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

Jacky Goddard, Mike Fisher

ECMWF

April 25, 2016

\footnote{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
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.

\[
J(x) = \frac{1}{2}(x_0 - x_b)^T B^{-1}(x_0 - x_b)
\]

\[
+ \frac{1}{2} \sum_{k=0}^{N} (\mathcal{H}_k(x_k) - y_k)^T R_k^{-1} (\mathcal{H}_k(x_k) - y_k)
\]

\[
+ \frac{1}{2} \sum_{k=0}^{N} (\mathcal{M}(x_{k-1}) - x_k)^T Q_k^{-1} (\mathcal{M}(x_{k-1}) - x_k)
\]
Weak-Constraint 4DVar with model error forcing

\[
J(x_0, \eta) = \frac{1}{2} (x_0 - x_b)^T B^{-1}(x_0 - x_b) + \frac{1}{2} \sum_{k=0}^{N} (H_k(x_k) - y_k)^T R_k^{-1}(H_k(x_k) - y_k) + \frac{1}{2} \eta^T Q^{-1} \eta
\]

- with \( x_k = M(x_{k-1}) + \eta_k \)
- \( \eta_k \) is propagated by the model
- \( \eta_k \) 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:

\[
J(x_0, \eta) = \frac{1}{2}(x_0 - x_b)^T B^{-1}(x_0 - x_b) \\
+ \frac{1}{2} \sum_{k=0}^{N} (\mathcal{H}_k(x_k) - y_k)^T R_k^{-1}(\mathcal{H}_k(x_k) - y_k) \\
+ \frac{1}{2}(\eta - \eta_b)^T Q^{-1}(\eta - \eta_b)
\]

- \( \eta_b \) is like a background to the model error
- In the following experiments a constant forcing over the assimilation window is used.
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
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

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 $\alpha Q$

![Cost Function varying alphaQ](image)

**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 $Q$ matrix
- CY41R1
Scores - against operational analysis

500hPa geopotential
Anomaly correlation
NHem Extratropics (lat 20.0 to 90.0, lon -180.0 to 180.0)
Date: 20140101 00UTC to 20140330 12UTC

rdx_an rd lwda | Mean method: fair

12h weak g8bc
12h weak g9jq alphaQ =0.3
12h weak gade alphaQ =0.2
12h weak gadf alphaQ =0.4
12h weak gazq alphaQ =0.15

Forecast Day

rdx_an rd lwda | Mean method: fair

0 1 2 3 4 5 6 7 8 9 10

Forecast Day

Jacky Goddard (ECMWF)
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

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.
Zonal Means of Analysis Increment and Estimated Model Error Forcing

(a) Model Error Zonal Mean (Weak)

(b) Analysis Increment Zonal Mean (Weak)

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

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

500hPa geopotential
Anomaly correlation
NHem Extratropics \((\text{lat } 20.0 \text{ to } 90.0, \text{ lon } -180.0 \text{ to } 180.0)\)
Date: 20140201 00UTC to 20140801 00UTC

rdx_an rd lwda 00UTC | Mean method: fair

12h strong gcjw 41R2
12h weak gchc, ME \(\sim 40.5\text{hPa to } \sim 9.8\text{hPa}\)

Jacky Goddard (ECMWF)
Operational resolution CY41R2

Figure: Temperature analysis increments and model error forcing comparison - July 2015
Figure: Model error forcing $\text{ms}^{-1}\text{hr}^{-1}$
### Operational resolution CY41R2 - MJJA own analysis

<table>
<thead>
<tr>
<th></th>
<th>ccf</th>
<th>msef</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n.hem</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r 700hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>100hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>500hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>850hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>1000hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td><strong>s.hem</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r 700hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>t 100hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>t 500hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>t 850hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>t 1000hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td><strong>tropics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r 700hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>t 100hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>t 500hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>t 850hPa</td>
<td>▼</td>
<td></td>
</tr>
<tr>
<td>t 1000hPa</td>
<td>▼</td>
<td></td>
</tr>
</tbody>
</table>

**Legend**
- ▼: Positive correlation
- ▲: Negative correlation
- ▼: Better performance
- ▲: Worse performance

---

**Source:** ECMWF
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?