New covariance statistics of model error for use in weak-constraint 4DVar

Jacky Goddard, Mike Fisher[1]

ECMWF

April 25, 2016

[1]Thanks: Yannick Trémolet, Elías Hólm, Simon Lang, Martin Leutbecher, Lars Isaksen, Massimo Bonavita

Outline

1. Motivation
2. Calculating the Model Error Covariance Matrix
3. Experimentation and Results

Weak-Constraint 4DVar

- Data assimilation for NWP has reached a level of accuracy where model errors can no longer be neglected.
- Taking model error into account in 4D-Var requires that we specify a covariance matrix for model error.

$$
\begin{aligned}
J(\mathbf{x}) = \frac{1}{2}(x_0 - x_b)^{\mathrm{T}}\mathbf{B}^{-1}(x_0 - x_b) \\
+ \frac{1}{2}\sum_{k=0}^{N}(\mathcal{H}_k(x_k) - y_k)^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathcal{H}_k(x_k) - y_k) \\
+ \frac{1}{2}\sum_{k=0}^{N}(\mathcal{M}(x_{k-1}) - x_k)^{\mathrm{T}}\mathbf{Q}_k^{-1}(\mathcal{M}(x_{k-1}) - x_k)
\end{aligned}
$$

Weak-Constraint 4DVar with model error forcing

$$J(\mathbf{x}_0, \eta) = \frac{1}{2}(x_0 - x_b)^{\mathrm{T}}\mathbf{B}^{-1}(x_0 - x_b) + \frac{1}{2}\sum_{k=0}^{N}(\mathcal{H}_k(x_k) - y_k)^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathcal{H}_k(x_k) - y_k) + \frac{1}{2}\eta^{\mathrm{T}}\mathbf{Q}^{-1}\eta$$

- with $x_k = \mathcal{M}(x_{k-1}) + \eta_k$
- η_k is propagated by the model
- η_k represents the instantaneous model error

Weak-Constraint 4DVar with cycling term

- Model error is both random and systematic.
- For the systematic part the cost function is:

$$J(\mathbf{x}_0, \eta) = \frac{1}{2}(x_0 - x_b)^{\mathrm{T}}\mathbf{B}^{-1}(x_0 - x_b) + \frac{1}{2}\sum_{k=0}^{N}(\mathcal{H}_k(x_k) - y_k)^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathcal{H}_k(x_k) - y_k) + \frac{1}{2}(\eta - \eta_b)^{\mathrm{T}}\mathbf{Q}^{-1}(\eta - \eta_b)$$

- η_b is like a background to the model error
- In the following experiments a constant forcing over the assimilation window is used.

Weak-Constraint 4DVar

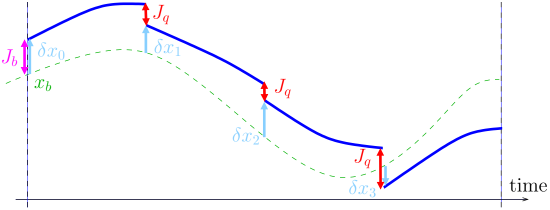


Figure: longwindow.

- 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.

EPS Experiment

- 50 member ensemble + control
- T_L 399 resolution
- 12 hour forecast
- Cycle 40R3
- 20 days of forecasts: 2013083100 - 2013091900
- Identical initial conditions (ensemble members are not perturbed)
- Stochastic parametrisation SPPT and SKEB

Methodology

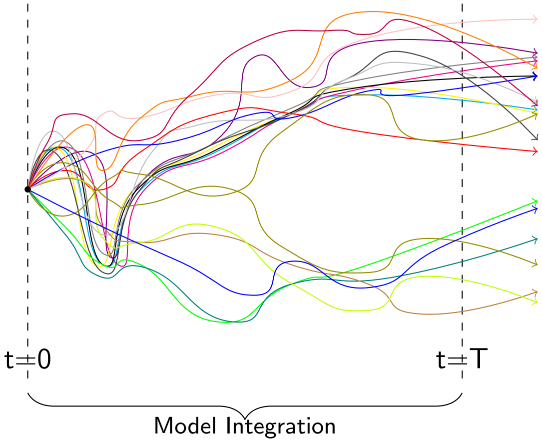
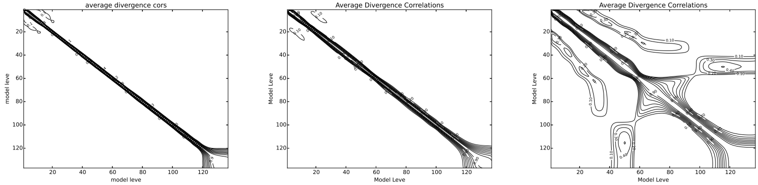


Figure: Cartoon of EPS members with identical initial conditions but different realisations of model error. In this experiment T is chosen to be 12 hrs because this is the length of the 4DVAR assimilation window.

Average Divergence Correlation



(a) Background (b) SPPT & SKEB **Q** (c) 4D-Var Model Error Estimation

Figure: Comparison between background, EPS model error at 12 hrs and 4D-Var Model Error Estimation. Maximum off diagonal correlation contour 0.5.

Choice of αQ

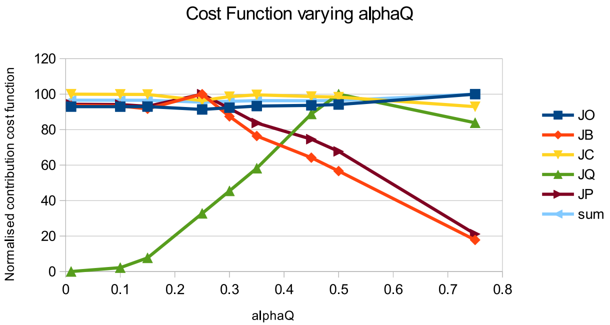


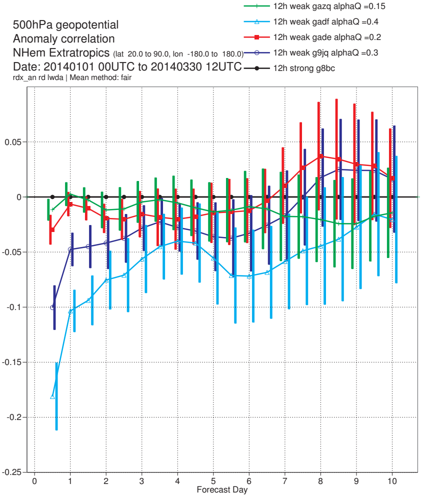
Figure: Normalised cost function contributions

- choose to experiment with $\alpha = 0.15, 0.2, 0.3$ and 0.4
- We want the Q term to be significant but not to dominate over other terms

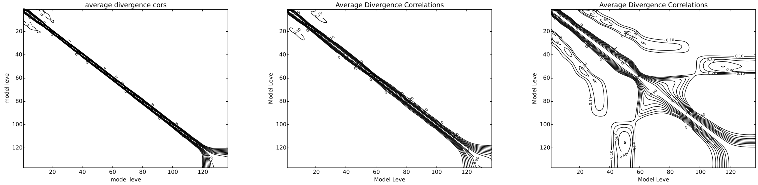
Weak-constraint 4DVar experimentation:

- 10 day forecast
- 12 hour weak constraint 4DVar with model error forcing
- $\alpha = 0.15, 0.2, 0.3$ and 0.4
- 3 months JFM 2014
- Using new stochastic $\mathbf{Q}$ matrix
- CY41R1

Scores - against operational analysis



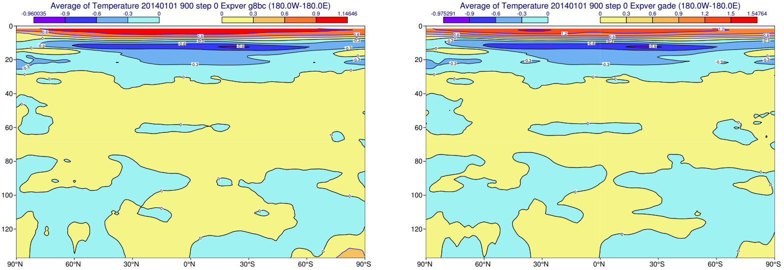
Average Divergence Correlation (Repeat slide)



(a) Background (b) SPPT & SKEB **Q** (c) 4D-Var Model Error Estimation

Figure: Comparison between background, EPS model error at 12 hrs and 4D-Var Model Error Estimation. Maximum off diagonal correlation contour 0.5.

Zonal Means of Analysis Increment and Estimated Model Error Forcing



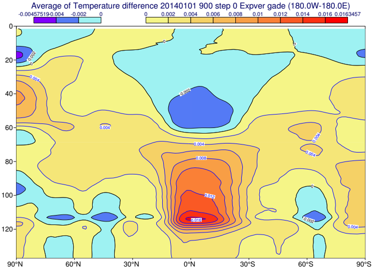
(a) Analysis Increment Zonal Mean (Strong)

(b) Analysis Increment Zonal Mean (Weak)

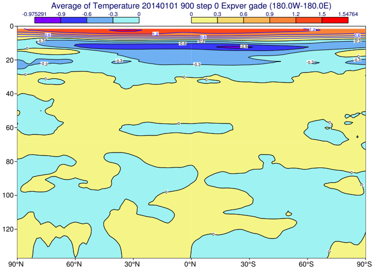
Figure: Fig (a) Analysis increment for strong constraint min value -0.96 K, max value 1.15 K, fig (b) Analysis increment for weak constraint min value -0.98 K, max value 1.55 K.

ECMWF

Zonal Means of Analysis Increment and Estimated Model Error Forcing



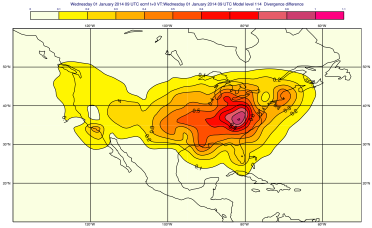
(a) Model Error Zonal Mean (Weak)



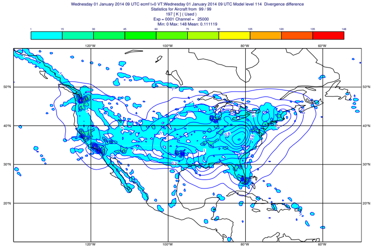
(b) Analysis Increment Zonal Mean (Weak)

Figure: Fig (a) Model error zonal mean for weak constraint min value -0.0046 K/hr, max value 0.0163 K/hr, fig (b) Analysis increment for weak constraint min value -0.98 K, max value 1.55 K.

Misinterpretation of AIREP Data



(a) Divergence Covariances



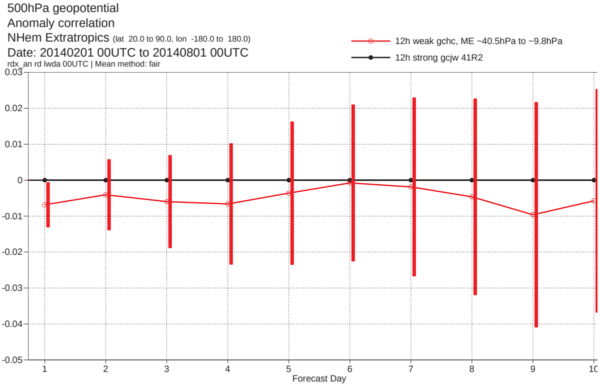
(b) AIREP 250hPa - 500hPa

Figure: (a) Divergence covariances between level 52 and 114 over the USA (multiplied by 10E14) (b) AIREP temperature data (averaged over January 2014) overlaid with (a).

Misinterpretation of AIREP Data

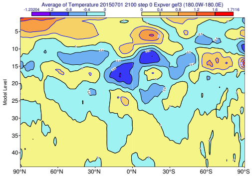
- Observation errors misinterpreted as model error
- To avoid erroneous aliasing of errors restrict model error forcing to above $\sim$ 40.5 hPa

Scores - against own analysis

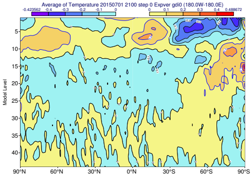


ECMWF

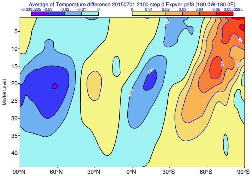
Operational resolution CY41R2



(b) Weak-constraint analysis inc



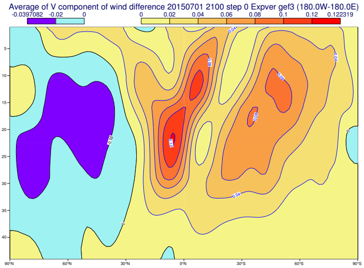
(c) strong-weak analysis inc



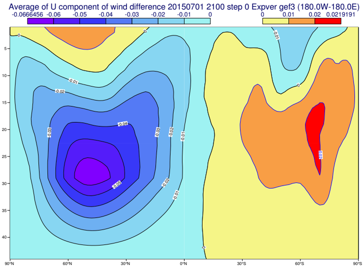
(d) Model error forcing

Figure: Temperature analysis increments and model error forcing comparison - July 2015

Operational resolution CY41R2



(a) V wind

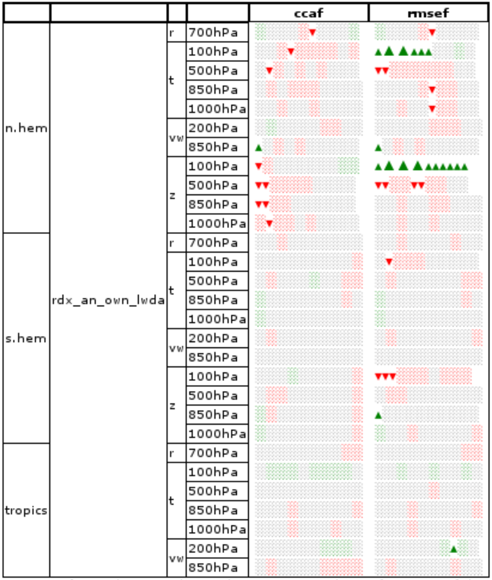


(b) U wind

Figure: Model error forcing $\mathrm{ms}^{-1}\mathrm{hr}^{-1}$

ECMWF

Operational resolution CY41R2 -MJJA own analysis



Conclusions

- Overall results are fairly neutral with some positives above 100hPa
- RMS forecast error is being shifted above 100hPa
- Model error forcing is trying to fix large scale circulation errors in the stratosphere
- Very difficult to verify results but gpsro verification may be an option

. . . Any Questions?