Variational Data Assimilation
Weak Constraint 4D-Var

Yannick Trémolet

ECMWF

JCSDA Summer Colloquium on Data Assimilation
July 2009

Outline

4D Variational Data Assimilation

4D-Var comprises the minimisation of:

$$
\begin{aligned}
J(\mathbf{x}) &= \frac{1}{2}[\mathcal{H}(\mathbf{x}) - \mathbf{y}]^T \mathbf{R}^{-1} [\mathcal{H}(\mathbf{x}) - \mathbf{y}] \\
&+ \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\mathcal{F}(\mathbf{x})^T \mathbf{C}^{-1} \mathcal{F}(\mathbf{x})
\end{aligned}
$$

- $\mathbf{x}$ is the 4D state of the atmosphere over the assimilation window.
- $\mathcal{H}$ is a 4D observation operator, accounting for the time dimension.
- $\mathcal{F}$ represents the remaining theoretical knowledge after background information has been accounted for (balance, DFI...).
- Control variable reduces to $\mathbf{x}_0$ using the hypothesis: $\mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_{i-1})$.
- The solution is a trajectory of the model $\mathcal{M}$ even though it is not perfect...

Weak Constraint 4D-Var

- Typical assumptions in data assimilation are to ignore:
  - Observation bias,
  - Observation error correlation,
  - Model error (bias and random).
- The perfect model assumption limits the length of the analysis window that can be used to roughly 12 hours.
- Model bias can affect assimilation of some observations (radiance data in the stratosphere).
- In weak constraint 4D-Var, we define the model error as

$$\eta_i = \mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1}) \quad \text{for } i = 1, \ldots, n$$

and we allow η_i to be non-zero.

Outline

Weak Constraint 4D-Var

- We can derive the weak constraint cost function using Bayes' rule:

$$p(\mathbf{x}_0 \cdots \mathbf{x}_n | \mathbf{x}_b; \mathbf{y}_0 \cdots \mathbf{y}_n) = \frac{p(\mathbf{x}_b; \mathbf{y}_0 \cdots \mathbf{y}_n | \mathbf{x}_0 \cdots \mathbf{x}_n) p(\mathbf{x}_0 \cdots \mathbf{x}_n)}{p(\mathbf{x}_b; \mathbf{y}_0 \cdots \mathbf{y}_n)}$$

- The denominator is independent of $\mathbf{x}_0 \cdots \mathbf{x}_n$.
- The term $p(\mathbf{x}_b; \mathbf{y}_0 \cdots \mathbf{y}_n | \mathbf{x}_0 \cdots \mathbf{x}_n)$ simplifies to:

$$p(\mathbf{x}_b | \mathbf{x}_0) \prod_{i=0}^{n} p(\mathbf{y}_i | \mathbf{x}_i)$$

- Hence

$$p(\mathbf{x}_0 \cdots \mathbf{x}_n | \mathbf{x}_b; \mathbf{y}_0 \cdots \mathbf{y}_n) \propto p(\mathbf{x}_b | \mathbf{x}_0) \left[\prod_{i=0}^{n} p(\mathbf{y}_i | \mathbf{x}_i) \right] p(\mathbf{x}_0 \cdots \mathbf{x}_n)$$

Weak Constraint 4D-Var

$$p(\mathbf{x}_0 \cdots \mathbf{x}_n | \mathbf{x}_b; \mathbf{y}_0 \cdots \mathbf{y}_n) \propto p(\mathbf{x}_b | \mathbf{x}_0) \left[\prod_{i=0}^{n} p(\mathbf{y}_i | \mathbf{x}_i) \right] p(\mathbf{x}_0 \cdots \mathbf{x}_n)$$

- Taking minus the logarithm gives the cost function:

$$J(\mathbf{x}_0 \cdots \mathbf{x}_n) = -\log p(\mathbf{x}_b | \mathbf{x}_0) - \sum_{i=0}^{n} \log p(\mathbf{y}_i | \mathbf{x}_i) - \log p(\mathbf{x}_0 \cdots \mathbf{x}_n)$$

- The terms involving $\mathbf{x}_b$ and $\mathbf{y}_i$ are the background and observation terms of the strong constraint cost function.
- The final term is new. It represents the *a priori* probability of the sequence of states $\mathbf{x}_0 \cdots \mathbf{x}_n$.

Weak Constraint 4D-Var

- Given the sequence of states $\mathbf{x}_0 \cdots \mathbf{x}_n$, we can calculate the corresponding model errors:

$$\eta_i = \mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1}) \quad \text{for } i = 1, \ldots, n$$

- We can use our knowledge of the statistics of model error to define

$$p(\mathbf{x}_0 \cdots \mathbf{x}_n) \equiv p(\mathbf{x}_0; \eta_1 \cdots \eta_n)$$

- One possibility is to assume that model error is uncorrelated in time. In this case:

$$p(\mathbf{x}_0 \cdots \mathbf{x}_n) \equiv p(\mathbf{x}_0) p(\eta_1) \cdots p(\eta_n)$$

- If we take $p(\mathbf{x}_0) = const.$ (all states equally likely), and $p(\eta_i)$ as Gaussian with covariance matrix $\mathbf{Q}_i$, weak constraint 4D-Var adds the following term to the cost function:

$$\frac{1}{2} \sum_{i=1}^{n} \eta_i^T \mathbf{Q}_i^{-1} \eta_i$$

Weak Constraint 4D-Var

- For Gaussian, temporally-uncorrelated model error, the weak constraint 4D-Var cost function is:

$$
\begin{aligned}
J(\mathbf{x}) &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) \\
&+ \frac{1}{2}\sum_{i=0}^{n} [\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i]^T \mathbf{R}_i^{-1} [\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i] \\
&+ \frac{1}{2}\sum_{i=1}^{n} [\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1})]^T \mathbf{Q}_i^{-1} [\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1})]
\end{aligned}
$$

- Do not reduce the control variable using the model and retain the 4D nature of the control variable.
- Account for the fact that the model contains some information but is not exact by adding a model error term to the cost function.
- Model $\mathcal{M}$ is not verified exactly: it is a weak constraint.
- If model error is correlated in time, the model error term contains additional cross-correlation blocks.

Outline

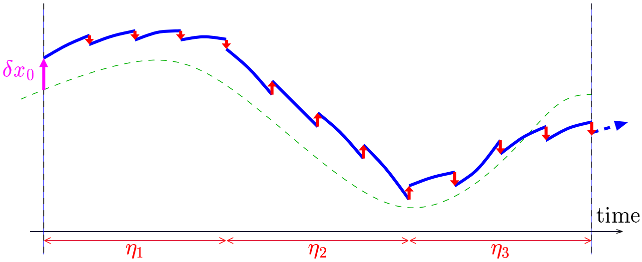
4D-Var with Model Error Forcing

$$
\begin{aligned}
J(\mathbf{x}_0, \eta) &= \frac{1}{2} \sum_{i=0}^{n} [\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i]^T \mathbf{R}_i^{-1} [\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] \\
&+ \frac{1}{2} (\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2} \eta^T \mathbf{Q}^{-1} \eta
\end{aligned}
$$

$$
\text{with } \mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_{i-1}) + \eta_i.
$$

- η_i has the dimension of a 3D state,
- η_i represents the instantaneous model error,
- η is constrained by the fact that it is propagated by the model.
- All results shown later are for constant forcing over the length of the assimilation window, i.e. for correlated model error.

4D-Var with Model Error Forcing



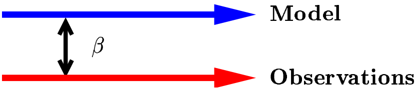
- TL and AD models can be used with little modification,
- Information is propagated between obervations and IC control variable by TL and AD models.

Outline

4D-Var with Model Bias

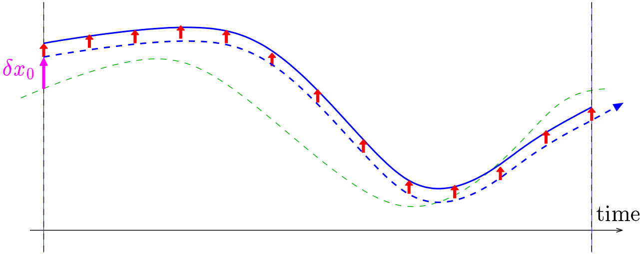
$$
\begin{aligned}
J(\mathbf{x}_0, \beta) &= \frac{1}{2}\sum_{i=0}^{n}[\mathcal{H}(\mathbf{x}_i^m + \beta_i) - \mathbf{y}_i]^T \mathbf{R}_i^{-1}[\mathcal{H}(\mathbf{x}_i^m + \beta_i) - \mathbf{y}_i] \\
&+ \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\beta^T \mathbf{Q}_\beta^{-1}\beta
\end{aligned}
$$

with $\mathbf{x}_i^m = \mathcal{M}_{i,0}(\mathbf{x}_0)$ and $\mathbf{x}_i = \mathcal{M}_{i,0}(\mathbf{x}_0) + \beta_i$.



- β_i is 3D state-like,
- The model is not perturbed,
- β sees global (model − all observations) bias,
- Does not correct for bias of one subset of observations against another subset of observations.

4D-Var with Model Bias



- Bias added to forecast at post-processing stage,
- Makes sense if β is slowly varying or constant ($\beta_i = \beta$),
- Information is propagated between obervations and IC control variable by TL and AD models (not modified),
- Model bias is represented by additional parameters, without entering the model equations,
- Optimisation problem is very similar to strong constraint 4D-Var.

Outline

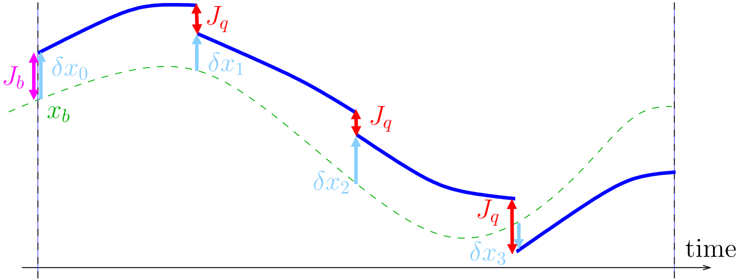
4D State Control Variable

- Use $\mathbf{x} = \{\mathbf{x}_i\}_{i=0,\ldots,n}$ as the control variable.
- Nonlinear cost function:

$$
\begin{aligned}
J(\mathbf{x}) &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) \\
&+ \frac{1}{2} \sum_{i=0}^{n} [\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i]^T \mathbf{R}_i^{-1} [\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i] \\
&+ \frac{1}{2} \sum_{i=1}^{n} [\mathcal{M}(\mathbf{x}_{i-1}) - \mathbf{x}_i]^T \mathbf{Q}_i^{-1} [\mathcal{M}(\mathbf{x}_{i-1}) - \mathbf{x}_i]
\end{aligned}
$$

- In principle, the model is not needed to compute the J_o term.
- In practice, the control variable will be defined at regular intervals in the assimilation window and the model used to fill the gaps.

4D State Control Variable



- Model integrations within each time-step (or sub-window) are independent:
  - Information is not propagated across sub-windows by TL/AD models,
  - Natural parallel implementation.
- Tangent linear and adjoint models:
  - Can be used without modification,
  - Propagate information between observations and control variable within each sub-window.

Control Variable in Weak Constraint 4D-Var

4D-Var	4D-Var$_x$	4D-Var$_\eta$	4D-Var$_\beta$
$\mathbf{x}_0$	$\mathbf{x}$	$\mathbf{x}_0,\ \eta$	$\mathbf{x}_0,\ \beta$
$\mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_{i-1})$	$\mathbf{x}_i \approx \mathcal{M}_i(\mathbf{x}_{i-1})$	$\mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_{i-1}) + \eta_i$	$\mathbf{x}_i = \mathcal{M}_{i,0}(\mathbf{x}_0) + \beta_i$
$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$
3D Initial Condition	4D State	3D I.C. + Model Error Forcing	3D I.C. + Model Bias

Outline

Model error covariance matrix

- The usual choice is $\mathbf{Q} = \alpha\mathbf{B}$.
- Linearisation in incremental formulation gives:

$$\delta\mathbf{x}_n = \mathbf{M}_n \ldots \mathbf{M}_1 \delta\mathbf{x}_0 + \sum_{i=1}^{n} \mathbf{M}_n \ldots \mathbf{M}_{i+1} \eta_i$$

- $\delta\mathbf{x}_0$ can be identified with η_0.
- The solution of the analysis equation satisfies:

$$\delta\mathbf{x}_0 = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x}_b))$$

$$\eta = \mathbf{Q}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{Q}\mathbf{H}^T)^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x}_b))$$

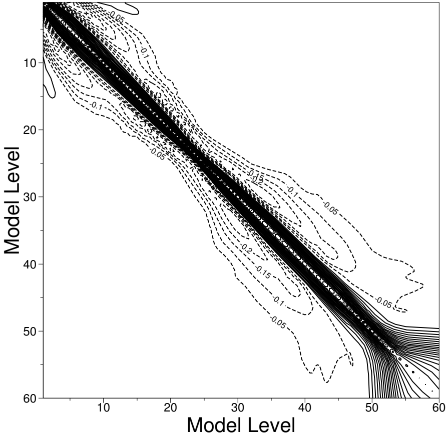
- If $\mathbf{Q}$ and $\mathbf{B}$ are proportional, $\delta\mathbf{x}_0$ and η are constrained in the same directions, may be with different relative amplitudes.
- They both predominantly retrieve the same information: $\mathbf{Q} = \alpha\mathbf{B}$ is too limiting.

Generating a Model Error Covariance Matrix

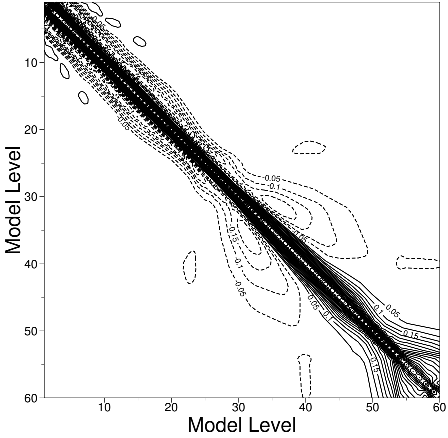
- **B** is estimated from an ensemble of 4D-Var assimilations.
- Considering the forecasts run from the 4D-Var members:
  - At a given step, each model state is supposed to represent the same *true* atmospheric state,
  - The tendencies from each of these model states should represent possible evolutions of the atmosphere from that same *true* atmospheric state,
  - The differences between these tendencies can be interpreted as possible uncertainties in the model or realisations of *model error*.
- **Q** can be estimated by applying the statistical model used for **B** to tendencies instead of analysis increments.
- **Q** has narrower correlations and smaller amplitudes than **B**.

Average Temperature Vertical Correlations

Background Error

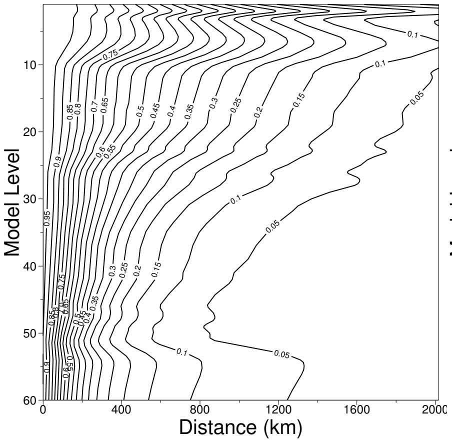


Model Error

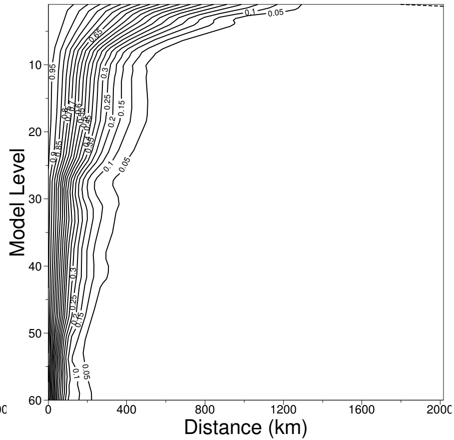


Temperature Horizontal Correlations

Background Error



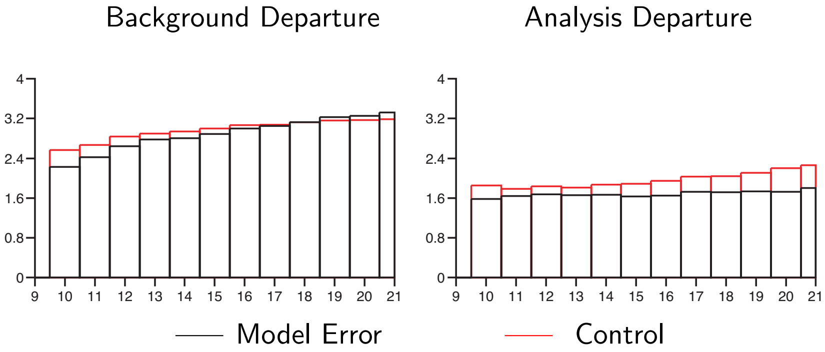
Model Error



Outline

Results: Fit to observations

`AMprofiler-windspeed Std Dev N.Amer`

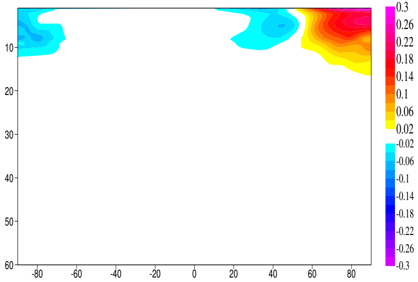


- Fit to observations is more uniform over the assimilation window.
- Background fit improved only at the start: error varies in time ?

Model Error Forcing

Zonal Mean Temperature
July 2004

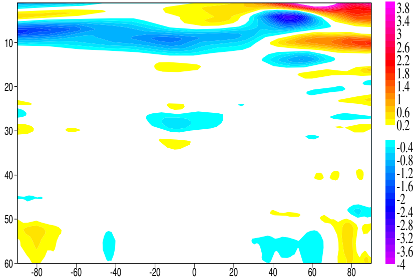
M.E. Forcing ⟶

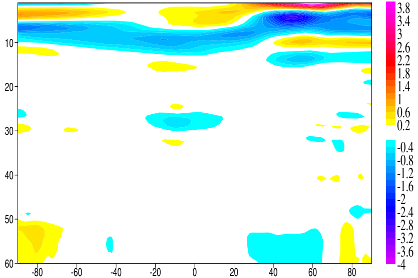


M.E. Mean Increment ↘

Control Mean Increment

↓



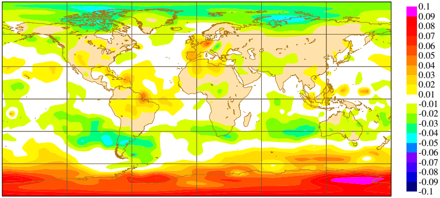


Mean Model Error Forcing

Temperature
Model level 11 (≈5hPa)
July 2004

Mean M.E. Forcing ⟶

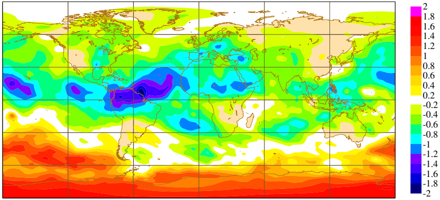
Wednesday 30 June 2004 21UTC ©ECMWF Mean Model Error Forcing (eptg)
Temperature, Model Level 11
Min = -0.05, Max = 0.10, RMS Global=0.02, N.hem=0.01, S.hem=0.03, Tropics=0.01



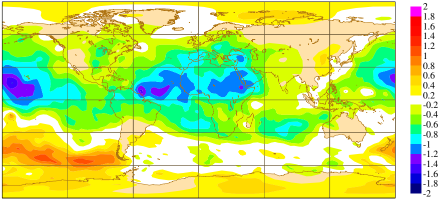
M.E. Mean Increment ↘

Control Mean Increment
↓

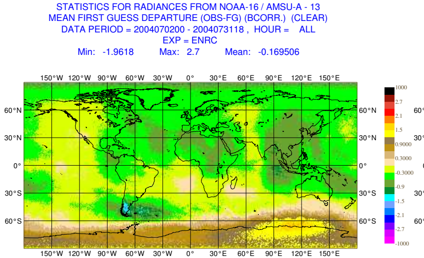
Monday 5 July 2004 00UTC ©ECMWF Mean Increment (enrc)
Temperature, Model Level 11
Min = -1.97, Max = 1.61, RMS Global=0.66, N.hem=0.54, S.hem=0.65, Tropics=0.77

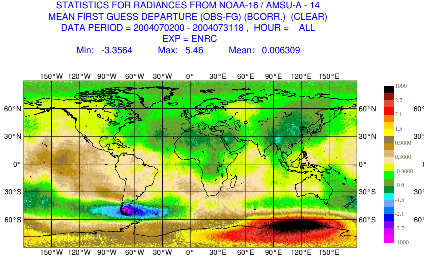


Monday 5 July 2004 00UTC ©ECMWF Mean Increment (eptg)
Temperature, Model Level 11
Min = -1.60, Max = 1.15, RMS Global=0.55, N.hem=0.51, S.hem=0.41, Tropics=0.69

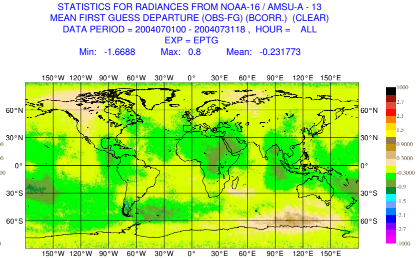


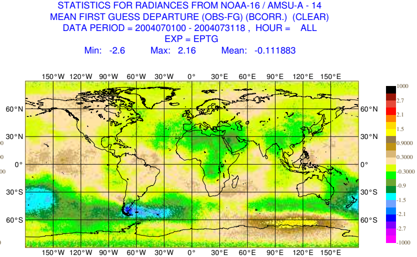
AMSU-A First Guess Departures





Strong Constraint

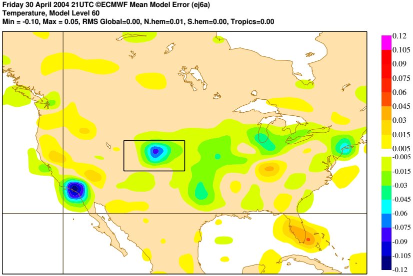


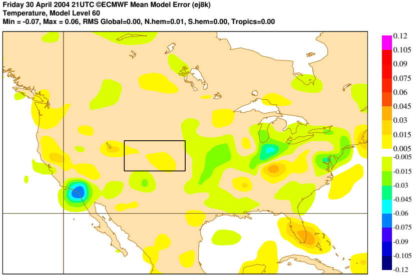


Weak Constraint

Outline

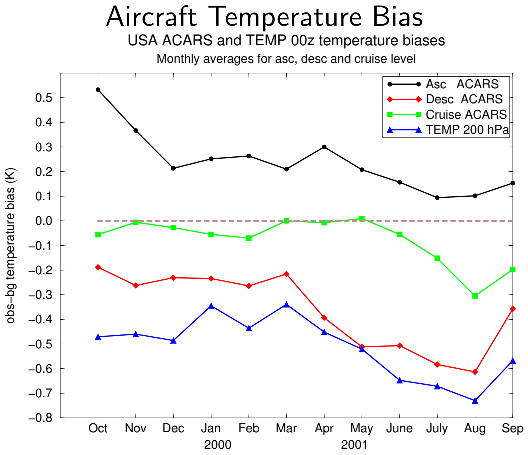
Low Level Mean Model Error Forcing





- The only significant source of observations in the box is aircraft data (Denver airport).
- Removing aircraft data in the box eliminates the spurious forcing.

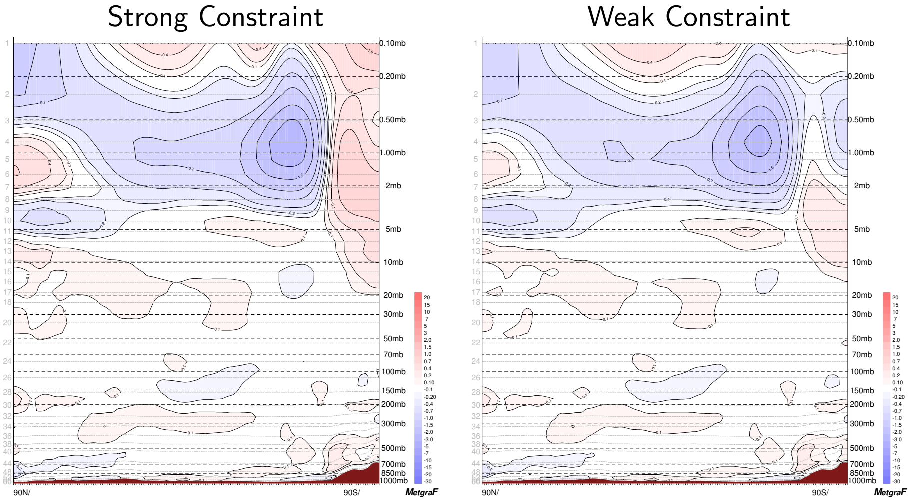
Is it model error?



Observations are biased.

Figure from Lars Isaksen

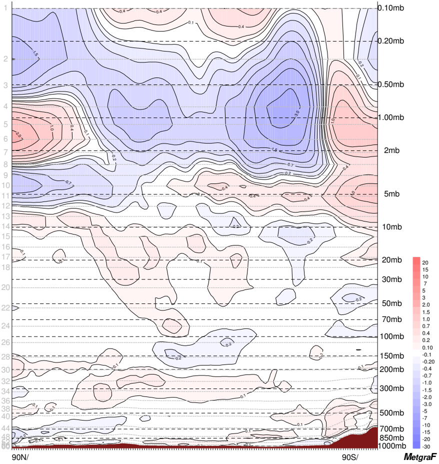
Is it model error?



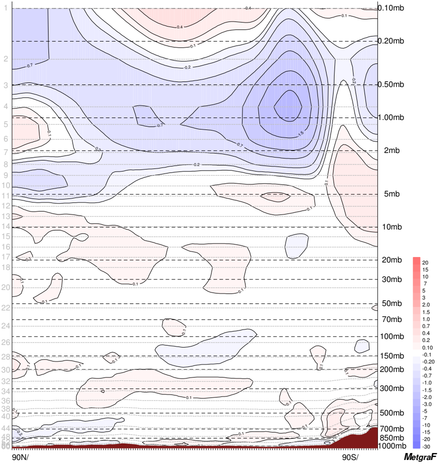
The mean temperature increment is smaller and smoother with weak constraint 4D-Var (Stratosphere only, June 1993).

Is it model error?

ERA interim



Weak Constraint

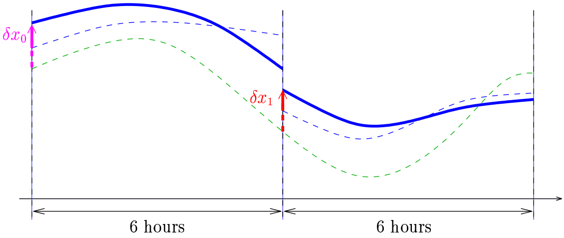


The work on model error has helped identify other sources of error in the system.

Outline

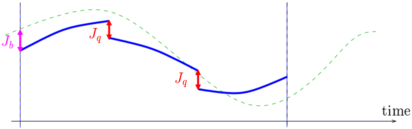
Weak Constraint 4D-Var Configurations

- 6-hour sub-windows:



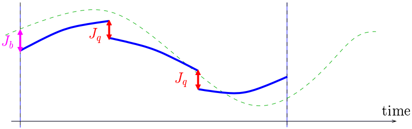
  - Better than 6-hour 4D-Var: two cycles are coupled through J_q,
  - Better than 12-hour 4D-Var: more information (imperfect model), more control,
  - $\mathbf{Q} = \alpha\mathbf{B}$ could be used in that case.
- Single time-step sub-windows:
  - Each assimilation problem is instantaneous = 3D-Var,
  - Equivalent to a string of 3D-Var problems coupled together and solved as a single minimisation problem,
  - Approximation can be extended to non instantaneous sub-windows.

Weak Constraint 4D-Var: Sliding Window

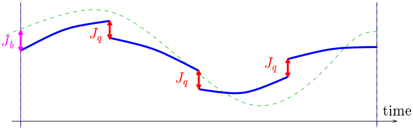


(1) Weak constraint 4D-Var

Weak Constraint 4D-Var: Sliding Window

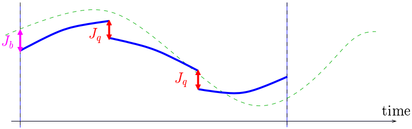


(1) Weak constraint 4D-Var

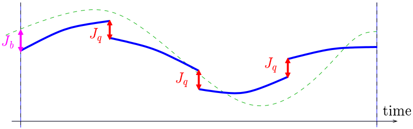


(2) Extended window

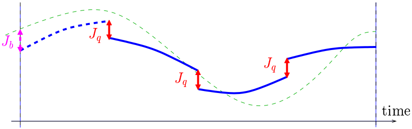
Weak Constraint 4D-Var: Sliding Window



(1) Weak constraint 4D-Var

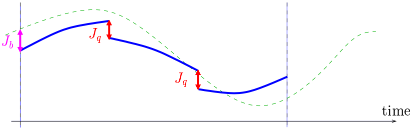


(2) Extended window

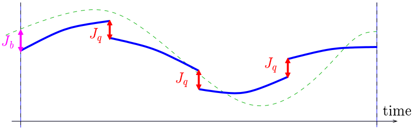


(3) Initial term has converged

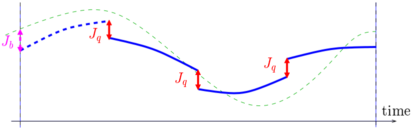
Weak Constraint 4D-Var: Sliding Window



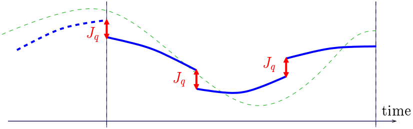
(1) Weak constraint 4D-Var



(2) Extended window

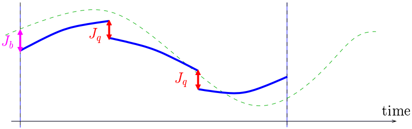


(3) Initial term has converged

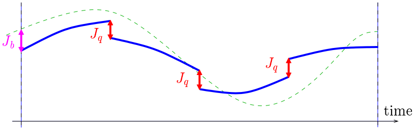


(4) Assimilation window is moved forward

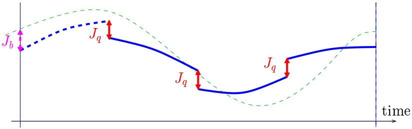
Weak Constraint 4D-Var: Sliding Window



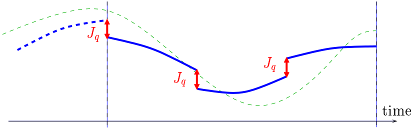
(1) Weak constraint 4D-Var



(2) Extended window



(3) Initial term has converged



(4) Assimilation window is moved forward

- This implementation is an approximation of weak contraint 4D-Var with an assimilation window that extends indefinitely in the past...
- ...which is equivalent to a Kalman smoother that has been running indefinitely.

4D State Control Variable: Minimization

- The Hessian of the cost function is the sum of
  - the Hessian of the observation term $\mathbf{G} = diag(\cdots, \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i, \cdots)$
  - the tri-diagonal Hessian of $J_b + J_q$ is:

$$\begin{pmatrix} \mathbf{B}^{-1} + \mathbf{M}_1^T \mathbf{Q}_1^{-1} \mathbf{M}_1 & -\mathbf{M}_1^T \mathbf{Q}_1^{-1} & & & 0 \\ -\mathbf{Q}_1^{-1} \mathbf{M}_1 & \mathbf{Q}_1^{-1} + \mathbf{M}_2^T \mathbf{Q}_2^{-1} \mathbf{M}_2 & & & \\ & -\mathbf{Q}_2^{-1} \mathbf{M}_2 & \ddots & -\mathbf{M}_{n-1}^T \mathbf{Q}_{n-1}^{-1} & \\ & & & \mathbf{Q}_{n-1}^{-1} + \mathbf{M}_n^T \mathbf{Q}_n^{-1} \mathbf{M}_n & -\mathbf{M}_n^T \mathbf{Q}_n^{-1} \\ 0 & & & -\mathbf{Q}_n^{-1} \mathbf{M}_n & \mathbf{Q}_n^{-1} \end{pmatrix}$$

- The off-diagonal terms propagate the information between the sub-windows.
- Accounting for correlated model error, the matrix becomes full.

4D State Control Variable: Properties

- Preconditioning with $B^{-1/2}, Q_1^{-1/2}, \cdots, Q_n^{-1/2}$
- Over one time step, $M_i \approx I$:

$$\hat{J}'' \approx \begin{pmatrix} 2I & -I & & & 0 \\ -I & 2I & -I & & \\ & -I & \ddots & \ddots & \\ & & \ddots & 2I & -I \\ 0 & & & -I & I \end{pmatrix} + \hat{J_o}''$$

- The largest eigenvalue is:

$$\lambda_{max} \approx 4 + 2 n_{obs}/n \max\left[(\sigma_b/\sigma_o)^2, (\sigma_q/\sigma_o)^2\right]$$

- Approximately the same as the maximum eigenvalue of strong constraint 4D-Var for the sub-windows.
- But the smallest eigenvalue is $\lambda_{min} \propto 1/n^2$.

4D State Control Variable: Properties

- Condition number:

$$\kappa \approx 2\, n\, n_{obs}\, (\sigma_q/\sigma_o)^2$$

  - Larger than for strong constraint 4D-Var,
  - Increases with the number of sub-windows (it takes n iterations to propagate information).
- Simplified Hessian of the cost function is close to a Laplacian operator: small eigenvalues are obtained for constant perturbations which might be well observed and project onto eigenvectors of J_o" associated with large eigenvalues.
- Using the square root of this tri-diagonal matrix to precondition the minimisation is equivalent to using the initial state and forcing formulation.
- Can we combine the benefits of treating sub-windows in parallel with efficient minimization?

Outline

Weak Constraints 4D-Var: Summary

- In strong constraint 4D-Var, we can use the constraints to reduce the minimization problem to an initial value problem.
- Weak constraint 4D-Var with a model error forcing term is very similar to an initial value problem with parameter estimation (parameters happen to represent model error).
- Weak constraint 4D-Var has already taught us about observation bias and errors in the balance operators,
- Weak constraint 4D-Var with constant model error forcing in the stratosphere should become operational in summer 2009.
- Weak constraint 4D-Var with a 4D state control variable is a fully four dimensional problem where J_q acts as a coupling term between sub-windows.

Weak Constraints 4D-Var: Open Questions

- Weak Constraint 4D-Var allows the perfect model assumption to be removed.
- it requires knowledge of the statistics of model error, and the ability to express this knowledge in the form of covariance matrices.
- What is the best model error covariance matrix?
- 4D-Var can handle correlated model error. What type of correlation model should be used?
- How can we distinguish model error from observation bias or other errors in the system?
- The statistical description of model error is one of the main current challenges in data assimilation.

Weak Constraints 4D-Var: Open Questions

- Weak Constraint 4D-Var allows the perfect model assumption to be removed.
- This allows longer windows to be contemplated.
- How much benefit can we gain from long window 4D-Var? How far from the optimal is 4D-Var with a 12h-window?
- Long window weak constraint 4D-Var is equivalent to a full rank Kalman smoother: it could be an efficient algorithm to implement it.
- Although the two weak constraint 4D-Var approaches are mathematically equivalent, they lead to very different minimization problems, with different possibilities for preconditioning. It is not yet clear which approach is the best.
- Can we combine the benefits of treating sub-windows in parallel with efficient minimization?
- Formulation of an incremental method for weak constraint 4D-Var remains a topic of research.