Variational Data Assimilation

ALATNET Seminar on Data Assimilation

Gourdon/France, 11th-22nd June 2001

Claude Fischer/András Horányi/Wafaa Sadiki

Toulouse/Budapest/Casablanca

Contents

- General aspects: motivations, history, general formulation of the cost function
- Practical Implementation: where the problems start and where refinements are needed
- Limited Area and Lateral Boundary Coupling

I. General aspects

Motivations

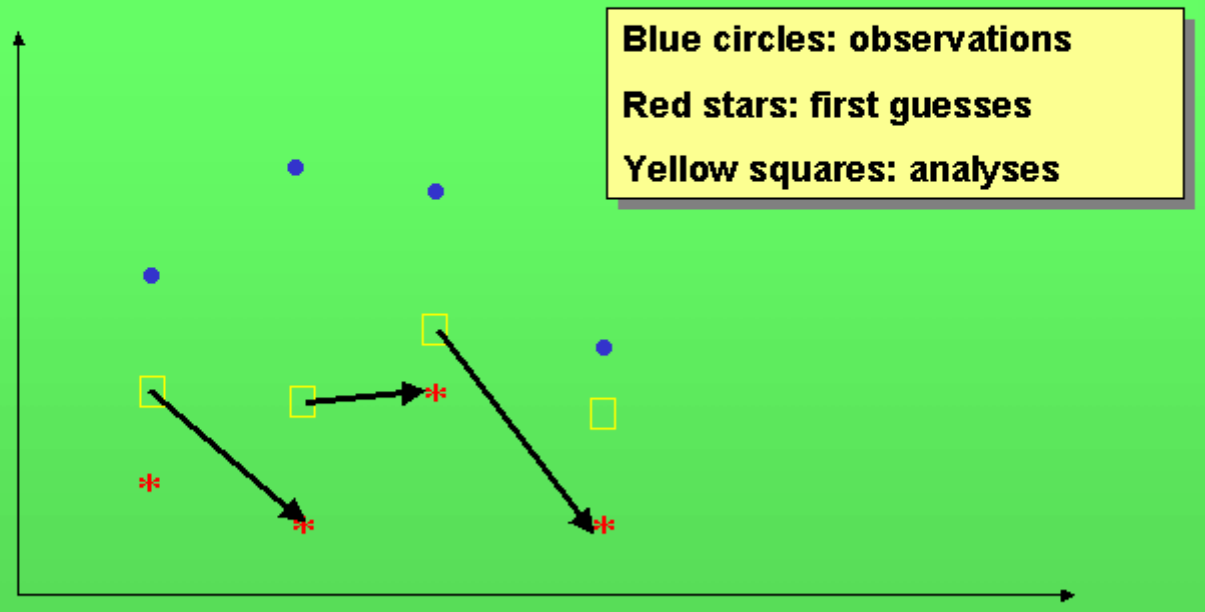
History: from local interpolation to global/variational

Cost functions in 3D-VAR and 4D-VAR

What is data assimilation ?

- The goal is to produce an atmospheric state as close as possible to the reality and at the same time dynamically consistent, taking into account all the available information: observed data, model, physical constraints, climatology
- The tools: any type of objective (as opposed to subjective) data analysis: Optimal Interpolation, three-dimensional and/or four-dimensional variational assimilation (3D-VAR/4D-VAR), Kalman filter. Other approaches such as Nudging won't be discussed in this lecture.
- The output is a set of meteorological fields on the model 's geometry (e.g. a geometrically regular grid or spectral coefficients, etc.) – the Analysis.
- Important aspect: cycling i.e. process of permanent assimilation of data

Example for cycling: sequential data assimilation



A quick historical overview

- Empirical techniques such as succesive correction (Cressman, 1959): the weights Wi given to the different observations to analyse a point Aj are computed by using only an empirical function of the distance between Aj and the observations Oi.

- Multivariate Optimal Interpolation (Gandin, 1963, Rutherford, 1972, Lorenc, 1981): based on least square linear statistical estimation theory, this technique becomes popular in the late 70ies. At that time, the number of observations starts to increase dramatically, and the types of observations become much more various. The use of statistical models to deal with the characteristics of model and observation errors allows OI to treat the more and more heterogeneous observations

Historical overview (2)

- Variational techniques: A cost function is defined and the solution is achieved through the minimisation of this cost function, in practice, the implementation of OI necessitates a number of approximations, such as the reduction to a gridpoint local objective analysis. In the middle of the 80ies, the *adjoint technique* is recognized as a powerful tool for computing efficiently one main ingredient of Var: the gradient of the cost function J(X) with respect to X (Le Dimet & Talagrand, 1986, Lewis & Derber, 1985)
- Other tools needed for Var become available: cost and memory efficient minimization algorithms, dramatic increase of the computers' incore memory
- Late 80ies: A new breath for OI !: the Kalman filter (Kalman, 1960, Kalman & Bucy, 1961, Ghil, 1989, Gauthier, Courtier & Moll, 1993) = OI with an explicit time evolution of the error statistics

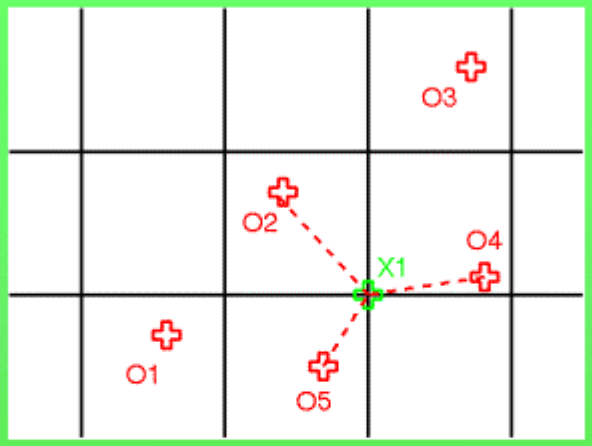
Motivations/Advantages for Variational

- Solve a high-dimensional minimization problem: no explicit computation of the whole system is possible.
- Use a large number of various observation types (satellites, surface, sondes, radar …).
- The operators that link the model variables to the observations need not be linear in 3D(4D)-VAR.
- Treat all the (model) control vector X in one go – 3D-VAR –> « global »
- Keep a dynamical consistency between analysis and model – 4D-VAR – reducing the spin-up effects.
- Respect the time stamps of the observations – 4D-VAR -.

From Optimal Interpolation to 3DVAR

Optimal Interpolation is based on simplifying assumptions: compute model variables locally, use a limited set of observed data, use only linear observation operators (Hi(Xj) -> Hi.Xj).

Aj = Xj + Σi Wi.(Hi(Xj)-Oi)



The move from a « local », low-dimensional problem to a global, variational problem of 3D-VAR type: use powerful minimization algorithms based on descent strategies (Le Dimet-Talagrand, 1986), use the adjoint technique to compute the needed gradients, take benefit of the increased incore memory capacities of « modern » computers.

Variational analysis: a conditional probability problem

- State vector with dimension Nx: X; Observations with dimension Ny: Y
- The best estimate of Y is H(X), where H is the observation operator (might be non-linear)
- Analysis: inverse problem Y -----> X

$$P_a(X) = P(X = X_t \mid Y = Y_o) \propto P(Y = Y_o \mid X = X_t) P(X = X_t)$$

- Bayes' theorem, where subscript „t" refers to the truth and subscript „o" to the observations
- Pa(X) is a probability density function with dimension Nx
- The best estimate: minimum variance, maximum likelihood, etc.

Variational analysis: a conditional probability problem (2)

- Taking into account the background information:

$$P(X = X_t) = P_b(X - X_b)$$

- The observation also contains errors:

$$P(Y = Y_o \mid Y_1 = Y_t) = P_o(Y_o - Y_t)$$

$$P(Y = Y_o \mid X = X_t) = P_{of}(Y_o - Y_f), \text{ where } Y_f = H(X_t)$$

- The observation operator also contains errors:

$$P(Y_1 = Y_t \mid X = X_t) = P_f(Y_1 - Y_f)$$

Variational analysis: a conditional probability problem/Representativeness error (3)

- Sources: every source of error other than directly coming from observation measurement errors
  - Examples
    - grid interpolations
    - truncation errors
    - observation operator errors due to
      - » simplifications in the operators
      - » Usage of such variables, which have their own error (e.g. surface temperature from first guess)
- **Practically:** representativeness error is embedded in the general observation standard deviation ("sigma_o")

Variational analysis: a conditional probability problem (4)

- Substituting all this into the original conditional probability:

$$P_a(X) \propto \left[\iint P_o(Y_1 - Y_o) P_f(Y_t - H(X) dY_1\right] P_b(X - X_b) = P_{of}(Y_o - Y_f) P_b(X - X_b)$$

  - For the analysis one needs the 3 probability density function: Pb, Po, Pf respectively
  - If these are Gaussian functions (Lorenc, 1988):

$$P_b(X - X_b) \propto \exp(-\frac{1}{2}(X - X_b)^T B^{-1}(X - X_b))$$

$$P_o(Y - Y_o) \propto \exp(-\frac{1}{2}(Y - Y_o)^T O^{-1}(Y - Y_o))$$

$$P_f(Y - Y_f) \propto \exp(-\frac{1}{2}(Y - Y_f)^T F^{-1}(Y - Y_f))$$

$$P_{of}(Y_o - Y_f) \propto \exp(-\frac{1}{2}(Y_o - Y_f)^T (O + F)^{-1}(Y - Y_f))$$

Variational analysis: a conditional probability problem (5)

- For maximising this probability one has to ensure the minimum of:

$$J = (Y_o - H(X))^T (O + F)^{-1} (Y_o - H(X)) + (X - X_b)^T B^{-1} (X - X_b)$$

  - Cost function: Jo + Jb
  - R =O + F

More details about the cost function

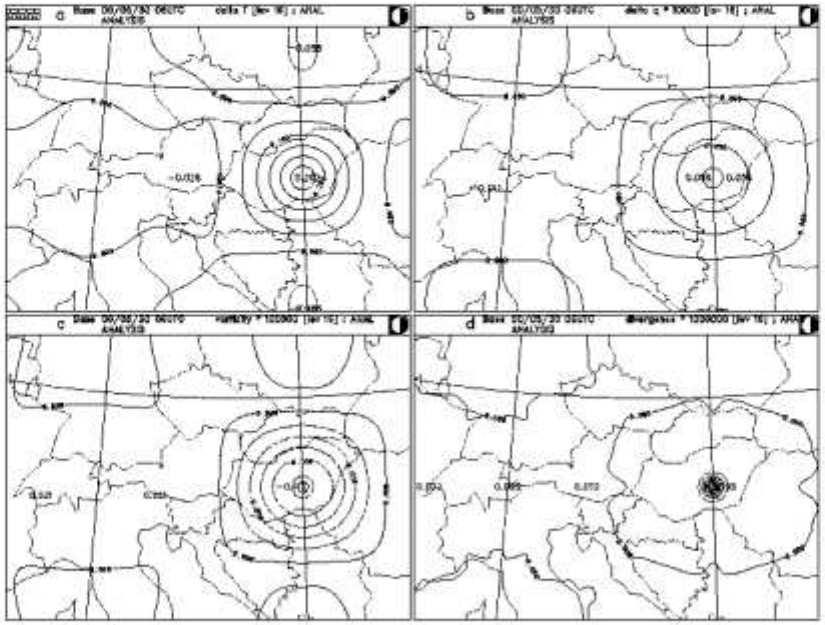
- J(X)=Jb+Jo=« distance to the background » + « distance to the observations »
- In 3DVAR, on the contrary to OI, the obs operator H need not be linear.
- There can be an extra cost function, Jc, *weak constraint term*, that forces the minimizing solution to respect some balance properties (eg, normal mode tendencies close to 0 or digital filter initialization, Gauthier *et al.*, 1999 / see the lecture by D. Giard). An alternative is to specify dynamical balance constraints directly in the Jb term, *strong constraint* approach.

Specification of the error covariance matrices

- Pb = Background error covariance matrix, computed by comparison to observations (Hollingsworth-Lönnberg, 1986), sampling over model forecasts (Parrish-Derber, 1992) in the so-called NMC method, stochastic Kalman filter (Bouttier, 1994, Fischer et al., 1998)

- Pb (and R as well): the specification of the covariance matrices controls directly the structure functions (see figure) of the analysis, that is the way the information extracted from the observations will spread spatially in the vicinity of each obs.

Structure functions/single obs. experiment



Specification of the error covariance matrices (2)

- To be *statistically* consistent with the properties of the innovation vector (di), Pb and R should reflect the actual error covariances of the innovation:

$$R + H'^{*}_{i} P_b H'_i = E[d_i^t . d_i]; \quad d_i = Y_i - H_i(X_b)$$

(See Wafaa 's advanced lecture on « a posteriori validation »)

Specification of the error covariance matrices (3)

- R = Observation error covariance matrix, split into block-diagonals for each observing system. The use of a growing number of new types of observations imposes to take into account vertical error correlations (eg, satellite or GPS data).

- Don't forget !: R contains both the error of measurement and the error of representativeness

- The structures of the *actually* used Pb and R need to be simple enough to allow the computation of their inverse.

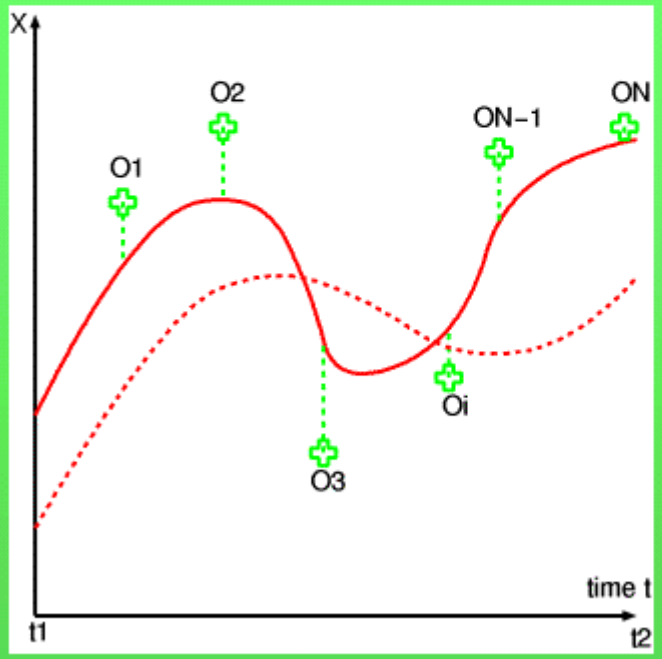
4D-Variational assimilation: find X(t1) such as the green dashed « distance » is minimum

Solid red: model trajectory from the minimizing solution X(t1)

Dashed red: initial model forecast

Green crosses: observations over the [t1,t2] assimilation window

Observations: direct measurements of elements of X, indirect ones, model dynamical constraints (distance to geostrophism)



4D-VAR cost function

In 4D-VAR, the distance to the observations is computed by Transporting all the innovations throughout the assimilation Window via the direct model integration:

$$J(X)=(X-Xb)^t.P_b^{-1}.(X-Xb)+$$
$$\sum_{i=0}^{M}(y_i-H(X_i))^t.R^{-1}.(y_i-H(X_i))$$

Observations are in practice sampled over M time slots.
In 3D-VAR, i=M=0 (one single slot).

4D-VAR cost function (2)

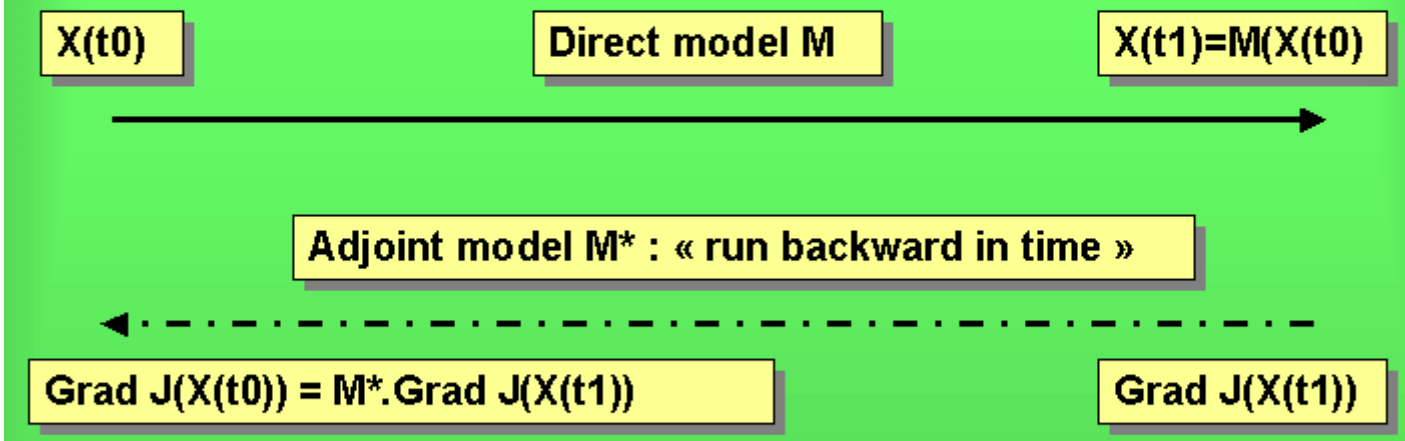
The minimization algorithms will need the gradient of J(X)
With respect to X:

$$\nabla_X J = (X - X_b^0)^t . P_b^{-1} + \sum_{i=0}^{M} (y_i - H(X_i))^t . R^{-1}$$

The adjoint model integration is needed to compute the gradient of the departure between yi and H(Xi) at the observation times(=time slots) starting with Xs for example:

$$(y_i - H(X_i))^t \approx (y_i - H(X_b^i))^t - (X_s - X_b^0)^t . M_{0,i}^{*} . H'^t$$

Gradient: use the adjoint model



Equivalences / Advantages of OI and (3D/4D)VAR

- If *all* observations are used, and if Jb and Jo are kept the same, that is OI and 3D-VAR use the very same covariance matrices Pb and R, then OI=3D-VAR. On the contrary, OI#3D-VAR if only a local subset of the observations is analyzed in OI.
- If the forecast model is *linear* and does not possess *model errors,* then 4D-VAR=Kalman filter.
- Advantages of VAR: 3D and 4D can be developped as one single code, there exist wide possibilities on the specification of Jb, observation operators H need not be linear (for the innovation).
- Advantages of OI: provides so far the only tractable assimilation method for surface fields

Preconditioning

- Necessary for speeding up the minimization process (Courtier et al., 1994)

- Computation of the second derivative of the cost function (the Hessian)

In practice, most variational algorithms use the change of variables that makes Jb trivial:

$$J_b(\delta X) = \delta X^t . P_b^{-1} . \delta X \Rightarrow J_b(\chi) = \chi^t . \chi$$

$$\chi = P_b^{-1/2} . \delta X$$

Minimization algorithms

Most algorithms use an iterative method:

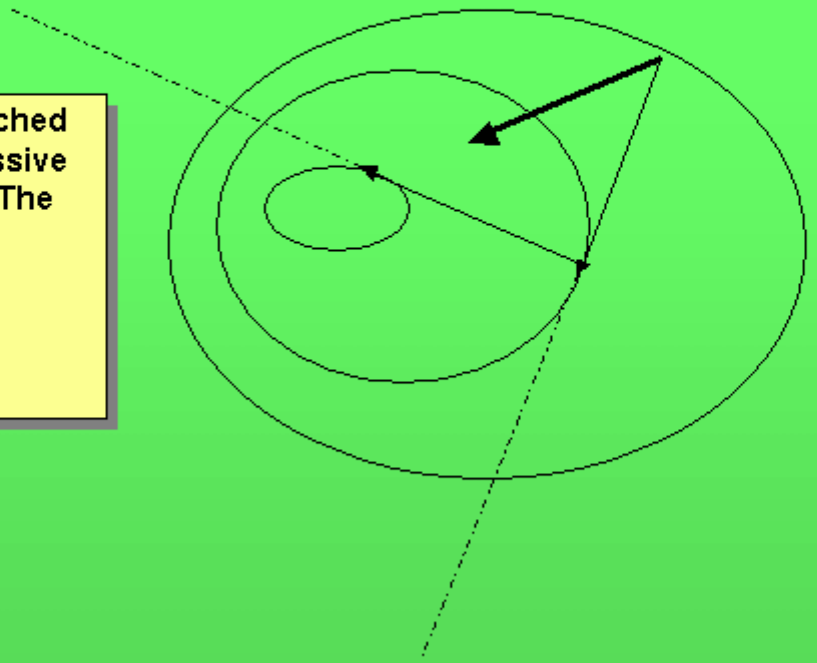
1. Choose the starting point (Xi=X0)
2. Compute the cost function J(Xi) and its gradient Grad(J(Xi)) with respect to X
3. Search the new value Xi+1=f(Xi) such as J(Xi+1)<J(Xi)

Steps 2 and 3 are repeated until a convergence criterion is fulfilled. Different algorithms exist:

Steepest descent, conjugate gradient, quasi-Newton

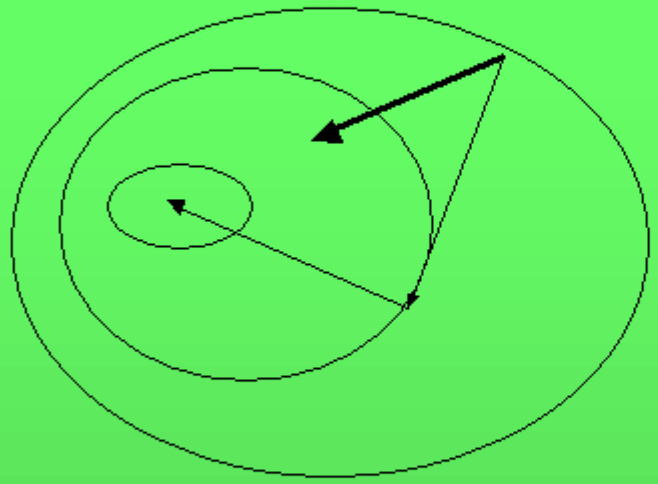
Steepest descent

The minimum is reached following the successive gradient directions. The optimal stepsize is adjusted so as to descend along the tangent for the next iteration.



Conjugate gradient

The minimum is reached following successive, orthogonal, directions. Thus, in a space of dimension N, N iterations are sure to give the result. Here, 2 loops are enough.



Quasi-Newton method

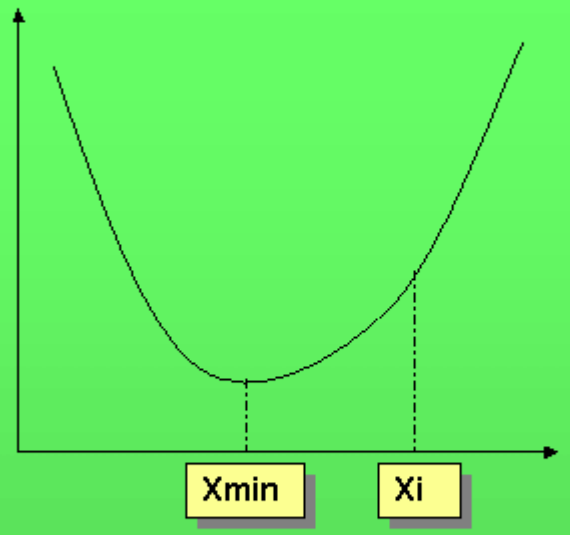
The minimum is reached by successive application of a Taylor formula of order 1. Also, we state that, at the minimum, the gradient must be 0:

$J'(X_{min}) = 0$

$J'(X_{i+1}) = J'(X_i) + (X_{i+1} - X_i)\, J''(X_i)$

Where J'' is the second-order derivative of the cost function (=the Hessian matrix).

$X_{i+1} = X_i - J'(X_i) / J''(X_i)$



Summary/Generalities

- What is data assimilation?
- From local OI to global variational
- Variational analysis as a conditional probability problem
- Cost function and its gradient in (3D/4D)VAR
- Error covariance matrices (observation, representativeness, background)
- Minimization algorithms

II. Practical implementation

(Multi-)incremental approach

Quasi-continuous approach

The example of Meteo-France 4D-VAR

Flow-dependant structure function

Incremental formulation

Instead of analyzing the full model variable, only analyze its departure from a « reliable » reference state. In practice: X is replaced by $\delta X = X - Xb$ (it ensures automatically that the estimation becomes unbiased).

A new set of assumptions is then derived to write the variational problem (Courtier et al., 1994):

- model forecast: $M(X) \approx M(Xb) + Ls.\delta X$
- observation operators: $H(X) \approx H(Xb) + H'.\ \delta X$

Ls is a *simplified* tangent linear version of the full model M.

Advantages of the incremental formulation

- Minimizing the cost function at full forecast resolution is out of the bounds of the computational possibilities. Thus, the variational problem is solved at lower resolution, and, for a global model, the emphasis is put on the analysis of the planetary/synoptic/sub-synoptic weather elements
- Cost function J(X) is kept quadratic with respect to the control variable δX
- Compute the innovations (obs minus background) with the full, possibly non-linear observation operator H(.)
- « Simplifications »

« Simplification » actions

Simplifications can be of various types:

- lower resolution
- simpler (and smoother) physics (Janisková et al., 1999)
- different advection scheme in Ls than in M
- filtered increments

Both M(.) and H(.) are linearized with respect to X, and their tangent linear operators are computed in the vicinity of Xb.

Incremental cost function

$$J(\delta X) = \chi^t . \chi +$$

$$\sum_{i=0}^{M} (d_i - H_i' . \delta X_i)^t . R^{-1} . (d_i - H_i' . \delta X_i) + J_c$$

with $\chi = P_b^{-1/2} . \delta X_0$ $\qquad d_i = y - H(X_i)$

and $(d_i - H_i' . \delta X_i)^t = d_i^t - \delta X_0^t . L^*_{s\,0,M} H_i'^t$

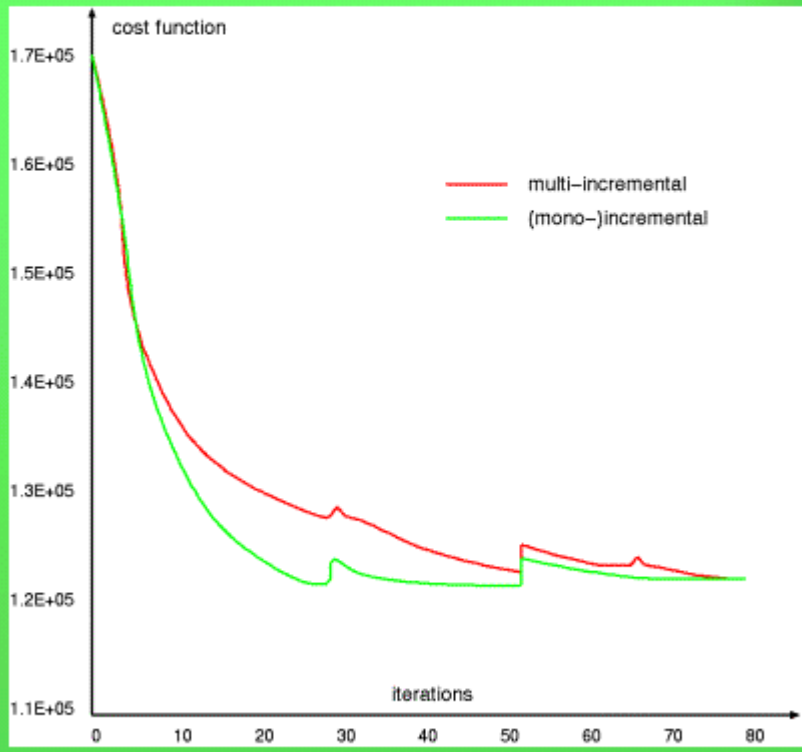
Multi-incremental

The *simplification* and the minimization are performed several times in a loop. Thus, the set of simplifications can be changed and a new reference trajectory, along with a new reference state Xb new, can be computed.

Presently, at Météo-France, 3 refreshments are done: T42 and T63 (with an almost adiabatic adjoint model Ls*) and T95 (with simplified physics, Janisková *et al.*, 1999).

Multi-incremental (2)

This hierarchical approach is cheaper than a (mono-) incremental computation, since the truncations are reduced and the adjoint model is highly *simplified*. The final results are close, because the first iterations of (mono-)incremental 4D-VAR are basically only adapting the large scales.

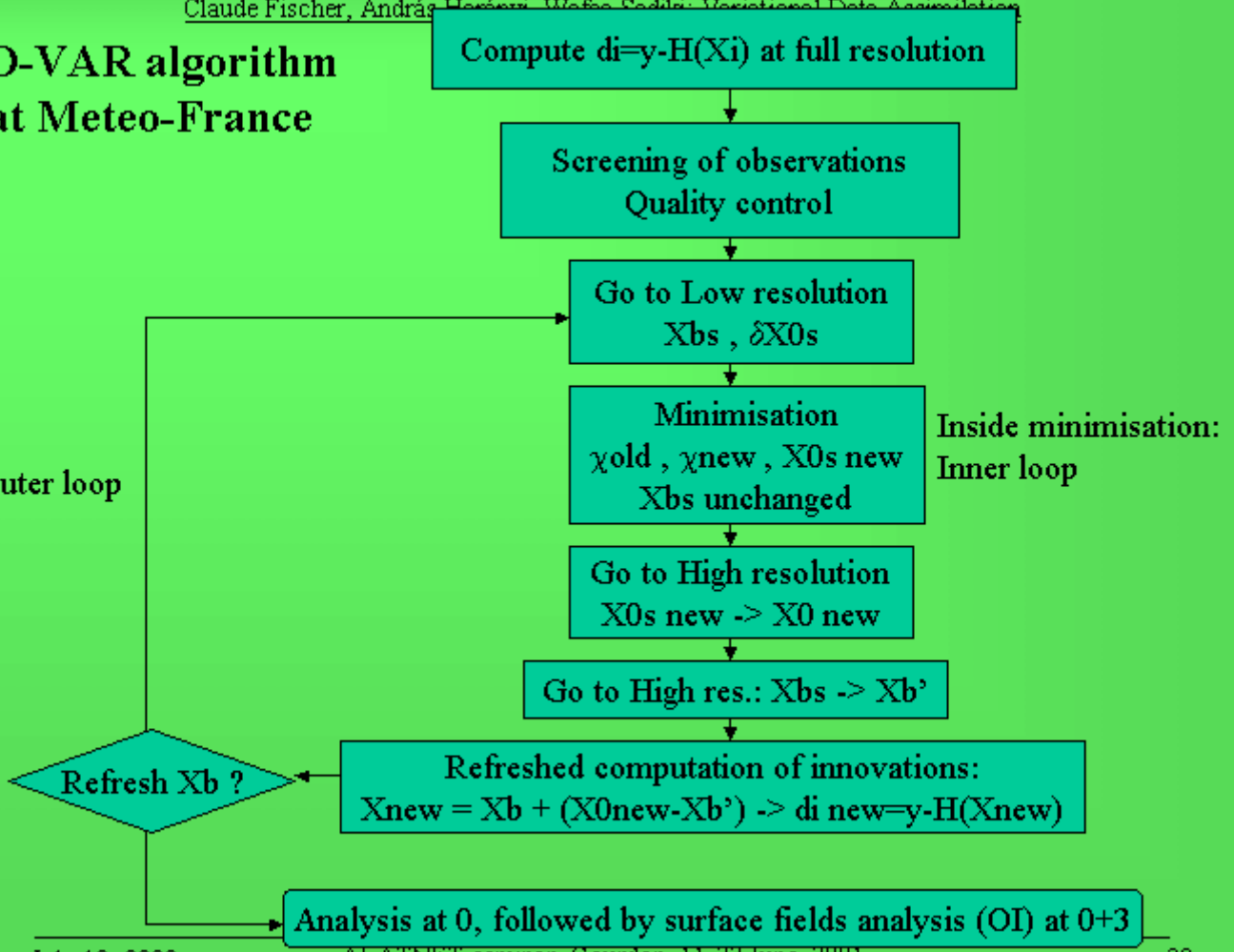


Quasi-continuous

The idea is to split the Jo term into pieces, and perform several successive minimizations.
The split can be motivated by the amount of data or by the large spread in time of the data (eg, minimize first with available data, then wait for new obs and start from the previous analysis as background for the next minimization):

- A1: Min J for J=Jb+Jo1
- A2: Min J for J=Jb+Jo1+Jo2
- An: Min J for J=Jb+$\sum_{i=1,n}$ Joi

4D-VAR algorithm at Meteo-France



Flow-dependant structure functions

- 3DVAR: static approach
  - The structure functions are originating from the background error statistics used
- 4DVAR: dynamic approach
  - Possibility for flow-dependant structure functions (implicit through the transport of observation increments in the TL/AD models in Jo)

Flow-dependant structure functions (2)

- Semi-geostrophic coordinate transform (Desroziers-Lafore, 1996)
- Change of vertical coordinate (in discussion for global variational systems, eg. ECMWF, WRF …)
- Reduced Rank Kalman filter (ECMWF)

Summary/Practice

- (Multi-)incremental formulation
- Quasi-continuous approach
- The operational 4DVAR algorithm at Météo France
- Flow-dependant structure functions

III. Limited Area and Lateral Boundary Coupling

Scale selection and localization

Possible tools

Nested assimilation cycles

Boundaries and coupling in 4D-VAR

Scale selection and localization

What scales to be analysed?

Effect of first guess

Effect of background errors

Effect of DFI

Incremental?

What scales to be analysed?

- Depending on the application: those scales are interesting, which are already not resolved by the forcing model

- Is there a lower limit? One might test VAR systems for meso-γ scales (δx =1 to 5 km) and for processes that are rapidly developing (time scales of 10 or 30 minutes) and non-linear (threshold, saturation) $\Rightarrow$ what is the risk of relying to a partially linear analysis system?

What scales to be analysed? (2)

- Some tools for controlling the analysed scales:
  - » first guess
  - » background error statistics
  - » initialisation
  - » incremental or not?

Scale selection/Effect of first guess

- What first guess to choose?
  - Synoptic scale: short range (6h) forecast of the model
  - Mesoscale:
    - » large scale analysis - most recent data is taken into account
    - » short range (6h) forecast - small scales are adapted from the model (NB: for example a 6h mesoscale forecast would provide useful dynamically adapted small scales, but might have missed some important forecast elements -- e.g. convective clusters…….--)
    - » mixed (blended) fields - advantage of both, but dependence on the blending method

Scale selection/Effect of background errors

- Comparison with observations (Hollingsworth and Lönnberg, 1986)
  - Independent from the assimilation system (+)
  - observation errors (-)
  - observation network can be dense (+) or sparse (-)
- Stochastic-dynamic, Kalman filter methods (Bouttier, 1994, Fischer et al., 1998)
  - approximate model error term (+)
  - non-linear error saturation (-)
  - expensive (-)
  - non-isotropic structure functions (+)

Scale selection/Effect of background errors (2)

- NMC method (Parrish and Derber, 1992)
  - rather simple to implement & link with model dynamics (+)
  - no theoretical justification (but it works) (-)
  - need for re-scaling (-)
  - underestimated variances in data sparse areas and overestimation in data-dense areas (-)
  - possible adaptations to LAM (+)

  (see the talk of Maria Siroka)

Control on the structure functions

- Convolutions of increments (IFS)
- Second Order AutoRegressive (SOAR) functions (UKMO)

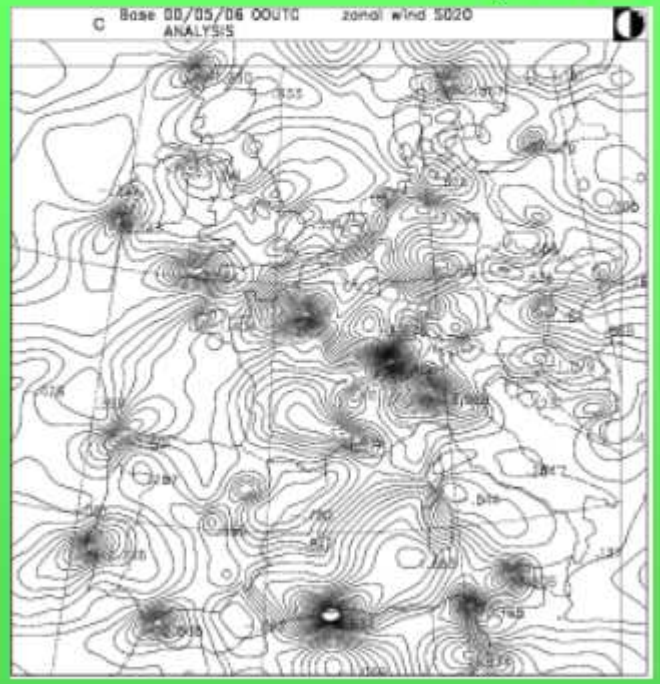
Other aspects

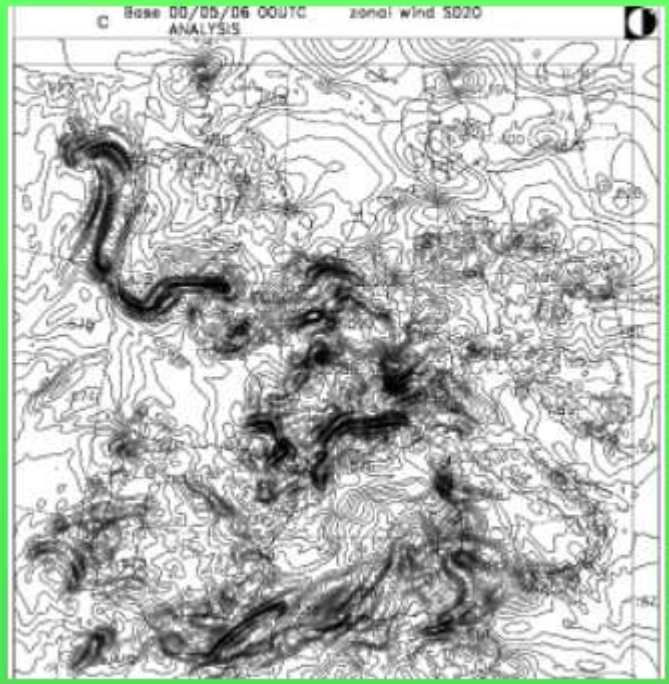
- Relative importance of the different terms in the cost function: ratio of Jo and Jb (re-tuning of sigma_o to take into account the grid projection representativeness error (in a LAM compared to a global model)

  /see the presentation of Wafaa/

- Effect of initialization (see presentations on DFI)
  - DFI might destroy some useful signal

Analysis increments without and with DFI (zonal wind)





Is an incremental approach meaningful for mesoscales or small scales?

- PROS:
  * Cheaper minimization (especially less memory)
  * Possibility to filter out strongly non-linear processes from the adjoint model
- CONS:
  * Miss some important features from the background or the dense observation network
  * The results saying that "large scales converge first" in the minimization might be invalid below some scale due to too important downscale/upscale interactions (e.g. small convective updrafts compensated by large scale subsidence, 3D turbulence)

Possible tools

- 3DVAR:
  * Cheap (+)
  * Probably efficient in terms of "quality" of observations to background fit (+)
  * Analysis only at one moment in time, while high density observations could be also available at high frequency representing rapidly evolving systems (-)

$\Longrightarrow$ At one stage in future, meteorologists probably have to answer to the question: "given the typical time scales of (convective/turbulent) processes, what is the optimal frequency of instantaneous 3DVAR analyses?"

Possible tools (2)

- 4DVAR:
  * Very expensive (-)
  * Raises the problem of lateral boundary conditions (-)
  * Could deal with ageostrophic, large frequency data (+)
  * Provided the non-linearities involved are not critical for the adjoint step (-)
  * For linear and smooth non-linear processes provided dynamically consistent increments (e.g. humidity in ascents and descents) (+)

Possible tools (3)

- ALTERNATIVES:
  - * Kalman filter type of approaches
  - * Forget about variational……..
  - * OI would become an "intelligent interpolation" device
    - * observation operators must be linear
    - * high density of observations is possible…… BUT difficult to implement (cluster of obs. ?)
    - * possibility to adapt the empirical statistical model to the desired spatial ranges

Nested assimilation cycles

Best scenario from global to local data assimilation?

Need for extra data at small scales

Targetted observations

Best scenario/Why nested assimilation?

- Every LAM is driven by a global (or regional) coupling model
- No need to reanalyse those scales already analysed in the global model
- If data assimilation cycles of global & LAM models were uncoupled there would be a risk of the LAM drifting away from the global cycle

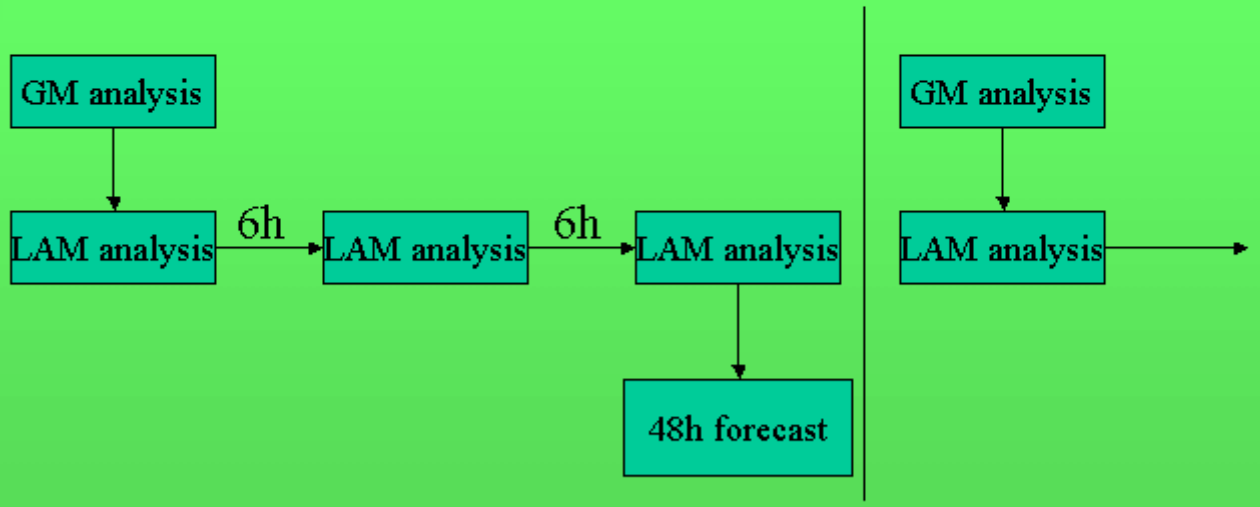
Best scenario/Why nested assimilation? (2)

- Since the global (regional) model provides the large scales for the coupling data, there exists potentially a risk of mismatch between the LAM analysis and the following coupling data

⇒One has very probably to ensure some consistency between the global (regional) & the LAM data assimilation

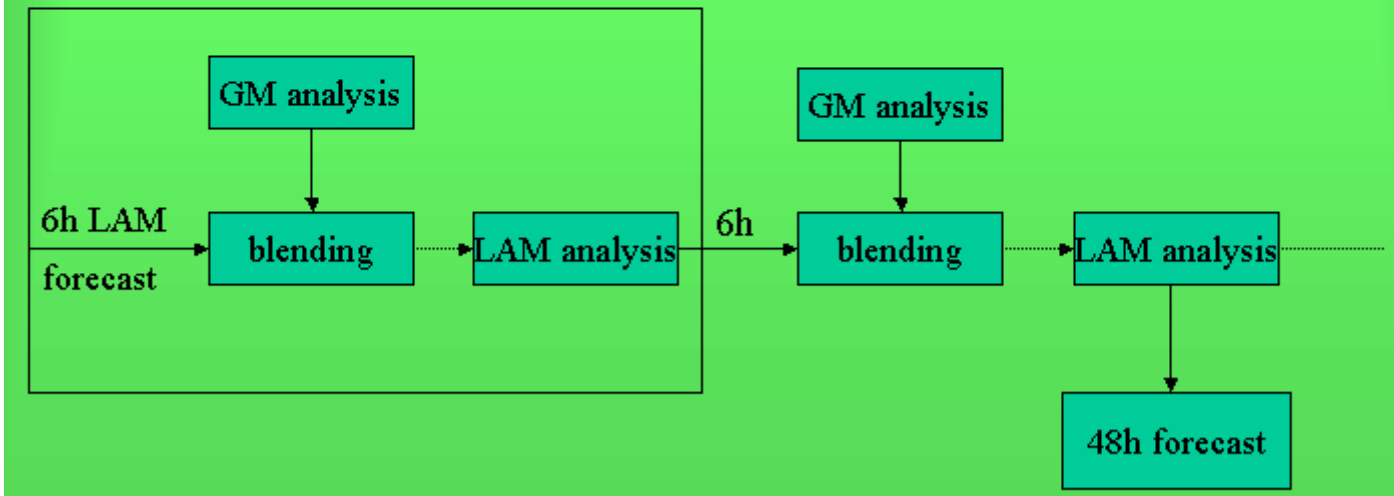
Best scenario

- Examples: the “12h” spin-up cycle (NMC, CMC)



Best scenario (2)

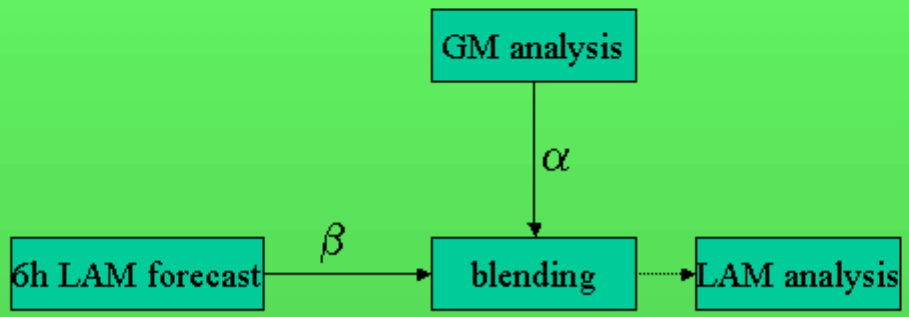
- Examples: the blendvar cycle (ask Marcsa for details)



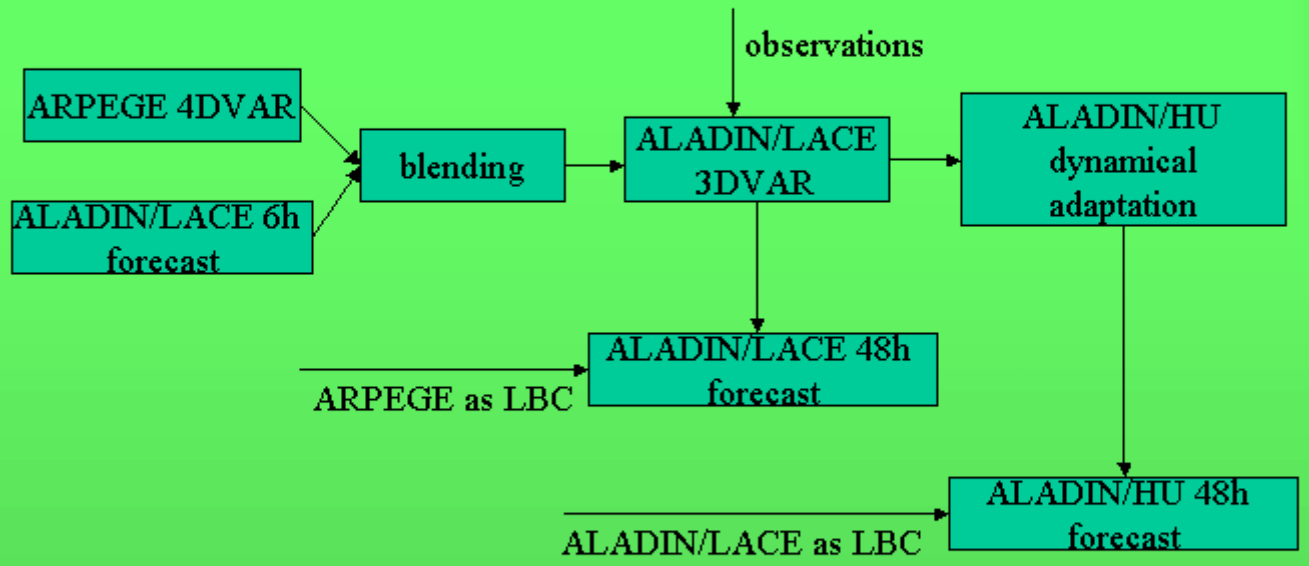
Best scenario/blending

- Large scales: from global analysis (most recent)
- Small scales: from LAM forecast (adapted small scales)
- Blending: optimal "mixture" (choice of α and β)

/see the presentation of Radmila/



Example of a nested cycle in practice



Data to be used

- Key issue: what data sources to be used globally, regionally and locally?
  - Global: everything available in GTS + satellite, new data sources (GPS,…….)
  - Regional and local: radar reflectivities & wind, local (conventional) surface data outside GTS, new data sources (GPS surface delay), satellite high resolution profiles, raw radiances, scatterometers...
- Priorities: so far all meteorological centres have decided on very pragmatic basis (e.g. data and know-how available at home)

Data to be used (2)

- Related issues
  - Monitoring and Quality Control
- Conclusion: extra data must be seen as a multi-dimensional problem
  - Needs competencies on network
  - Regular monitoring & control
  - Effort on observation operators

BUT: it is not only a key issue for the success of LAM/VAR, but also for forecast validation

Targetted observations on mesoscale?

- Point to the data in weather-sensitive areas (e.g. radar reflectivities in and around a convective system)
- Reduce the amount of data to the set that is most likely to give a benefit to the analysis
- Define strategies for future high-density networks (location, frequency, type of measurements)

Summary/LAM & LBC

- Scale selection (first guess, background errors, structure functions, incremental)
- Best scenarii for nested assimilation cycles
- Observations used
- Targetted observations?

Boundaries and coupling in 4DVAR (more questions than answers)

Problem of increments near the borders

Coupling in 4DVAR

Control through boundaries

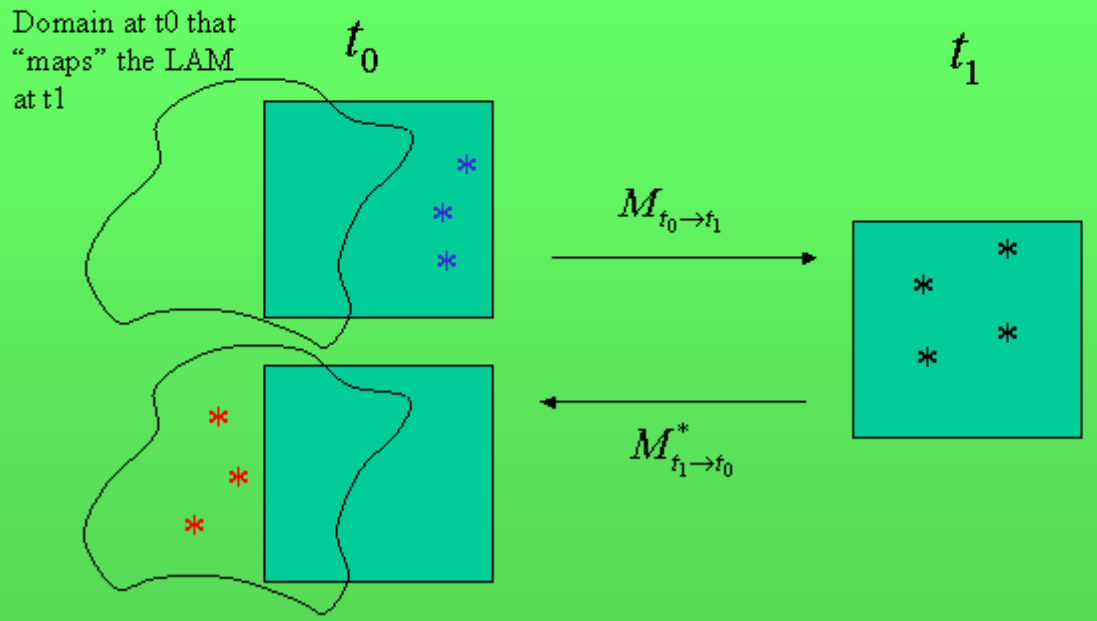
Coupling strategy for the trajectory

Problem of increments near the borders

- Static case (3DVAR):
  - Projection: set the increments to zero at the boundaries
  - Remove observations near the boundaries
  - Large enough extension zone (possible in bi-Fourier spectral models, e.g. HIRLAM)
  - Do nothing…… (probably acceptable in 3DVAR (??))

Problem of increments near the borders (2)

- Advective case (4DVAR): "Sweeping effect"



Problem of increments near the borders (3)

- Ideally:
  - Inflow area → upstream observations should be taken into account in M*
  - Outflow area → downstream innovations should be advected out of the domain (considered in M)
- Consequences in LAM:
  - Inflow observations are not used → stronger impact of first guess
  - Outflow observations are leaving the domain → useless analysis of some innovations
  - Possible mismatches between analysis increments and LBC?

Coupling in 4DVAR/TL + AD models

- Very few theoretical approaches (Lu & Browning, 2000), otherwise mainly pragmatic considerations
  - Well-posed LBC formulations are required (possible in non-hydrostatic models, Eulerian equations……)
  - Hydrostatic primitive equation models always use ill-posed LBC (e.g. Davies coupling in ALADIN) $\Rightarrow$ over-deterministic
  - Even in a well-posed framework (i.e. the amount of information from the coupler is equal to the amount of variables to be coupled) the assumption of perfect LBC data (associated to zero increments at the boundaries) is not consistent: only "inflow direct and adjoint waves" can be set to zero. "Outflow direct and adjoint waves" should evolve freely.

Coupling in 4DVAR/TL + AD models (2)

- Pragmatic approach:

$$\Psi = \alpha\Psi_i + (1-\alpha)\Psi_{LBC}$$

Ψ_i interior model solution before coupling

Ψ_{LBC} coupling value (external forcing)

$$\Psi^*_{LBC} = (1-\alpha)\Psi^*_i + \Psi^*_{LBC}$$

$$\Psi^* = \alpha\Psi^*_i$$

Ψ^*_{LBC} adjoint of coupling

- The adjoint of coupling: sensitivity with respect to LBC ("LBC sensitivity")

Coupling in 4DVAR/TL + AD models (3)

- Question: what type of coupling to be used for the better consideration of the interior increments in the TL/AD models?
  - No coupling: bio-periodic propagation or wall (fully reflective) $\Rightarrow$ unrealistic
  - Impose zero increments (equivalent to assume perfect LBC)

    $\Rightarrow$ can lead to strong, unrealistic increments in the coupling area (Zupanski, 1993)

    $\Rightarrow$ results might depend on the quality of the chosen grid-point coupling: radiative and dissipative properties,……..

Coupling in 4DVAR/TL + AD models (4)

- Question: what type of coupling to be used for the better consideration of the interior increments in the TL/AD models?
  - Spectral space coupling to be revisited for spectral models?

    TL/AD models might be better designed for wave-space coupling, because they are not so complex, non-linear, than the direct complete forecast model. Thus, energy cascades and large scale forcing of the increments could be represented "correctly" in the linear simplified models.
- To which extent can the coupling in TL/AD models be defined differently from the one used in the direct model?

Control through boundaries

- Question: Should LBC sensitivities be included in the control of the analysis solution, and if yes by what method?

  Few sensitivity studies have dealt with control through LBC:

  $\Rightarrow$ relatively modest or no impact (Errico & Vukicevic, 1992, Soci, 2000)

  $\Rightarrow$modest impact at high price (LBC is refreshed at every timestep and time filter, Gustafsson et al., 1998)

Control through boundaries (2)

- Question: Should LBC sensitivities be included in the control of the analysis solution, and if yes by what method?

  Inside LAM/4DVAR: hardly ever tested!!

  * no LBC at all in the control ⇒spurious increments at the coupling zone (Zupanski, 1993)

  * LBC controlled by relaxation towards the large scale analysis ⇒modified LBC would be needed to have the LBC consistent with the modified interior analysis (Zou & Kuo, 1996)

Control through boundaries (3)

- What method??
  - Modify LBC and re-compute the trajectory (and then redo a new minimisation)
  - Add LBC to control variable: never done anywhere!!!

    Is it even sound mathematically??

    It would need, strictly speaking, the computation of the LBC error covariance matrix

Coupling strategy for the 4DVAR trajectory

- Question: Keep always the initial LBC or use the refreshed ones provided by the adjoint of coupling?
    - Refreshment could provide LBC data more consistent with the changes produced in the (interior) analysis

BUT

- It might also introduce numerical noise or imbalances
- Does it impact on the quality of the convergence of a multi-incremental 4DVAR?

Coupling strategy for the 4DVAR trajectory (2)

- If really coupling and LBC are proven to be important for mesoscale 4DVAR in a LAM, then they would be part of the design of a multi-incremental approach, as being one of the driving factors for the refreshment of the trajectory

Other questions/General aspects on 4DVAR/coupling & LBC

- Most studies on coupling in LAM/VAR have dealt with sub-synoptic ranges ($\delta x \geq 30$ km, time span 6-12 h in 4DVAR, forecast range ≥ 36 h)
  - How important is this issue at meso-β/meso-γ scale scales ($\delta x \leq 10$ km, time span ≤ 3h in 4DVAR, forecast range ≤ 36 h)?
  - What is more important?
    - **Physically consistent LBC (radiative…….)**
    - **Design-consistent with the control/assimilation system (e.g. spectral coupling if spectral control variable)?**

Other questions/General aspects on 4DVAR/coupling & LBC (2)

- Regional 4DVAR ⇒ cost issue
- Multiply nested data assimilation (global, regional, local)? ⇒new data is needed at every step
- Equivalence: {no LBC in control} ⇔ {zero boundary increments imposed through (any) coupling in TL/AD models}
- 2-way coupling/nesting (coupler and coupling models exchange information at their boundaries)
- Pass information from the coupled (LAM) adjoint model back to the coupler (global) adjoint model

Summary, overview

- Generalities - Practice - LAM & LBC
- Generalities:
  - Data assimilation
  - Optimal interpolation
  - Variational data assimilation
    - Cost function and its gradient
    - Error covariance matrices
    - Minimization

Summary, overview (2)

- Practical implementation:
  - (Multi)-incremental formulation
  - Quasi-continuous approach
  - The operational 4DVAR algorithm at Météo France
  - Flow-dependant structure functions

Summary, overview (3)

- LAM & LBC:
  - Scale selection
    - Choice of first guess
    - Control on the structure functions (background errors)
    - Other aspects (initialization, incremental?)
  - Possible tools for high resolution analysis
  - Nested assimilation cycles
    - Scenarii, examples
    - Observations

Summary, overview (4)

- LAM & LBC (continued):
  - Boundaries and coupling
    - Problem of increments near the borders
    - Coupling in 4DVAR
    - Control through boundaries