Towards a longer assimilation window in 4D-Var

Yannick Trémolet

Thanks to Paul Poli

ECMWF

October 2011
Outline

1. What we want to do

2. What we can do

3. Results
   - Model Error Aspects
   - 24h Window: Operational System
   - 24h Window: Re-analysis System

4. Final Comments
Outline

1 What we want to do

2 What we can do

3 Results
   - Model Error Aspects
   - 24h Window: Operational System
   - 24h Window: Re-analysis System

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

\[
J(x) = \frac{1}{2} (x_0 - x_b)^T B^{-1} (x_0 - x_b) \\
+ \frac{1}{2} \sum_{i=0}^{n} [H_i(x_i) - y_i]^T R_i^{-1} [H_i(x_i) - y_i] \\
+ \frac{1}{2} \sum_{i=1}^{n} [x_i - M_i(x_{i-1})]^T Q_i^{-1} [x_i - M_i(x_{i-1})]
\]

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):

![](image)
This implementation is an approximation of weak constraint 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 $\mathbf{B}$ is a problem of the past! Only the error characteristics of the fundamental ingredients of the DA problem remain.
Outline

1 What we want to do

2 What we can do

3 Results
   - Model Error Aspects
   - 24h Window: Operational System
   - 24h Window: Re-analysis System

4 Final Comments
In practice, weak constraint 4D-Var is still difficult to implement (in the IFS).

Change of variable:

$$J(x_0, \eta) = \frac{1}{2} \sum_{i=0}^{n} [H(x_i) - y_i]^T R_i^{-1} [H(x_i) - y_i]$$

$$+ \frac{1}{2} (x_0 - x_b)^T B^{-1} (x_0 - x_b) + \frac{1}{2} \sum_{i=1}^{n} \eta_i^T Q_i^{-1} \eta_i$$

with $x_i = M_i(x_{i-1}) + \eta_i$

- $\eta_i$ represents model error in a time step,
- $\eta_i$ has the same dimension as a 3D state.
4D-Var with Constant Model Error Forcing

- Approximation: model error is constant.

\[ J(x_0, \eta) = \frac{1}{2} \sum_{i=0}^{n} [H(x_i) - y_i]^T R_i^{-1} [H(x_i) - y_i] \]
\[ + \frac{1}{2} (x_0 - x_b)^T B^{-1} (x_0 - x_b) + \frac{1}{2} \eta^T Q^{-1} \eta \]

with \( x_i = M_i(x_{i-1}) + \eta \)

- \( \eta \) represents model error in a time step,
- \( \eta \) 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:

\[
J(x_0, \eta) = \frac{1}{2} \sum_{i=0}^{n} [\mathcal{H}(x_i) - y_i]^T R_i^{-1} [\mathcal{H}(x_i) - y_i] \\
+ \frac{1}{2} (x_0 - x_b)^T B^{-1} (x_0 - x_b) + \frac{1}{2} \eta^T Q^{-1} \eta
\]

- For systematic model error:

\[
J(x_0, \eta) = \frac{1}{2} \sum_{i=0}^{n} [\mathcal{H}(x_i) - y_i]^T R_i^{-1} [\mathcal{H}(x_i) - y_i] \\
+ \frac{1}{2} (x_0 - x_b)^T B^{-1} (x_0 - x_b) + \frac{1}{2} (\eta - \eta_b)^T Q^{-1} (\eta - \eta_b)
\]

- Test case: model bias in the stratosphere.
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 $HQH^T$ 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 $P^f = MP^aM^T + Q$?

Model error is correlated in time: $Q$ should account for time correlations. How?

How to account for flow dependence?
Outline

1. What we want to do

2. What we can do

3. Results
   - Model Error Aspects
   - 24h Window: Operational System
   - 24h Window: Re-analysis System

4. Final Comments
Outline

1. What we want to do

2. What we can do

3. Results
   - Model Error Aspects
     - 24h Window: Operational System
     - 24h Window: Re-analysis System

4. Final Comments
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.
Temperature zonal means, December 2010
Temperature zonal means, December 2010

Model error estimates vary rapidly in NH stratosphere.
Outline

1 What we want to do

2 What we can do

3 Results
   - Model Error Aspects
   - 24h Window: Operational System
   - 24h Window: Re-analysis System

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

Y. Trémolet
Long window 4D-Var
October 2011
Outline

1. What we want to do
2. What we can do
3. Results
   - Model Error Aspects
   - 24h Window: Operational System
   - 24h Window: Re-analysis System
4. Final Comments
Ps-only Re-analysis

Background and Analysis fit to Observations

2004-07-01 to 2005-04-09

![Graph showing time (h) vs. Ps observation fit (hPa) with lines for Overlapping 24h 4D-Var, 24h 4D-Var, and 12h 4D-Var.]
Ps-only Re-analysis

Forecast scores vs. operational analysis
Z500, NH, 2004-07-01 to 2005-04-09

Anomaly Correlation (%) vs. Forecast Range (days)

- Overlapping 24h 4D-Var
- 24h 4D-Var
- 12h 4D-Var
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

1 What we want to do

2 What we can do

3 Results
   - Model Error Aspects
   - 24h Window: Operational System
   - 24h Window: Re-analysis System

4 Final Comments
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.
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).