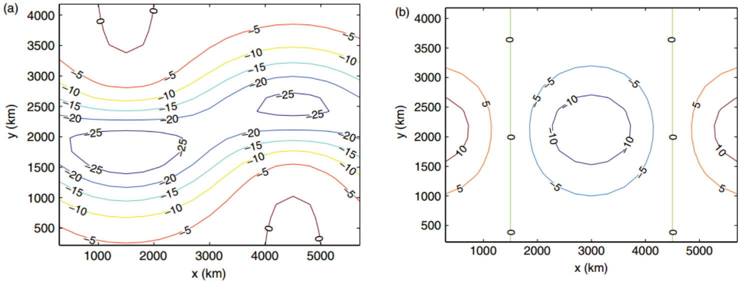


Fig.3 Contour plot for (a) u_0 and (b) v_0.

Numerical Results

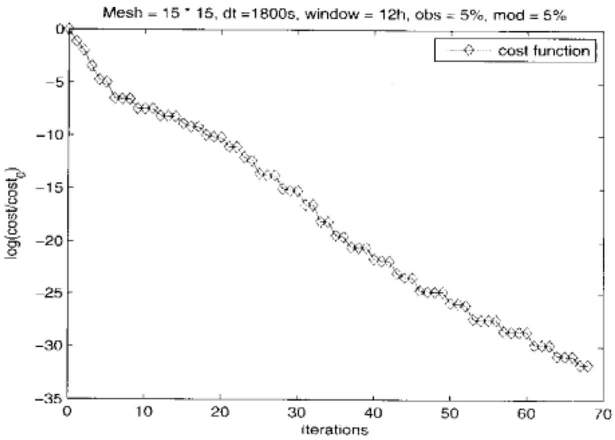
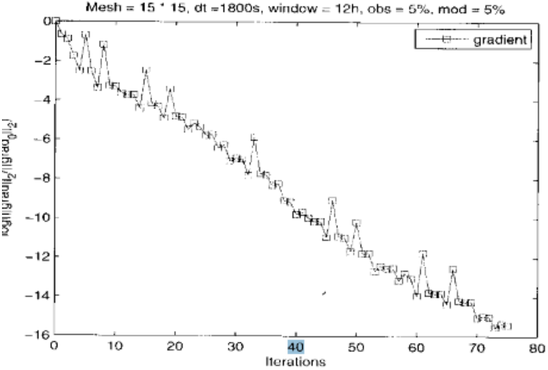


Fig.4 L-BFGS minimization: Data assimilation window = 12h, $\Delta x = \Delta y = 400km$, mesh resolution = 15 x 15, random perturbation = 5% (a) Normalized cost function scaled by initial cost function versus the number of minimization iterations

Numerical Results



(b) L2 norm of gradient

Fig.5 L-BFGS minimization: Data assimilation window = 12h, $\Delta x = \Delta y = 400km$, mesh resolution = 15 x 15, random perturbation = 5% (b) The norm of gradient scaled by initial nonn of the gradient versus the number of minimization iterations.

Numerical Results

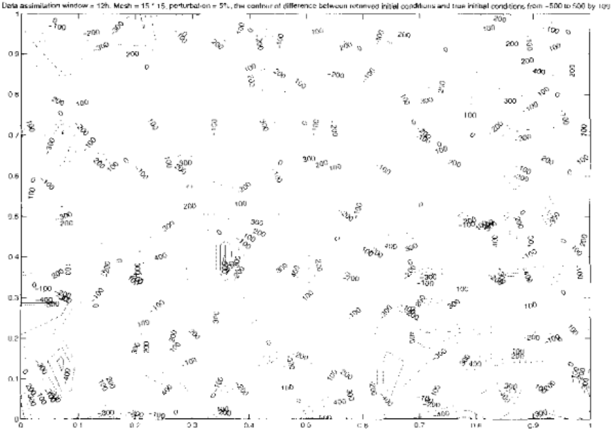


Fig.6 Data assimilation window = 12h, $\Delta x = \Delta y = 400km$, mesh resolution = 15 x 15, random perturbation = 5%. The contours of difference between retrieved initial geopotential and true initial geopotential are plotted.

Numerical Results

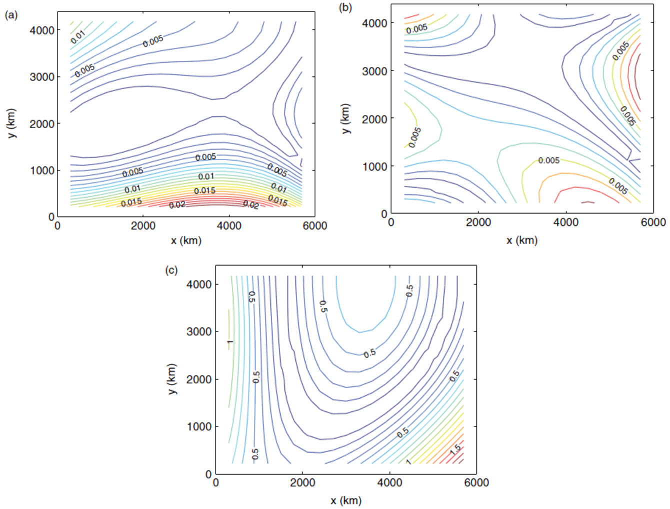


Fig.7 Final error of the computed versus exact solution. (a) u, contour interval 10^{-3}; (b) v, contour interval 10^{-3}; (c) ϕ, contour interval 10^{-1}

Numerical Results

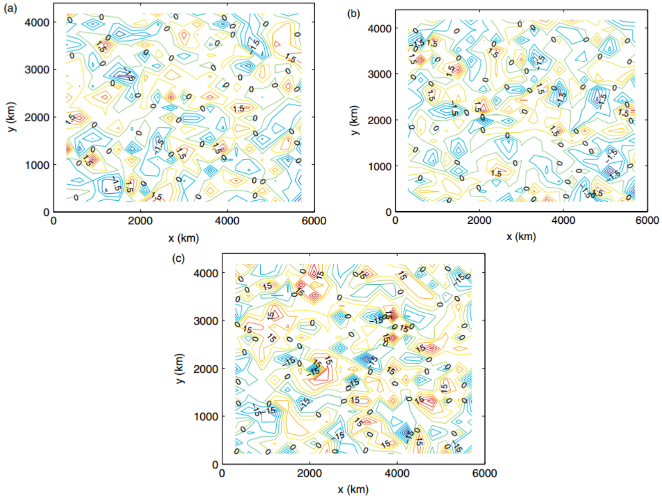


Fig.8 Observation error sample at t_0 for u, v and ϕ. (a) $\eta_{obs}^{(u)}$ sample, contour interval contour interval 0.5; (b) $\eta_{obs}^{(v)}$ sample, contour interval 0.5; (c)$\eta_{obs}^{(\phi)}$ sample, contour interval 5.

Second-order methods - Hessian

- The optimality system, the Euler-Lagrange equation, provides only a necessary condition for optimality.
- In the linear case, the solution is unique if the Hessian is definite positive.
- From a general point of view the information given by the Hessian is important for theoretical, numerical and practical issues.
- For operational models it is impossible to compute the Hessian itself, as it is a square matrix with 10^{14} terms nevertheless the most important information can be extracted from the spectrum of the Hessian which can be estimated without an explicit determination of this matrix.
- This information is of importance for estimating the condition number of the Hessian for preparing an efficient pre conditioning.

Second-order methods - Hessian

- A general method to get this information is to apply the techniques described above to the couple made by the direct and adjoint models (Le Dimet, Navon and Daescu (2002)and Wang, Navon, Le Dimet and Zou (1992)), leading to a so called second order adjoint.
- The following steps are carried out:
  1. Linearization of the direct and adjoint models with respect to the state variable. Since the system is linear with respect to the adjoint variable, no linearization is necessary.
  2. Introducing two second order adjoint variables.
  3. Transposition to exhibit the linear dependence with respect to the directions.

Second-order methods - Hessian

- If the model has the form:

$$\begin{cases} \dfrac{dX}{dt} = F(X) + B.U \\ X(0) = V \end{cases}$$

  U and V being the control variables.

- Considering the cost function usual definition and the corresponding adjoint equation, from a backward integration of the adjoint model the gradient is deduced:

$$\nabla J = \begin{pmatrix} \nabla_U J \\ \nabla_V J \end{pmatrix} = \begin{pmatrix} -B^T P \\ -P(0) \end{pmatrix}$$

Second-order methods - Hessian

- To calculate the second order derivative of J with respect to U and V we have to derive the optimality system (i.e. the model plus the adjoint system).
- By analogy to the first order case we introduce two so-called second order adjoint variables R and Q as the solution of the system:

$$\begin{cases} \frac{dR}{dt} = \left[\frac{\partial F}{\partial X}\right].R + B.\Lambda \\ \frac{dQ}{dt} + \left[\frac{\partial F}{\partial X}\right]^T.Q = -\left[\frac{\partial^2 F}{\partial X^2}.R\right]^T.P + C^T CR \end{cases}$$

where Λ has the dimension of U.

Second-order methods - Hessian

- If the Hessian of J is written

$$H(U,V) = \begin{pmatrix} J_{U,U} & J_{U,V} \\ J_{U,V} & J_{V,V} \end{pmatrix}$$

and if the system is integrated with the conditions :

$$Q(T) = 0$$
$$R(0) = \Theta$$

and $\Lambda = 0$, then :

$$J_{V,V}.\Theta = -Q(0)$$
$$J_{V,U}.\Theta = -B^T.Q$$

Second-order methods - Hessian

- If the system is integrated with the conditions :

$$Q(T) = 0$$
$$R(0) = 0$$

then we obtain:

$$J_{U,U}.\Lambda = -B^T.Q$$

Second-order methods - Hessian

- Therefore, without an explicit computation of the Hessian it is possible to compute the product of the Hessian by any vector and consequently , using classical methods of linear algebra, to evaluate its eigenvalues and eigenvectors and also to carry out Newton-type methods.
- It is worth pointing out that the R variable is the solution of the linear tangent model (when $\Lambda = 0$) and therefore no extra code has to be written in this case. The left hand side of the equation verified by Q is the adjoint model and only the code associated to its right hand side has to be written.

Second-order methods - Hessian

- In the case of the shallow water equations with the initial condition V as unique control vector (no model error), the state variable is $X = (u, v, \phi)$ and the adjoint variable is $P = (\tilde{u}, \tilde{v}, \tilde{\phi})$. For the second order, the variable R$= (\bar{u}, \bar{v}, \bar{\phi})^T$ is the solution of the linear tangent model, while the variable $Q = (\hat{u}, \hat{v}, \hat{\phi})^T$ is the solution of the equations

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

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

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

Second-order methods - Hessian

- From a formal point of view we see that first and second order differ by second order terms which do not take into account the adjoint variable.
- The computation of second derivatives requires storing both the trajectories of the direct and adjoint models.
- For very large models it could be more economical to recompute these trajectories.
- The system obtained, the second order adjoint, is used to compute the product of the Hessian by any vector.
- If we consider all the vectors of the canonical base, then it will be possible to get the complete Hessian.

Second-order methods - Hessian

- By using Lanczos type methods and deflation, it is possible to compute the eigenvectors and eigenvalues of the Hessian.
- To carry out second order optimization methods of Newton-type for equations of the form:

$$\nabla J(X) = 0.$$

- The iterations are:

$$X_{n+1} = X_n - H^{-1}(X_n) . \nabla J(X_n)$$

  where H is the Hessian of J.
- At each iteration a linear system should be solved. This is done by carrying out some iterations of a conjugate gradient method which requires computing the product Hessian-vector.

Numerical Results

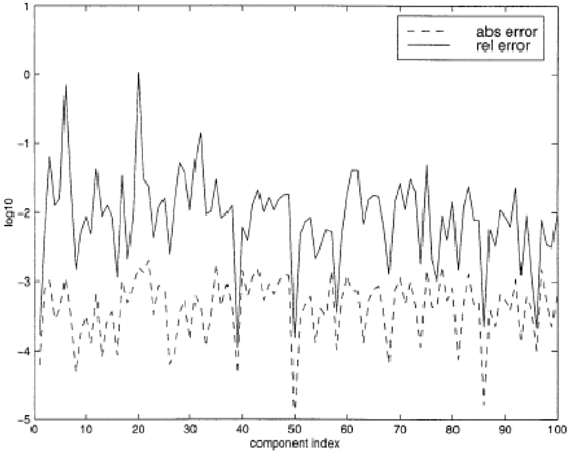


FIG. 1. The absolute (dashed line) and relative (solid line) differences between the Hessian/vector product computed with the SOA method and with the finite-difference method at the first iteration (initial guess state). The first 100 components are considered.

Numerical Results

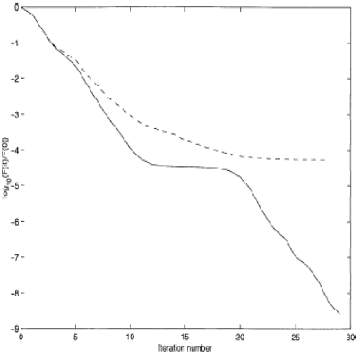


FIG. 2. The evolution of the normalized cost function during the minimization using the SOA method (solid line) and the finite-difference method (dashed line) to compute the Hessian/vector product.

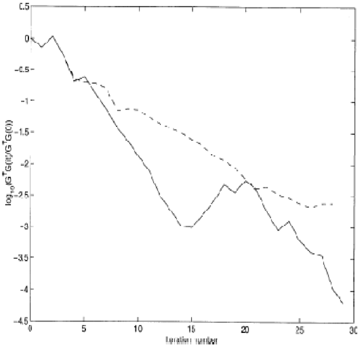


FIG. 3. The evolution of the normalized gradient norm during the minimization using the SOA method (solid line) and the finite-difference method (dashed line) to compute the Hessian/vector product.

The incremental method in 4-D Var and its formulation

- Following an idea by Derber, Courtier et al (1994) an incremental 4-D Var algorithm, which removes nonlinearities in the minimization by using a forward integration of a linear model instead of the non-linear one, was developed.
- This minimization can be carried out at a reduced model resolution which leads to an effective reduction of the computational cost.

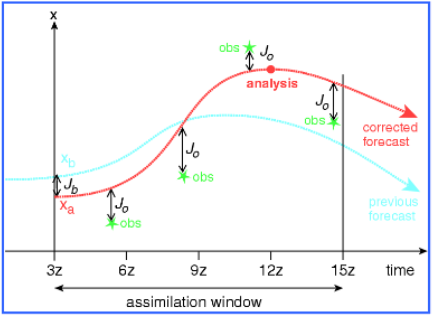
The incremental method in 4-D Var and its formulation

- Incremental 4D-Var algorithm
- Minimization algorithm
- Preconditioning

- In 4D-Var, each iteration of the minimization algorithm is very expensive
- It is crucial to use the most efficient minimization algorithms.

Exemple - 4D Variational Data Assimilation

- All observations within a 12 hour period are used simultaneously in one global estimation problem.
- 4D-Var finds the 12 hour forecast evolution that best fits the available observations.
- It does so by adjusting the initial condition for surface pressure and for the upper air fields of temperature, wind, humidity and ozone.



4D Variational Data Assimilation

- Model: M
- Observations: Y
- Background: X_b
- Observation operator: H
- Cost function to minimize

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

- In discrete form

$$J(X) = \tfrac{1}{2}(X - X_b)^T B^{-1}(X - X_b) + \tfrac{1}{2}\sum_{i=0}^{N}[Y_i - H_i(M_i(X))]^T R_i^{-1}[Y_i - H_i(M_i(X))],$$

The incremental method in 4-D Var and its formulation

- The incremental 4-D Var algorithm consists in minimizing a cost function expressed in terms of increments with respect to the background state $\delta X = X - X_b$.
- The observation operator H and the model M are linearized around $X_i = M_i(x_0)$.

$$J(\delta W_0) = \frac{1}{2}\delta W_0^T B^{-1} \delta W_0 + \\ \frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i\mathbf{M}_i\delta X_i - d_i)^T R_i^{-1}(\mathbf{H}_i\mathbf{M}_i\delta X_i - d_i),$$

where

- $\delta W_0 = S(X_0 - X_b)$ is the simplified increment at time t_0 and S is an operator transforming the field from high to low resolutions.
- $\delta X_i = X_i - X_{i-1}$ defines analysis increment of the i^{th} inner loop

The incremental method in 4-D Var and its formulation

- $\mathbf{H}_i$ and $\mathbf{M}_i$ are the linearized observation operator and model.
- $d_i,\ i = 0, .., N$ are the innovation departure or observation increment at time t_i.
- The innovations, which are the primary input to the assimilation, are always computed using the full observation operator and model to ensure the highest possible accuracy.
- Tangent linear approximation:

$$M(X + \delta X) \simeq M(x) + \mathbf{M}\delta x \quad \text{and} \quad H(X + \delta X) \simeq H(x) + \mathbf{H}\delta x$$

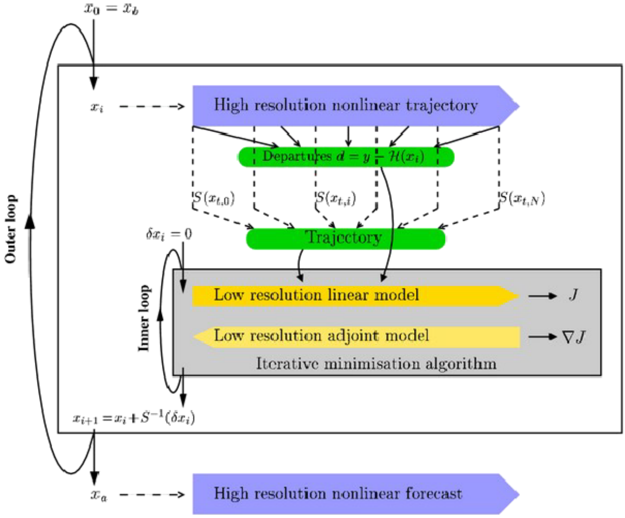
The incremental method in 4-D Var and its formulation

- Approximations to reduce cost: the tangent linear model (and its adjoint) is degraded with respect to the full model M:

  1. Lower resolution
  2. Simplified physics
  3. Simpler dynamics

- Shorter control vector and cheaper TL and AD models during minimization.

Incremental 4D-Var



The incremental method in 4-D Var and its formulation

- The solution δW_0^a of the cost function minimization is added to the background to obtain analysis at t_0

$$X_0^a = X_b + S^{-1}\delta W_0^a,$$

  where S^{-1} is an operator transforming the field from low to high-resolutions.
- It is a generalized inverse of the operator S that is going from high to low-resolutions.
- The minimization process is similar to the standard algorithm except that the control variable is the increment at time t_0, while the increment trajectory is obtained by the integration of the tangent linear model.
- The reference trajectory required by the linear and adjoint models is from the background integration and is not updated at each iteration.

The incremental method in 4-D Var and its formulation

- This drawback is overcome by an "outer-loop" updating the high resolution reference trajectory and the observation departures.
- Correspondingly, the iterative procedure of minimizing the incremental cost function is called the "inner-loop".
- It is possible to use a successively increased inner-loop resolution after each outer loop updated - referred to as "multi-incremental" algorithm.

The incremental method in 4-D Var and its formulation

- In this way we run the TLM and adjoint model at low-resolution in each inner loop iteration while in the outer loop iteration the nonlinear model is calculated at the higher resolution.
- **Computational savings made by implementing the inner loop in this way made incremental 4-D Var feasible for operational weather and ocean forecasting.**
- In practice very few outer loops steps are performed.
- ECMWF performs 3 outer loops while UK Met Office performs only one.

Minimization: Conjugate Gradient

- The Conjugate Gradient algorithm for solving linear systems can be used for minimizing a quadratic cost function with a symmetric positive-definite Hessian A:

$$J(x) = \frac{1}{2}x^T A x - b^T x + c$$

- The algorithm is
  1. $x_{k+1} = x_k + \alpha_k d_k,$
  2. $g_{k+1} = g_k + \alpha_k A d_k,$
  3. $d_{k+1} = -g_{k+1} + \beta_k d_k,$

  with $d_0 = -g_0, \quad \alpha_k = \frac{g_k^T g_k}{d_k^T A d_k}, \quad \beta_k = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}.$

Preconditioning

- The rate of convergence of the Conjugate Gradient algorithm depends on the condition number of A: $\kappa(A) = \lambda_1/\lambda_N$.
- The convergence can be accelerated by improving the condition number of the problem.
- Apply the change of variable $\hat{x} = Cx$:

$$J(\hat{x}) = \frac{1}{2}\hat{x}C^{-T}AC^{-1}\hat{x} - (C^{-T}b)^T\hat{x}.$$

- The difficulty is to find C so that

$$\kappa(C^{-T}AC^{-1}) \ll \kappa(A).$$

- If $\kappa(C^{-T}AC^{-1}) = 1$, the minimization would converge in one iteration.

4-D Var Preconditioning

- The Hessian of the 4D-Var cost function is $B^{-1} + H^T R^{-1} H$.
- Define L verifying:

$$B^{-1} = LL^T$$

- Change of variable $\chi = L^{-1}\delta X$
- The 4-D Var cost function becomes:
  $\hat{J}(\chi) = \frac{1}{2}\chi^T\chi + \frac{1}{2}(\mathbf{H}L\chi - d)^T R^{-1}(\mathbf{H}L\chi - d)$
- The Hessian of $\hat{J}$ is

$$\hat{J}'' = I + L^T\mathbf{H}^T R^{-1}\mathbf{H}L.$$

- The smallest eigenvalue is $\lambda_N = 1$

4-D Var Preconditioning

- $\sqrt{B^{-1}}$ preconditioning usually gives good results.
- First experiments with direct assimilation of Meteosat radiance data showed significant analysis differences far away from the observations.
- Differences disappeared when the number of iterations was increased: they were the result of insufficient convergence.
- A perfect preconditioner would be $L = \sqrt{(J'')^{-1}}$
- Approximation of the Hessian which can be easily inverted

Hessian Eigenvectors Preconditioning

- The Hessian can be written as: $J'' = \sum_{k=1}^{N} \lambda_k v_k v_k^T$ where λ_k and v_k are its eigenvalues and eigenvectors.
- Preconditioning based on the K leading eigenvectors of J'':

$$L^{-1} = I + \sum_{k=1}^{K} (\mu_k^{1/2} - 1) v_k v_k^T$$

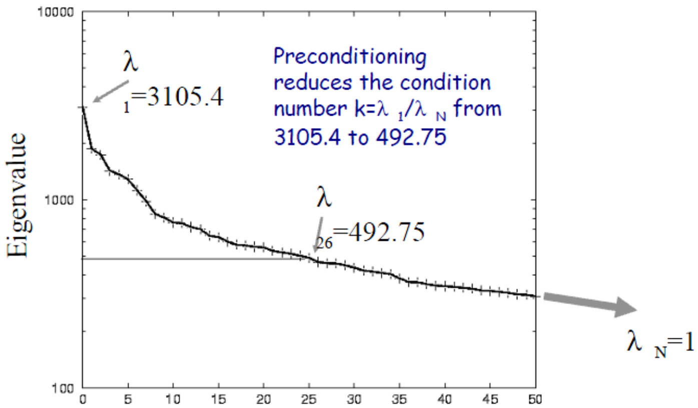
gives

$$\hat{J}'' = \sum_{k=1}^{K} \mu_k \lambda_k v_k v_k^T + \sum_{k=K+1}^{N} \lambda_k v_k v_k^T.$$

- Choose μ_k verifying $\mu_k \lambda_k < \lambda_{k+1}$, then

$$\kappa(\hat{J}'') = \lambda_{K+1} / \lambda_N = \lambda_{K+1}.$$

4D-Var Eigenvalues



Conjugate Gradient and Lanczos Algorithm

- The relation between Conjugate Gradient and Lanczos Algorithms allows to simultaneously:
  1. Minimize the cost function
  2. Compute the eigenvectors and eigenvalues of the Hessian.
- The connection can be exploited to improve the minimization:
  1. At each outer loop iteration, the eigenvectors and eigenvalues of the Hessian can be computed
  2. They are used to precondition the minimization in the following outer loop iteration.

Superlinear Convergence

- The convergence rate of conjugate gradients is largely determined by the condition number $\kappa = \lambda_{max}/\lambda_{min}$
- When λ_{max} or λ_{min} has converged in the Lanczos process, $J(x)$ has been fully minimized in the direction of the eigenvector.
- The minimization then behaves as if it were minimizing a problem with a smaller κ
- Conjugate Gradient converges superlinearly.

Numerical Results

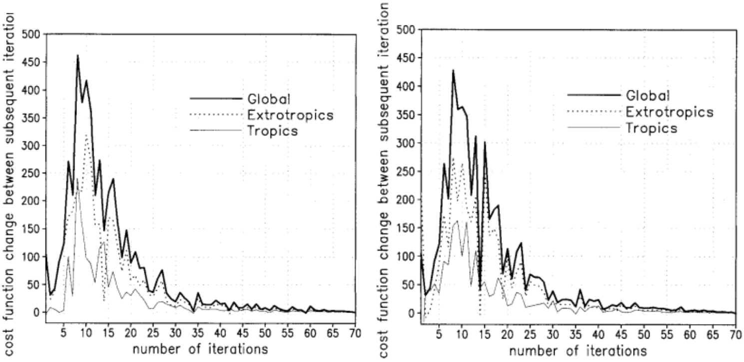


Fig.9 The cost function changes for each minimization iteration for the incremental 4D-Var(left) and standard 4D-Var (right). They are defined as the differences of values of the standard cost function between two subsequent iterations k and k+1.

Incremental 4-D Var summary

- The connection between Conjugate Gradient and Lanczos algorithm can be exploited to:
- Minimize the cost function
- Compute the eigenvectors and eigenvalues of the Hessian at negligible extra cost.
- Preconditioning the minimization for the following outer loop.
- Leads to multi-incremental algorithm
- It is possible because 4D-Var eigenvectors are large scale.
- The consequence is a more efficient and more robust 4D-Var minimization w.r.t. observation amounts and distribution.

Data Assimilation of lightning

- Compared to other types of satellite-derived data, assimilating lightning data into operational numerical models has received relatively little attention.
- NASA will launch in 2015 the GOES-R Lightning Mapper (GLM) that will provide continuous, full disc, high resolution total lightning (IC + CG) data.
- Previous efforts of lightning assimilation mostly have employed nudging.
- We developed a more sophisticated approach involving 3D Var and 1D+3D Var toward which both NCEP and NRL are moving.
- The early stages of our research will utilize existing ground-based lightning data that can be assimilated prior to the launch of GLM; later phases will utilize GLM proxy data that will mimic what GLM will detect.

Data Assimilation of lightning

- A major difficulty associated with this exercise is the complexity of the observation operator defining the model equivalent of the lightning.
- It is using Convective Available Potential Energy (CAPE) as a proxy between lightning data and model variables. This operator is highly nonlinear.
- Marecal and Mahfouf (2003) have shown that nonlinearities can prevent a direct assimilation of rainfall rates in the ECMWF 4D-VAR (using the incremental formulation proposed by Courtier et al. (1994)) from being successful.

Data Assimilation of lightning

- By adjusting the temperature lapse rate, we directly assimilate WTLN (Worldwide Total Lightning Network) total lightning data for the 2011 Tuscaloosa, AL tornado outbreak in a domain of 129×155 km^2 with a mesh resolution of 9 km in each horizontal direction and 60 vertical levels.
- Next we developed a new scheme similar to the one outlined in Marecal and Mahfouf (2002, 2003) i.e., use 1-D VAR to adjust rainfall rate from the moist physics (mass-flux convection scheme and large-scale condensation) closer to an observed value as in Mahfouf et al. (2005).
- The 1D-VAR vertical temperature columns retrievals are considered as new observations and are assimilated in the 3D-Var system .
- It minimizes the problem that nonlinearities of the moist convective scheme can introduce discontinuities in the cost function between inner and outer loops of the incremental 3D Var minimization.

Data Assimilation of lightning

- For WRFDA to assimilate lightning our choice depends on the horizontal resolution of the WRF model. At $9km$ resolution we can calculate flash rates based on ice fluxes. These are the approaches described by Barthe, C et al 2010. Thus we want to exploit the strong linear correlation between the maximum vertical velocity and the total flash rate:

$$f = 5 \cdot 10^{-7}(0.677 w_{max} - 17.286)^{4.55},$$

where f is the total flash rate and w_{max} the maximum vertical velocity.

Data Assimilation of lightning

- So we link the maximum vertical velocity to the lightning flash rate and then translate it in temperature lapse rate using CAPE (Convective available potential energy)
- $$w_{max} = \sqrt{2 \cdot CAPE},$$

  according with parcel theory, so we have to adjust for entrainment.

  $$\mathrm{CAPE} = \int_{z_f}^{z_n} g \frac{T_{\mathrm{parcel}} - T_{\mathrm{env}}}{T_{\mathrm{env}}} dz,$$

  with T_{parcel} the virtual temperature of the specific air parcel, T_{env} the environment temperature, z_f and z_n the heights of free convection and that of equilibrium (neutral buoyancy). Thus, our operator has the following form

  $$H(X) = 5 \cdot 10^{-7}(0.677\sqrt{2 \cdot CAPE(X)} - 17.286)^{4.55}.$$

Data Assimilation of lightning

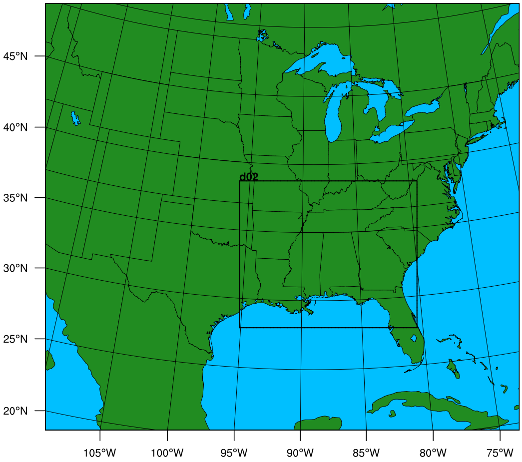


Fig.10 Tuscaloosa/Alabama storm 04/27/2011, Nested domains

Data Assimilation of lightning

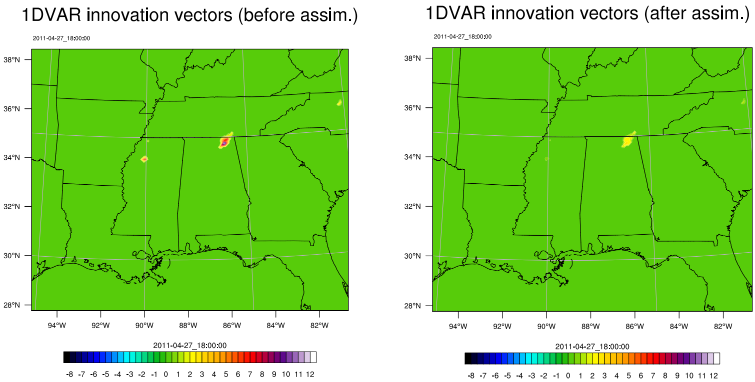


Fig.11 1DVAR innovation vectors before and after lightning assimilation. The after flash assimilation innovation vectors are calculated using the proxies profiles of temperatures

Data Assimilation of lightning

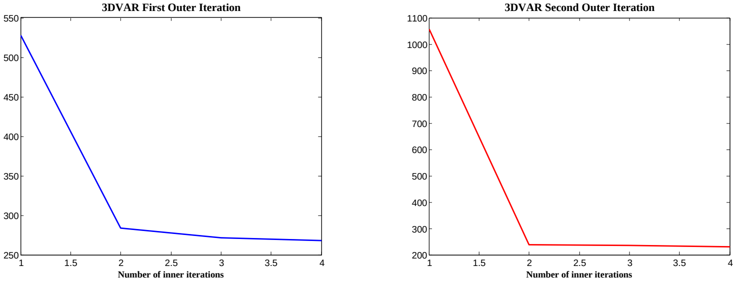
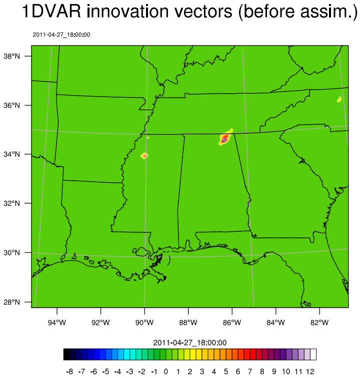


Fig.12 3D-Var direct assimilation of lightning. The cost functional minimization.

Data Assimilation of lightning



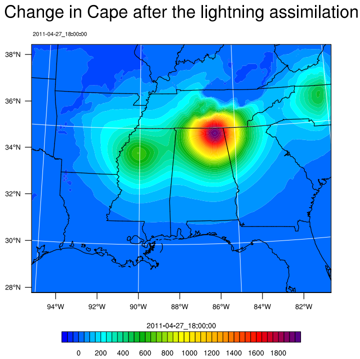


Fig.13 Change in CAPE before and after 3D-Var direct assimilation of lightning.

I.M. Navon, R. Ştefănescu

Data Assimilation of lightning

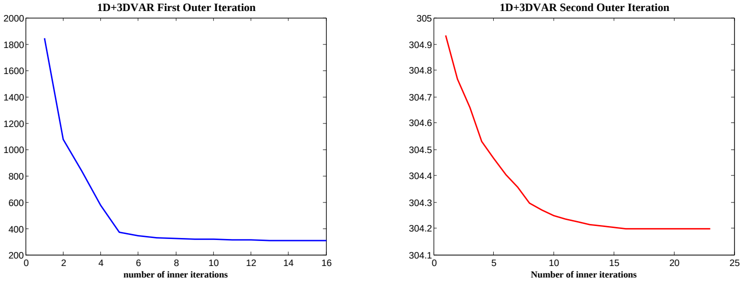


Fig.14 1D+3D-Var assimilation of lightning. The cost functional minimization.

I.M. Navon, R. Ştefănescu

Data Assimilation of lightning

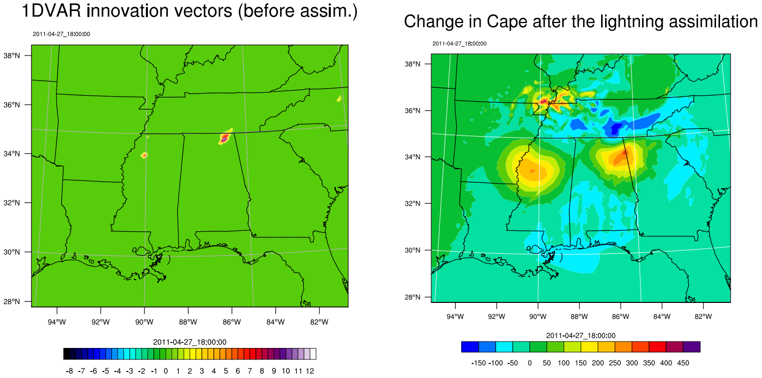


Fig.15 Change in CAPE before and after 1D+3D-Var assimilation of lightning.

Data Assimilation of lightning

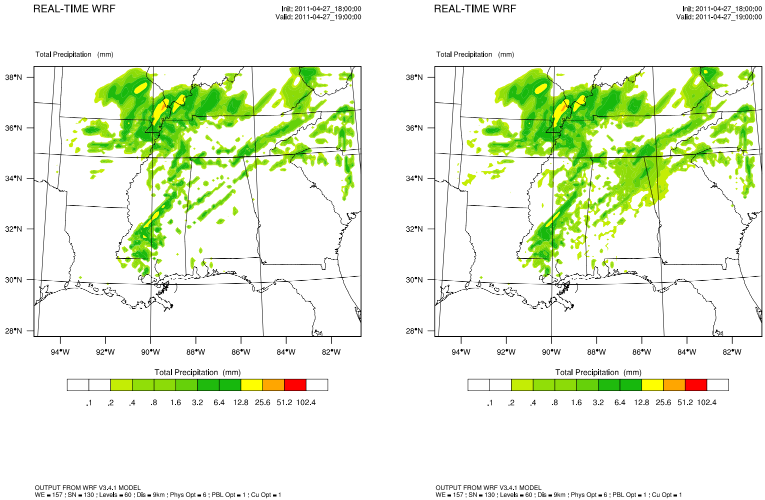


Fig.16 1 hour forecast of precipitation tendencies: Control (left), 1D+3D-Var (right)

Data Assimilation of lightning. Quantitative results

Change in CAPE at analysis time

	3D-Var	1D+3D-Var
Min	-114	-192
Max	1960	475

RMSE of 3D Var, 1D+3D Var and control 6 hour forecast states

Domain	Control	3D-Var	Percentage reduction	1D+3D-Var	Percentage reduction
Full	1.10515	1.10672	+0.14%	1.10236	-0.25%
Lightning	0.06949	0.06477	-6.79%	0.05312	-17.986%

Data Assimilation of lightning

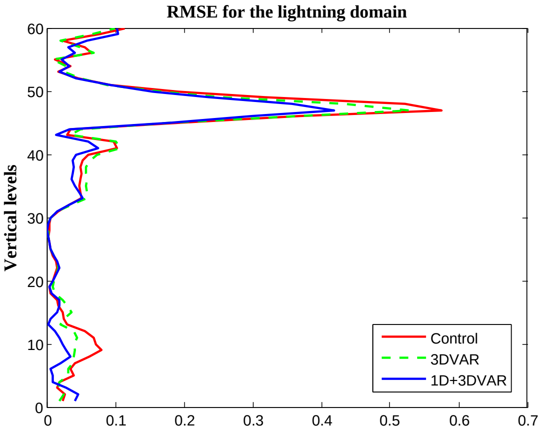


Fig.17 Vertical distribution of root mean square error differences for the temperature field between the true state and the control, 3DVar and 1D+3DVar 6 hours forecasts at 04/28/2012 00:00. The troposphere is covered by the first 34 vertical levels.

Nudging Method

- Nudging is a four-dimensional data assimilation (NDA) method that uses dynamical relaxation to adjust toward observations (observation nudging) or toward an analysis (analysis nudging).
- Nudging is accomplished through the inclusion of a forcing term in the model dynamics, with a tunable coefficient that represents the relaxation time scale. Computationally inexpensive, nudging is based on both heuristic and physical considerations.

I.M. Navon, R. Ştefănescu

Nudging Method

- The NDA method relaxes the model state toward the observations during the assimilation period by adding a non-physical diffusive-type term to the model equations.
- The nudging terms are defined as the difference between the observation and model solution multiplied by a nudging coefficient.
- The size of this coefficient is chosen by numerical experimentation so as to keep the nudging terms small in comparison to the dominating forcing terms in the governing equations, in order to avoid the rebounding effect that slows down the assimilation process, yet large enough to impact the simulation.

Adjoint parameter estimation of optimal nudging coefficients

- The application of the variational approach to determine model parameters is conceptually similar to that of determining the initial conditions.
- For the parameter estimation of the nudging coefficients, the cost function J can be defined as

$$J(G) = \int_0^T \langle W(X - X^o), X - X^o \rangle \, dt + \int_0^T \left\langle K(G - \hat{G}), G - \hat{G} \right\rangle dt,$$

where $\hat{G}$ denotes the estimated nudging coefficients and W and K are specified weighting matrices.

Parameter estimation of optimal nudging coefficients

- The second term plays a double role. On one hand it ensures that the new value of the nudging parameters is not too far away from the estimated quantity. On the other hand it enhances the convexity of the cost function since this term contributes a positive term K to the Hessian matrix of J.
- An optimal Nudging data assimilation (NDA) procedure can be defined by the optimal nudging coefficients G^* such that

$$J(G^*) \leq J(G), \quad \forall G.$$

Parameter estimation of optimal nudging coefficients

- The problem of extracting the dynamical state from observations is now identified as the mathematical problem of finding initial conditions or external forcing parameters that minimize the cost function.
- Due to the dynamical coupling of the state variables to the forcing parameters, the dynamics can be enforced through the use of a Lagrange function constructed by appending the model equations to the cost function as constraints in order to avoid the repeated application of the chain rule when differentiating the cost function. The Lagrange function is defined by

$$L(X, G, P) = J + < P, \frac{\partial X}{\partial t} - F(X) - G(X^o - X) >$$

where P is a vector of Lagrange multipliers.

Parameter estimation of optimal nudging coefficients

- The Lagrange multipliers are not specified but computed in determining the best fit.
- The gradient of the Lagrange function must be zero at the minimum point. The results are coupled in the following first order conditions:

$$\frac{\partial L}{\partial X} = 0 \quad \sim \quad \text{adjoint model forced by } \; 2\mathbf{W}(X - X^o)$$

$$\frac{\partial L}{\partial P} = 0 \quad \sim \quad \text{direct model}$$

$$\frac{\partial L}{\partial G} = 0 \quad \sim \quad \int < P, -\frac{\partial\left(F(X) + G(X^o - X)\right)}{\partial G} > dt + 2K(G - \hat{G}) = 0$$

Parameter estimation of optimal nudging coefficients

- The solution of previous equations is called a stationary point of L.
- Even if the dynamical evolution operator is nonlinear, the equations $\left(\frac{\partial L}{\partial X} = 0\right)$ will be the same as those derived by constructing the adjoint of the linear tangent operator;
- The linearization is automatic due to the Lagrange function L being linear in terms of the Lagrange multipliers P.

Parameter estimation of optimal nudging coefficients

- An important relation between the gradient of the cost function with respect to parameters G and the partial derivative of the Lagrange function with respect to the parameters is

$$\nabla_G J(G) = \frac{\partial L}{\partial G}\Big|_{\text{at stationary point}},$$

i.e. the gradient of the cost function with respect to the parameters is equal to the left hand side of third equation in the first order conditions system which can be obtained in a procedure where the model state P is calculated by integrating the direct model forward and then integrating the adjoint model backwards in time with the Lagrange multipliers as adjoint variables.

Parameter estimation of optimal nudging coefficients

- Using this procedure we can derive the following expressions of the adjoint equation and gradient formulation.

$$\begin{aligned} \frac{\partial P}{\partial t} + \left[\frac{\partial F}{\partial X}\right]^T P - G^T P &= W(X - X^o) \\ P(T) &= 0 \end{aligned}$$

and

$$\nabla_G J = -\int_0^T \langle (X^o - X), P \rangle \, dt + 2K(G - \hat{G}).$$

Parameter estimation of optimal nudging coefficients

- We see that the adjoint equation of a model with a nudging term added is the same as that without a nudging term except for the additional term $-G^T P$ added to the left hand side of the adjoint equation.
- Having obtained the value of cost function by integrating the model forward and the value of the gradient $\nabla_G J$ by integrating the adjoint equation backwards in time, any large-scale unconstrained minimization method can be employed to minimize the cost function and obtain an optimal parameter estimation.

Parameter estimation of optimal nudging coefficients

- If both the initial condition and the parameter are controlled, the gradient of the cost function for performing the minimization would be

$$\nabla J = (\nabla_U J, \nabla_G J)^T$$

where

$$\nabla_U J = -P(0).$$

Parameter estimation of optimal nudging coefficients

- Zou, Navon and Le Dimet (1992) have shown that estimated NDA, optimal NDA and Kalman filter (KF) differ from each other in the choice of the weight matrix often called the gain matrix:

$$G_n^* \equiv W_n^f H_n^T \left(H_n W_n^f H_n^T + R_n \right)^T .$$

- The VDA, on the other hand, takes both the model forecasts and the observations as perfect. It attempts to obtain an optimal initial condition which minimizes the cost function

$$J^f = E \left(X_n^f - X_n^o \right)^T \left(X_n^f - X_n^o \right) .$$

Parameter estimation of optimal nudging coefficients

- The theoretical framework of estimation and control theory provides the foundation of data assimilation techniques.
- The estimated NDA and the KF are closer to the estimation theory, the VDA to the optimal control aspect while optimal NDA is a combination of both (see also LORENC (1986)).
- See also work of Vidard, Piacentini and LeDimet (2003) on the optimal estimation of nudging coefficients.

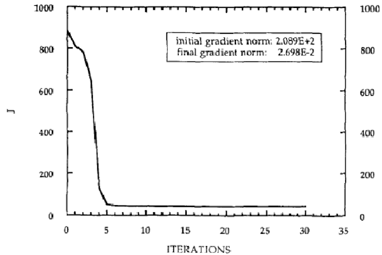


Figure 1. Variation of the cost function, $\mathcal{J}$, with the number of iterations. Both the initial conditions and the nudging coefficients, $G_{\ln p_s}$, G_D and G_ζ, serve as control variables. The norms of initial gradient, $\|g_0\|$, and final gradient, $\|g\|$, (30 iterations) are also given. The length of the assimilation window is 6 hours.

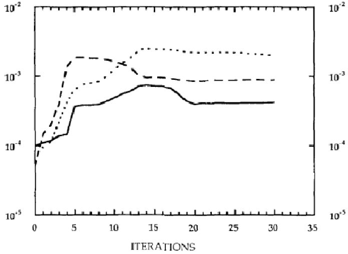


Figure 2. Variation of the nudging coefficients, $G_{\ln p_s}$ (——), G_D (– – –) and G_ζ (. . .), with the number of iterations. Both the initial conditions and the nudging coefficients $G_{\ln p_s}$, G_D and G_ζ serve as control variables.

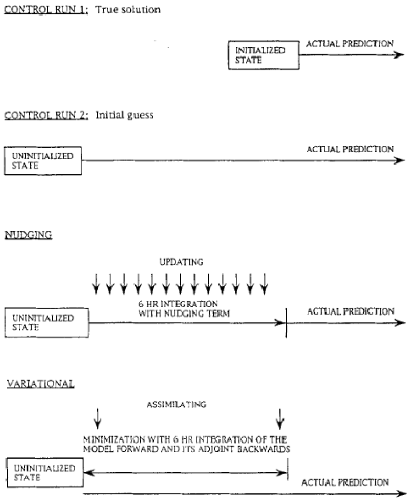


Figure 5. A schematic view of the assimilation-forecast cycle.

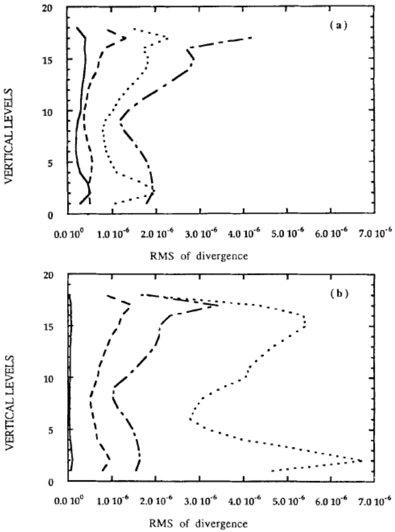


Figure 6. Distributions of the root-mean-square differences for the divergence field between the true solution and the assimilated fields at the end of the (a) 6-hour and (b) 12-hour assimilation intervals. Results are shown for different assimilation experiments: control run (—·—), estimated nudging (· · ·), variational data assimilation (- - -) and optimal nudging (—). The surface pressure, divergence and vorticity fields are nudged.

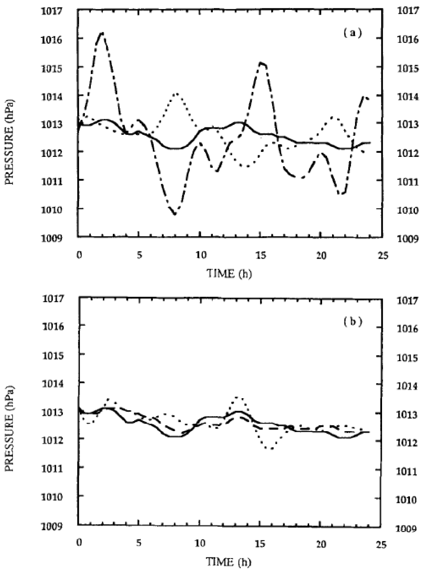


Figure 8. Same as Fig. 7 except the length of the assimilation window is 12 hours.

4-D Var issues

- Observations, preprocessing, quality control
- $H-$ new observation types (lightning), improving operators
- R - observational errors, tuning, correlated observational errors
- X_b - improve models
- $B-$Statistical models, flow dependent features minimization algorithms (Quasi-Newton, CG)
- Structure of model error covariance
- Weak constraint 4-D Var allows perfect assumption to be removed and use of lower assimilation time windows
- The statistical description of model error is one of the main current challenges in data assimilation.

- Forecast sensitivity to the R- and B-specification
- Adaptive sensor location using singular vectors and adjoint sensitivity
- Hybrid Variational/Ensemble Data Assimilation
- 1-D + 4-D VAR data assimilation of rainfall affected radiative data