Towards a longer assimilation window in 4D-Var

Yannick Trémolet

Thanks to Paul Poli

ECMWF

October 2011

Outline

Outline

Weak Constraint 4D-Var

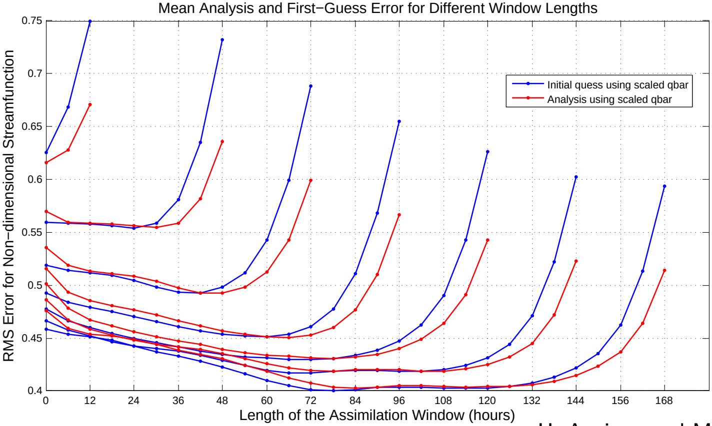
- For Gaussian, temporally-uncorrelated model error, the weak constraint 4D-Var cost function is:

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

- 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.
- This problem can be solved in parallel (saddle-point algorithm, no need for inverse of covariances, preconditioning is being investigated).

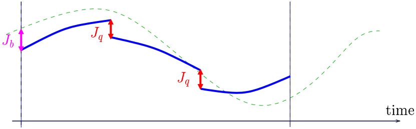
Longer is better

- Theory says: long window weak constraint 4D-Var is equivalent to a full rank Kalman smoother (Fisher *et al.*, 2005, Ménard and Daley, 1996).
- Long window weak constraint 4D-Var works for simple systems (Lorenz 95, QG):

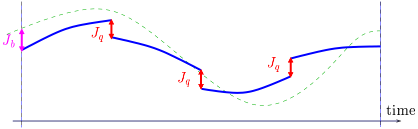


H. Auvinen and M. Fisher

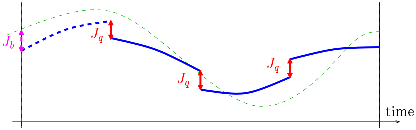
Long Window Weak Constraint 4D-Var



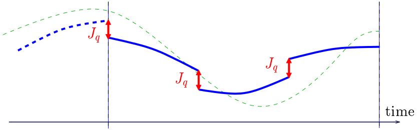
(1) Weak constraint 4D-Var



(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 (full rank) Kalman smoother that has been running indefinitely.
- And **B** is a problem of the past! Only the error characteristics of the fundamental ingredients of the DA problem remain.

Outline

4D-Var with Model Error Forcing

- In practice, weak constraint 4D-Var is still difficult to implement (in the IFS).
- Change of variable:

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

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

- η_i represents model error in a time step,
- η_i has the same dimension as a 3D state.

4D-Var with Constant Model Error Forcing

- Approximation: model error is constant.

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

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

- η represents model error in a time step,
- η has the same dimension as a 3D state.

- The number of degrees of freedom doubles.

Weak Constraints 4D-Var for Systematic Model Error

- For random model error, the 4D-Var cost function is:

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

- For systematic model error:

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

- Test case: model bias in the stratosphere.

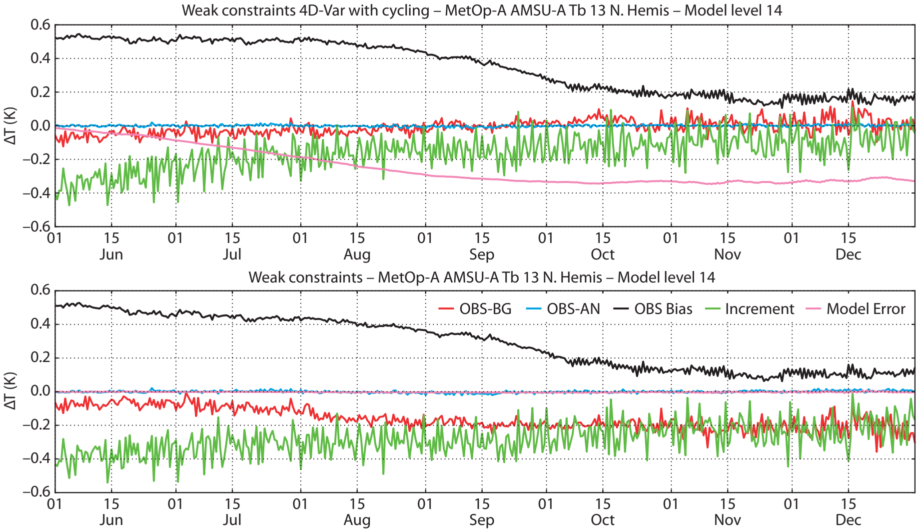
Model Error Covariance Matrix

- Currently, tendency differences between integrations of the members of an ensemble are used as a proxy for samples of model error.
- Statistics of model drift (for systematic model error).
- Use results from stochastic representation of uncertainties in EPS.
- It is possible to derive an estimate of $\mathbf{HQH}^T$ from cross-covariances between observation departures produced from pairs of analyses with different length windows (R. Todling).
- Is it possible to extract model error information using the relation $\mathbf{P}^f = \mathbf{MP}^a\mathbf{M}^T + \mathbf{Q}$?
- Model error is correlated in time: $\mathbf{Q}$ should account for time correlations. How?
- How to account for flow dependence?

Outline

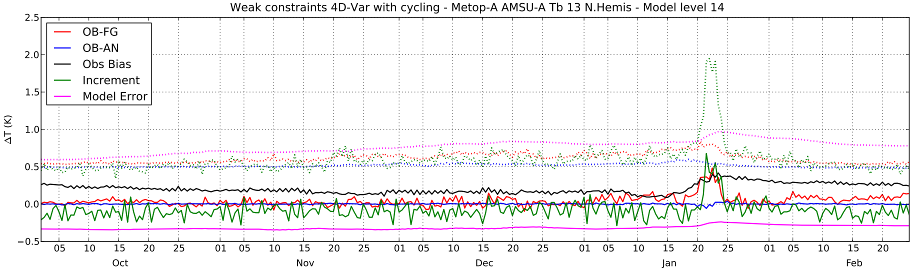
Outline

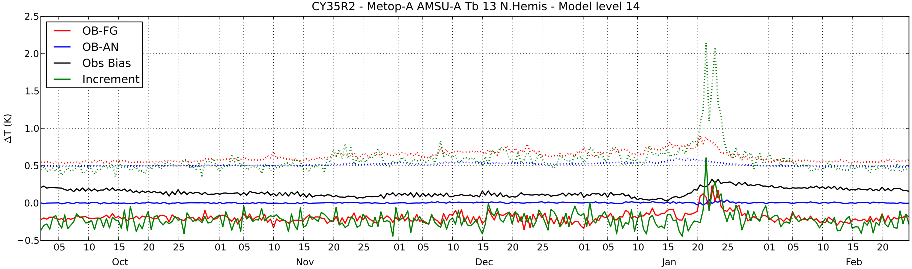
Weak Constraints 4D-Var with Cycling Term



The short term forecast is improved with the model error cycling.
Weak constraints 4D-Var can correct for seasonal bias (partially).

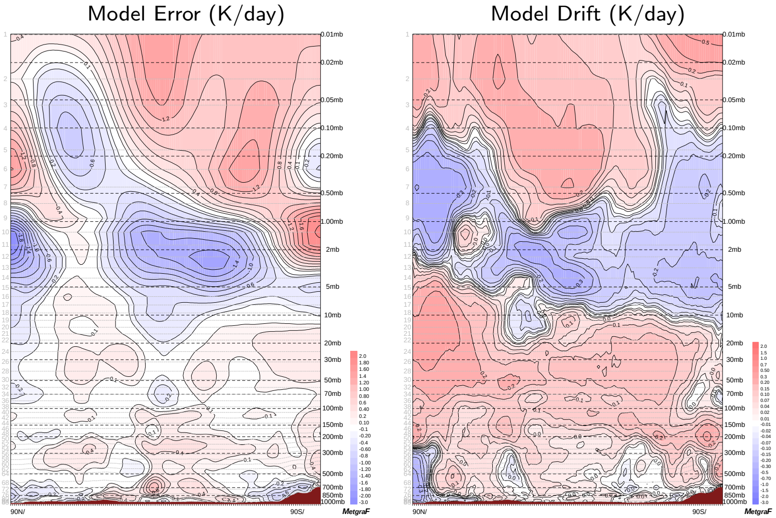
Observation Error or Model Error?





Observation error bias correction can compensate for model error.

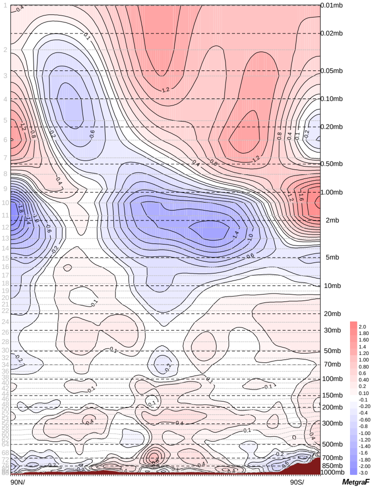
Weak Constraint 4D-Var



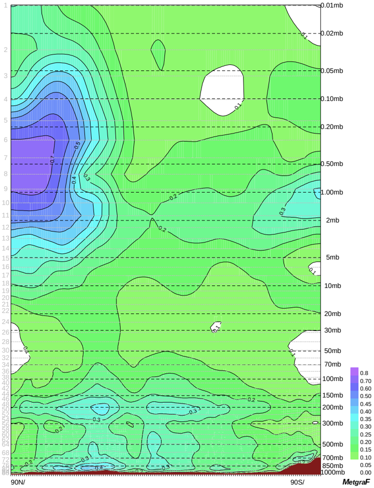
Temperature zonal means, December 2010

Weak Constraint 4D-Var

Mean (K/day)



Standard Deviation (K/day)

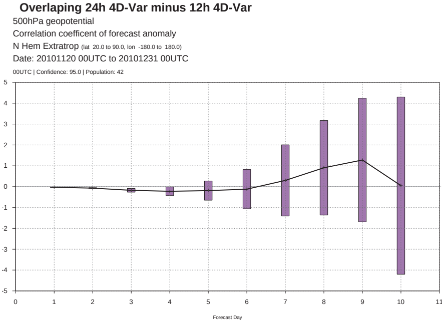


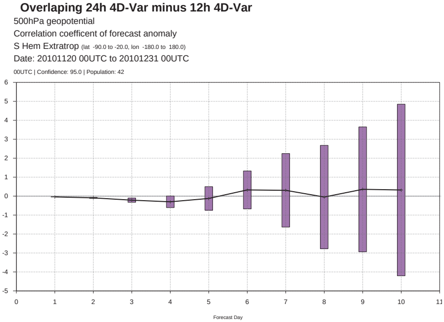
Temperature zonal means, December 2010

Model error estimates vary rapidly in NH stratosphere.

Outline

24h 4D-Var: Forecast Scores

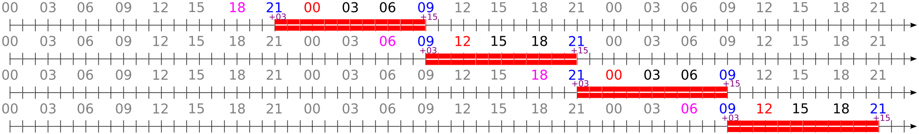




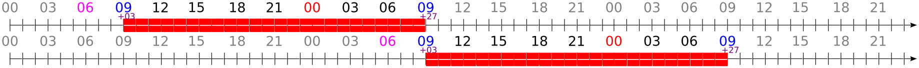
Forecast scores for overlapping 24h 4D-Var with respect to 12h 4D-Var.

Long Window 4D-Var Cycling

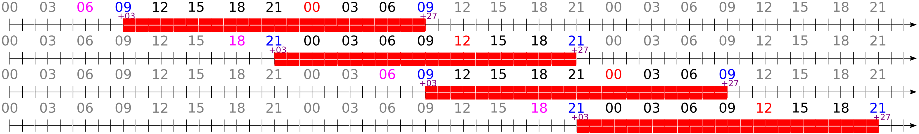
12-hour 4DVAR



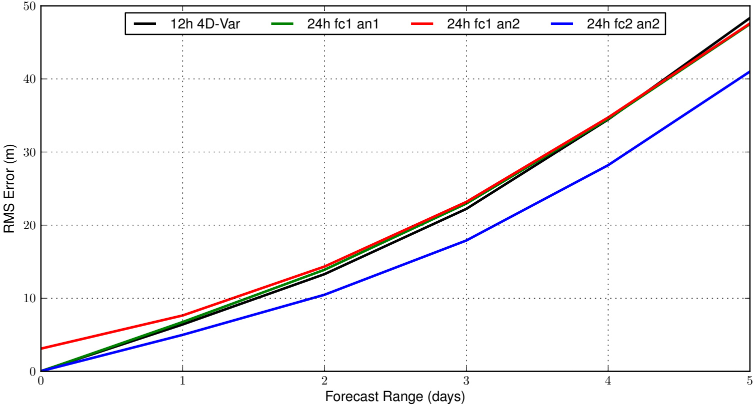
24-hour 4DVAR



24-hour overlapping 4DVAR

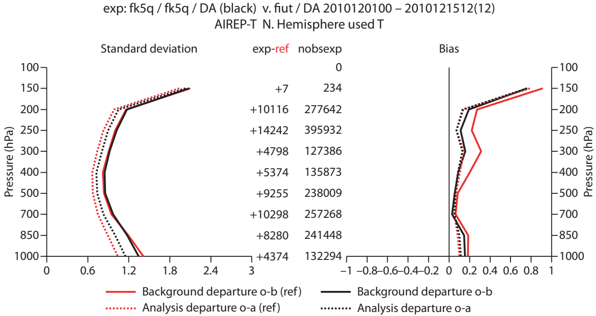


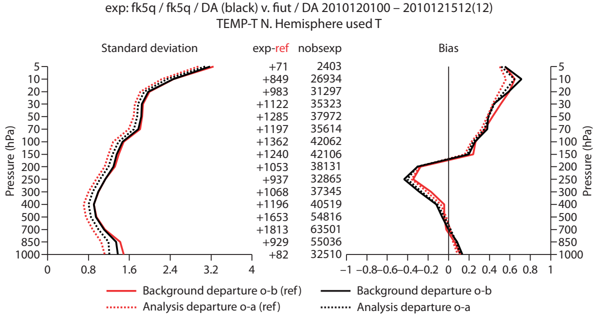
24h 4D-Var: Forecast Scores



- With overlapping analysis windows, there are several analyses to start the forecast from and to verify against!
- Warning: too few cases to draw conclusions from this figure.

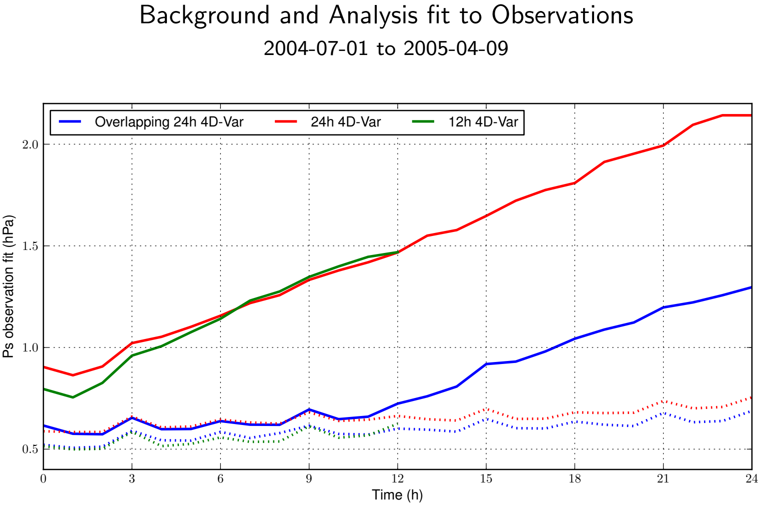
24h 4D-Var: Observation Statistics





Outline

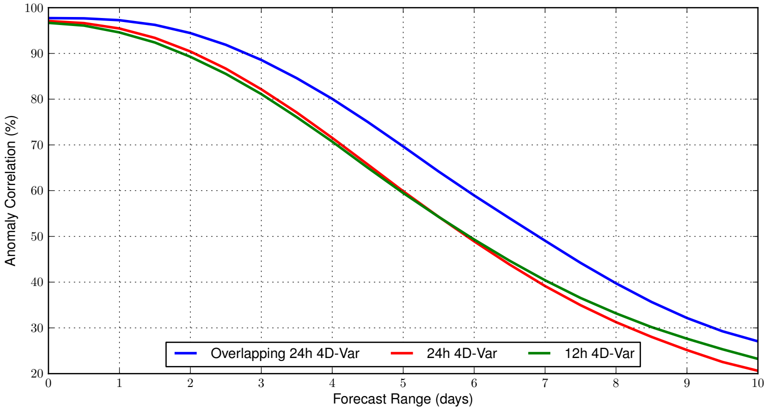
Ps-only Re-analysis



Ps-only Re-analysis

Forecast scores vs. operational analysis

Z500, NH, 2004-07-01 to 2005-04-09



Ps-only Re-analysis

- Verification against independent (unused) observations:
  - confirms positive results with overlapping windows,
  - shows that 24h 4D-Var without overlap is slightly better than 12h 4D-Var.

- 24h 4D-Var system has not been tuned.
  - Results should improve.

- Why is 24h 4D-Var better in Ps-only re-analysis context?
  - Model error is small relative to other errors,
  - Kalman smoother rather than Kalman filter (in part),
  - Not enough observations to fully constrain the analysis in 12h 4D-Var,
  - Full observing system constrains the analysis so tightly that the assimilation algorithm is not as important.

Outline

24h Weak Constraint 4D-Var

- In the current formulation of weak constraints 4D-Var (model error forcing):
  - Background term to address systematic error,
  - 24h assimilation window.
- Observation biases can be an issue.
  - Experiment with bias corrected aircraft observations is starting.
- Investigate physical meaning of model error estimates.
  - For the first time, we might be looking at model error!
- Weak Constraints 4D-Var requires better knowledge of the statistical properties of model error.
- Very good results in Ps-only experiments (re-analysis).
- Kalman smoother is better at least for re-analysis.

Long Window Weak Constraints 4D-Var

- Weak constraint 4D-Var with a 4D state control variable:
  - Four dimensional problem with a coupling term between sub-windows is a smoother over the whole assimilation period.

- Practical implementation is very difficult in current ECMWF system (code, scripts, archiving...).

- We are re-designing our data assimilation system to make it all possible: Object Oriented Prediction System (OOPS).
  - High level algorithms in C++,
  - Improved scalability, reliability, flexibility,
  - New algorithms are implemented (saddle point).