Treatment of model error in 4D-Var

Yannick Trémolet

ECMWF

4D-Var/EnKF workshop - Nov. 2008

Outline

Why account for model error?

After all, 4D-Var works quite well without it...

- Improve the state estimate and/or the forecast,
- Model bias can affect assimilation of some observations (radiance in the stratosphere),
- Typical assumptions in data assimilation are to ignore:
  - Observation bias,
  - Observation error correlation,
  - Model error,
- Long window weak constraint 4D-Var is equivalent to the full rank Kalman filter,
- Learn about the model.

Outline

4D Variational Data Assimilation

4D variational data assimilation is based on 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 relation: $\mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_{i-1})$.
- Used in operational 4D-Var implementations.
- Model $\mathcal{M}$ verified exactly although it is not perfect...

4D Variational Data Assimilation

4D variational data assimilation is based on 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...).
- Do not reduce the control variable using the model.
- Retain 4D nature of the control variable.
- Model $\mathcal{M}$ is not verified exactly: it is a weak constraint.

Control Variable in 4D-Var

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

Outline

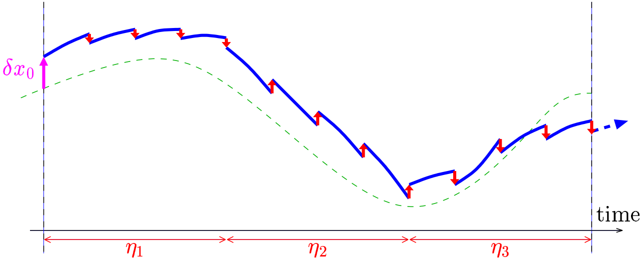
4D-Var with Model Error Forcing

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

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

- $\boldsymbol{\eta}_i$ has the dimension of a 3D atmospheric state,
- $\boldsymbol{\eta}_i$ represents the instantaneous model error,
- $\boldsymbol{\eta}$ is constrained by the fact that it is propagated by the model.

4D-Var with Model Error Forcing



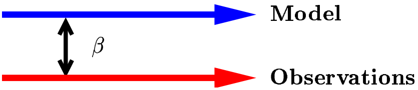
- TL and AD models can be used with little modification,
- Information is propagated between obervations and IC control variable by TL and AD models.

Outline

4D-Var with Model Bias

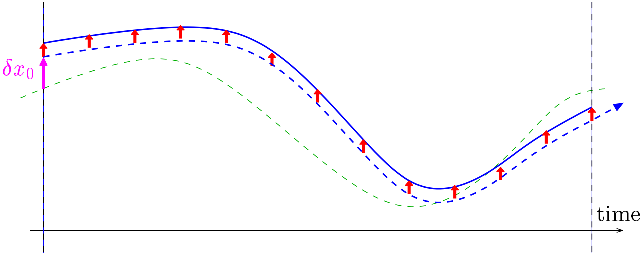
$$J(\mathbf{x}_0, \beta) = \frac{1}{2}\sum_{i=0}^{n}[\mathcal{H}(\mathbf{x}_i^m + \beta_i) - \mathbf{y}_i]^T \mathbf{R}_i^{-1}[\mathcal{H}(\mathbf{x}_i^m + \beta_i) - \mathbf{y}_i]$$
$$+ \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \beta^T \mathbf{Q}_\beta^{-1}\beta$$

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



- β_i is a 3D atmospheric state,
- 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 $\boldsymbol{\beta}$ is slowly varying or constant ($\boldsymbol{\beta}_i = \boldsymbol{\beta}$),
- 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

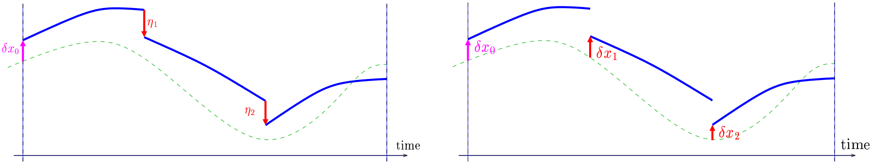
4D State Control Variable

- Use $\{\mathbf{x}_i\}_{i=0,\ldots,n}$ 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



Forcing Control Variable   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.

4D State Control Variable: Properties

- The Hessian of the cost function is the sum of
  - the Hessian of the observation term is $diag(\cdots, \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i, \cdots)$
  - 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

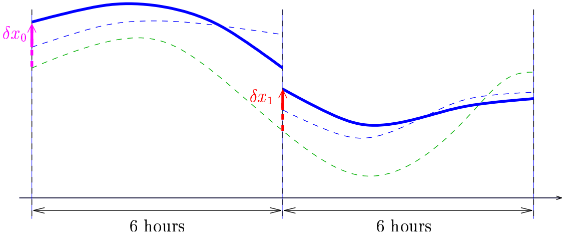
$$J'' = \hat{\mathbf{B}} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$$

- Because of the off-diagonal terms in $\hat{\mathbf{B}}$ the smallest eigenvalue of the system is smaller than 1.
- The condition number increases with the number of sub-windows.
- Using the square root of this tri-diagonal matrix to precondition the minimisation is equivalent to using the initial state and forcing formulation (in the inner loop only!).
- Can we combine the benefits of treating sub-windows in parallel with efficient minimization?

Outline

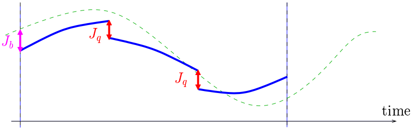
Weak Constraint 4D-Var: Examples

- 6-hour sub-windows:



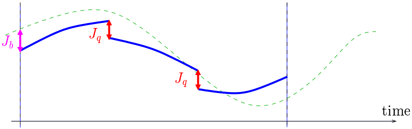
  - Better than 6-hour 4D-Var: two cycles are coupled through J_q,
  - Better than 12-hour 4D-Var: more information (imperfect model), more control,
  - $\mathbf{Q} = \alpha\mathbf{B}$ could be used in that case.
- One time step sub-windows:
  - Each assimilation problem is instantaneous = 3D-Var,
  - 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

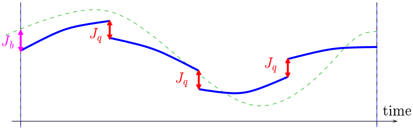


(1) Weak constraint 4D-Var

Weak Constraint 4D-Var: Sliding Window

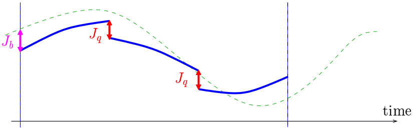


(1) Weak constraint 4D-Var

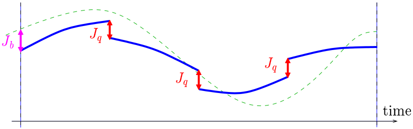


(2) Extended window

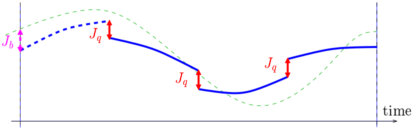
Weak Constraint 4D-Var: Sliding Window



(1) Weak constraint 4D-Var

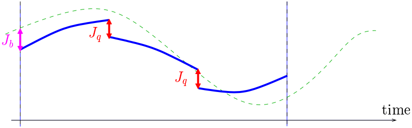


(2) Extended window

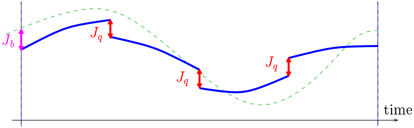


(3) Initial term has converged

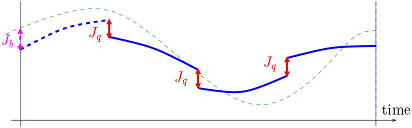
Weak Constraint 4D-Var: Sliding Window



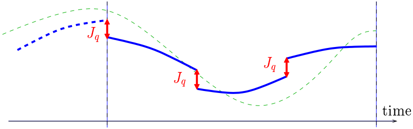
(1) Weak constraint 4D-Var



(2) Extended window

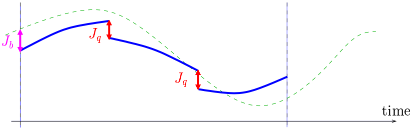


(3) Initial term has converged

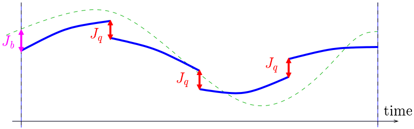


(4) Assimilation window is moved forward

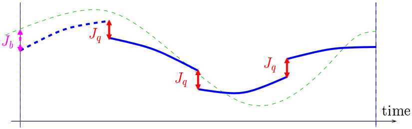
Weak Constraint 4D-Var: Sliding Window



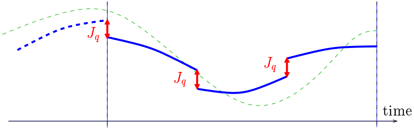
(1) Weak constraint 4D-Var



(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 filter that has been running indefinitely.

Outline

Model error covariance matrix

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

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

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

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

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

- If $\mathbf{Q}$ and $\mathbf{B}$ are proportional, $\delta \mathbf{x}_0$ and $\boldsymbol{\eta}$ 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.

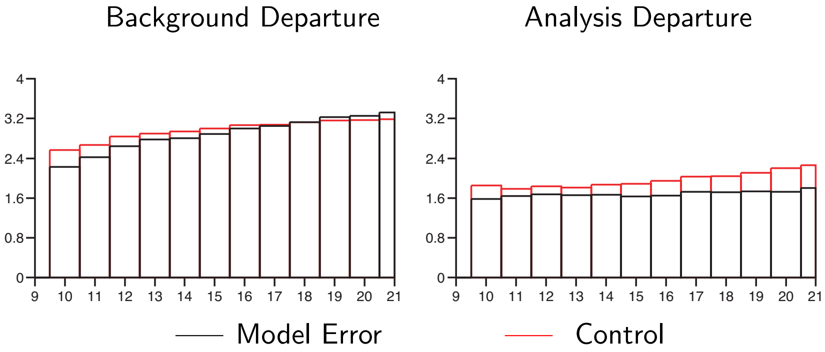
Generating a Model Error Covariance Matrix

- **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**.

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 ?

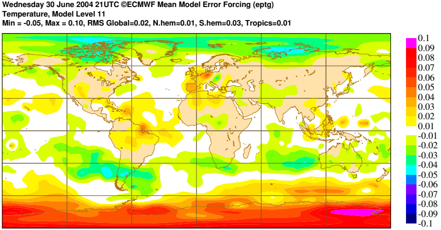
Mean Model Error Forcing

Temperature

Model level 11 (≈5hPa)

July 2004

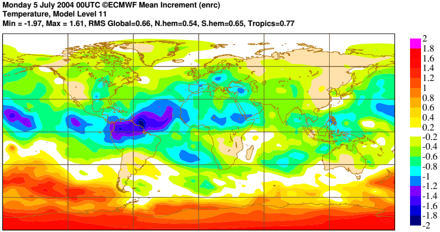
Mean M.E. Forcing ⟶

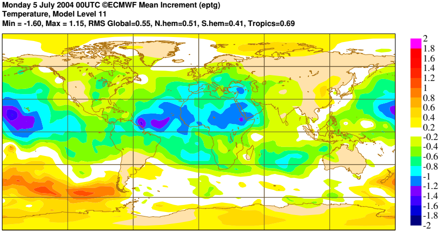


M.E. Mean Increment ↘

Control Mean Increment

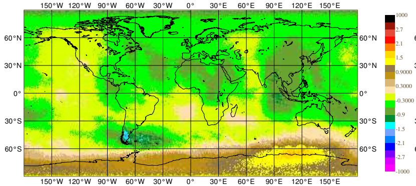
↓



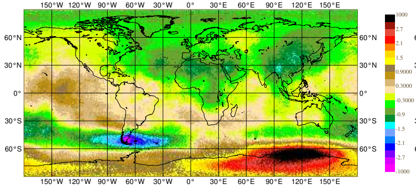


AMSU-A First Guess Departures

STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 13
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070200 - 2004073118 , HOUR = ALL
EXP = ENRC
Min: -1.9618 Max: 2.7 Mean: -0.169506

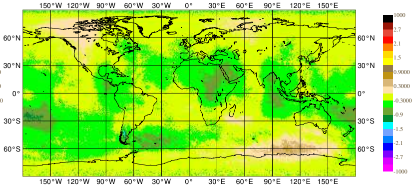


STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 14
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070200 - 2004073118 , HOUR = ALL
EXP = ENRC
Min: -3.3564 Max: 5.46 Mean: 0.006309

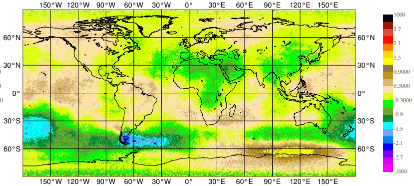


Strong Constraint

STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 13
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070100 - 2004073118 , HOUR = ALL
EXP = EPTG
Min: -1.6688 Max: 0.8 Mean: -0.231773

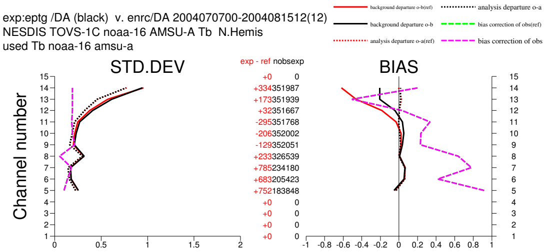


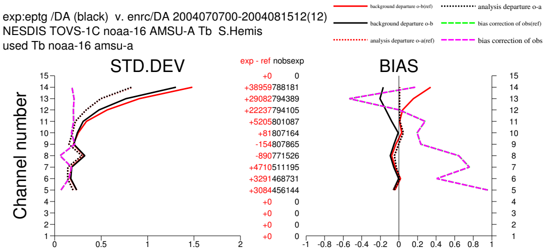
STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 14
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070100 - 2004073118 , HOUR = ALL
EXP = EPTG
Min: -2.6 Max: 2.16 Mean: -0.111883



Weak Constraint

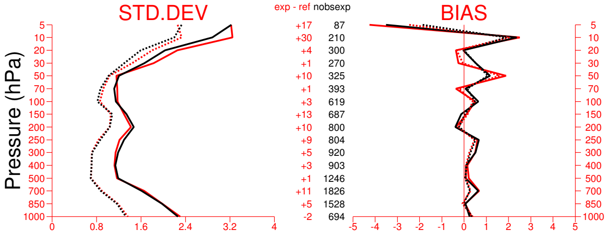
AMSU-A Statistics





- Bias is more uniform,
- BG std. dev. is reduced in SH,
- More data is used.

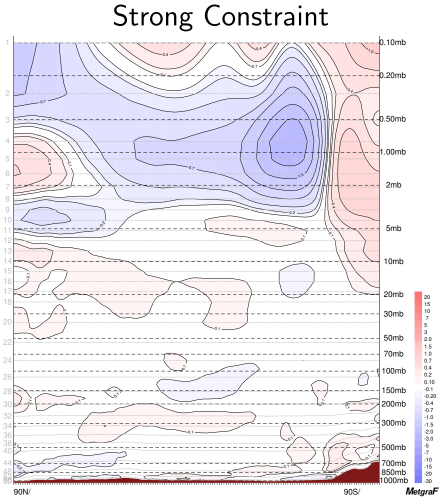
Fit to radiosonde data



- Oscillations in bias are reduced,
- Std. deviation is reduced above 50 hPa (bg and an).

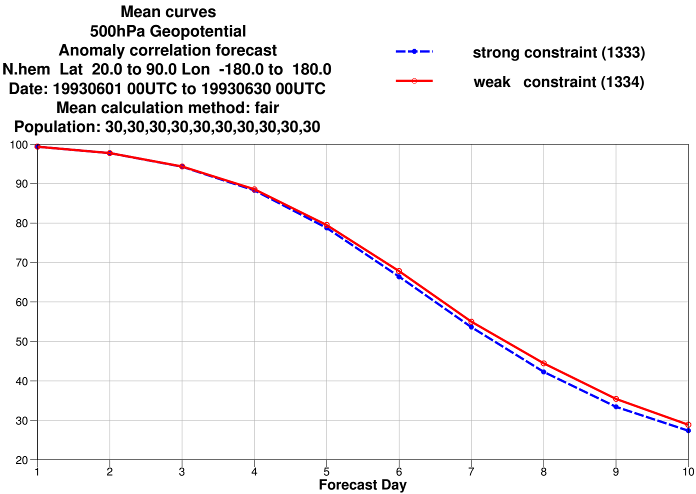
Model error in the stratosphere was reduced by the forcing term.

Weak constraint 4D-Var: Stratosphere



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

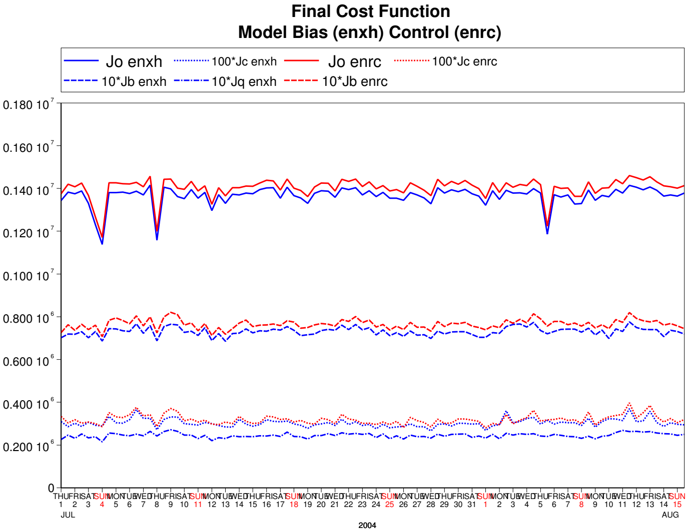
Weak constraint 4D-Var: Performance



Weak constraint 4D-Var with model error forcing term performs well
(Reanalysis, T255, 60 levels, June 1993, no balance in stratosphere)

Outline

Model Bias Experiments



Model Bias Experiments

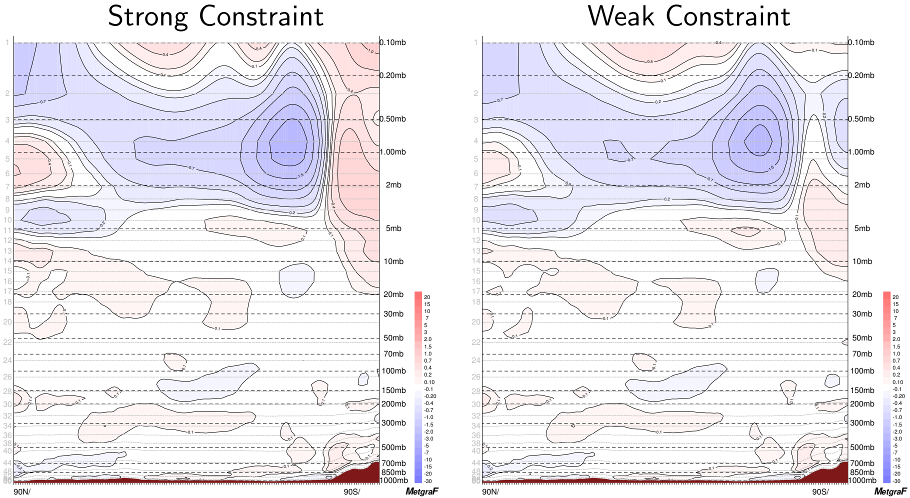
- Fits observations and background better,
- Observation statistics show some improvements and some degradations,
- Might absorb some of the signal,
- $\mathbf{Q}$ should have large scale correlations?

Model Bias Experiments

- Fits observations and background better,
- Observation statistics show some improvements and some degradations,
- Might absorb some of the signal,
- $\mathbf{Q}$ should have large scale correlations?
- Add bias to forecast at post-processing stage,
- We decided to focus on other formulations of weak constraint 4D-Var because of the difficulty to propose products to users.

Outline

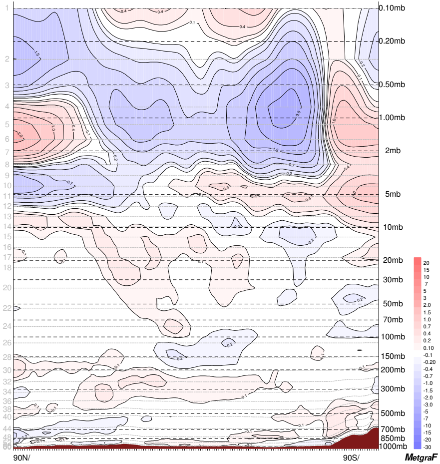
Is it model error?



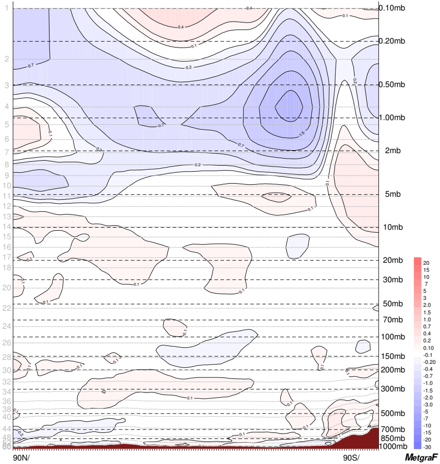
The mean temperature increment is smaller and smoother with weak constraint 4D-Var (June 1993).

Is it model error?

ERA interim

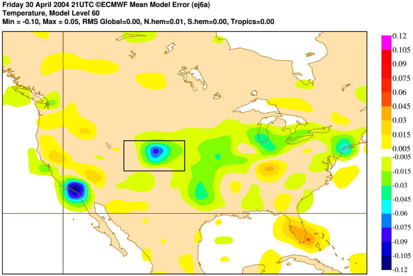


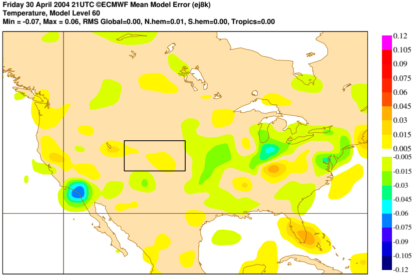
Weak Constraint



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

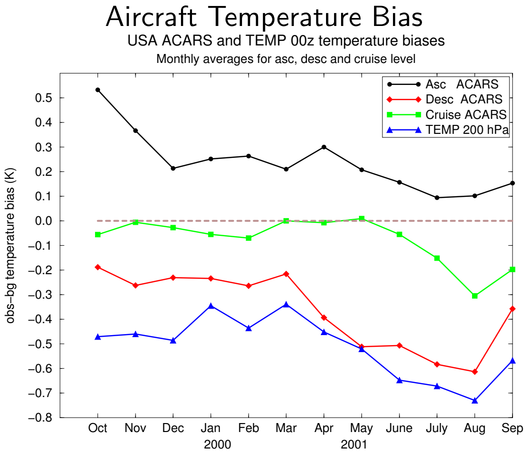
Low Level Mean Model Error Forcing





- 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

Outline

Results with the forcing term

- With constant model error forcing:
  - Fits the data more uniformly over the assimilation window (as expected),
  - Captures model errors (winter stratosphere),
  - Captures observation bias (ascending/descending aircrafts),
  - Captures errors in J_b.
- Two tools to define the errors we wish to capture:
  - Model error covariance matrix (scales, correlations...),
  - Choice of control variable which is a model for model error (constant 3D field has been used).
- Well suited for systematic model errors in (relatively) short windows.
- First target for operational implementation.

Weak Constraint 4D-Var

- Weak constraint 4D-Var with model bias or forcing model error is essentially an initial value problem with parameter estimation (parameters happen to represent model error).
- Weak constraint 4D-Var with a 4D state control variable is a fully four dimensional problem.
- J_q acts as coupling between sub-windows:
  - propagates information across sub-windows,
  - forgetting factor because model equations are not strictly verified.
- Long window weak constraint 4D-Var is equivalent to a full rank Kalman smoother: it is an efficient algorithm to implement it.
- The 4D state control variable is implemented in the IFS (with *sequential* change of variable in the inner loop).
- The first target for implementation could be a 24h window with 2 or 4 sub-windows.

Conclusions and Questions

- Weak constraint 4D-Var has already taught us about observation bias and errors in the balance operators.
- Weak constraint 4D-Var can be adapted to operational analysis and reanalysis.
- 4D-Var can handle correlated model error. What type of correlation model should be used?
- Can we always distinguish model error from observation bias or other errors in the system?
- How do we best determine the model error covariance matrix?
- EnKF has to account for model error: Can we learn from the EnKF approach to the problem?