Accounting for Model Error in 4D-Var

Yannick Trémolet

ECMWF
Shinfield Park, Reading, UK
yannick.tremolet@ecmwf.int

1 Introduction

The goal of data assimilation is to estimate the state of a system given all available information about that system. In practice, two sources of information are available: observations of the system and theoretical knowledge represented by a model.

Through cycling of the data assimilation system we introduce an auxiliary source of information: the background. It represents the knowledge of the system given the model and all past observations.

Uncertainty in the two main sources of information (observations and model) should in theory be accounted for so that the best data assimilation can do is determine the best estimate of the state of the system given the model, the observations and their error characteristics. In operational practice though, model error is difficult to take into account. It is most of the time hidden in the background error term which, although not strictly an input of the data assimilation system, makes data assimilation practical. We present below an approach to account for model error in 4D-Var and some results in the IFS.

Model error has bias and random components. In this paper, model errors that vary on a timescale longer than an assimilation window will be called systematic error while model errors that vary on a timescale shorter than an assimilation window will be called random error.

2 Weak constraint 4D-Var

Model error at a given time t_i is defined as:

$$\mathbf{q}_i = \mathbf{x}_i^t - \mathcal{M}(\mathbf{x}_{i-1}^t)$$

where $\mathbf{x}_i^t$ and $\mathbf{x}_{i-1}^t$ represent the (unknown) true state of the system at time t_i and t_{i-1} respectively. This definition is very general and does not rely on any assumption regarding the form of model error: if the true state is known, it can always be computed. However, in real systems, the true state is unknown and model error cannot be computed.

2.1 Formulation

Based on a maximum likelihood formulation, 4D-Var comprises the minimisation of a cost function to estimate the state of a system over a period of time called the assimilation window. In most implementations of 4D-Var, the four dimensional unknown state is reduced to a three dimensional state at the begining of the assimilation window, relying on the model to compute the states at other times from the initial condition. This makes 4D-Var possible in practice at the cost of assuming that the model is perfect over the length of the assimilation window.

Weak constraint 4D-Var does not rely on this assumption, keeping the four dimensional state as the unknown variable to be estimated. For Gaussian temporally-uncorrelated model error, the weak constraint 4D-Var cost function is:

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

where $\mathbf{x}_b$ is the background state, $\mathbf{B}$ is the background error covariance matrix, $\mathcal{H}$ is the observation operator, $\mathbf{R}$ is the observation error covariance matrix, $\mathcal{M}$ is the forecast model, $\mathbf{Q}$ is the model error covariance matrix and J_c represents additional constraints (digital filter for example). More details are given for example by Trémolet (2006). This formulation of 4D-Var accounts for the fact that the model contains information but is not exact by adding a model error term to the cost function.

2.2 Systematic model error

A fully four dimensional control variable implementation of 4D-Var, as described above, is not possible at the moment in the IFS. Several scientific and technical challenges must be addressed before reaching that goal, and simplifications are still necessary. For practical reasons, in particular in implementing the outer loop iterations, the weak constraint 4D-Var implementation in the IFS is limited to the "forcing" formulation, with model error assumed to be constant over the entire analysis window. With this simplification, the cost function becomes:

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

with $\mathbf{x}_i=\mathcal{M}_i(\mathbf{x}_{i-1})+\mathbf{q}$ and where $\mathbf{q}$ is now independent of i.

The assumption that model error is constant over a certain period of time makes weak constraint 4D-Var achievable. It also means that it is mainly the systematic component of model error, or model bias, that is estimated. An additional assumption is that systematic model error, or model bias, varies slowly from one data assimilation cycle to the next. The cost function is thus modified and re-written:

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

where $\mathbf{q}_b$ is the model error estimate from the previous assimilation cycle. Model bias is allowed to evolve slowly at a rate determined by the covariance matrix $\mathbf{Q}$.

3 Model error covariance matrix

A lot of research has gone into estimating the background error covariance matrix at all operational centres but almost nothing is known about the model error covariance matrix that appears in the weak constraint 4D-Var cost function and in the Kalman filter equations.

Indirectly related to the data assimilation context, some research has taken place to diagnose forecasting models. For example, some studies have recognised the link between systematic initial tendencies and model error for some time (see, e.g., Klinker and Sardeshmukh (1992)) and used it to assess climate models (Rodwell and Palmer (2007)) or to demonstrate an improvement in model aerosol climatology (Rodwell and Jung (2008)).

Based on short range model tendency statistics or longer range model drift statistics, two types of model error covariance matrices have been used so far in weak constraints 4D-Var experiments in the IFS, overall with similar results.

3.1 Model tendency covariances

The first model error covariance matrix to be used with the IFS is described by Trémolet (2007). In the current ECMWF data assimilation system, the background error covariance matrix **B** is estimated from an ensemble of 4D-Var. Considering the ensemble of forecasts run from the analyses produced by the 4D-Var ensemble, at any given step, the ensemble of model states represents a proxy for the probability distribution for the true atmospheric state. The model tendencies derived from the states of ensemble members should represent a distribution of the possible evolutions of the atmosphere from the true state. The differences between these tendencies can be interpreted as possible uncertainties in the model forcing or an ensemble of possible realisations of model error. This is the basis for constructing a model error covariance matrix.

This set of model error realisations can be fitted to a statistical model similar to the one used to represent **B**. The statistical model used here is isotropic, homogeneous and non separable. It is the same model (and code) that was used to specify the background error covariance matrix for the operational strong constraint 4D-Var at ECMWF until April 2005.

To determine the model error covariance matrix, forecasts were run from the 4D-Var ensemble at the spectral resolution of T319 and tendencies saved after 12h, 18h, 24h and 30h. Four tendencies spread over 24 hours were saved in order to avoid any diurnal cycle signal in the statistics and to increase the sample size. The first hours of the forecast were also avoided to reduce the dependence on the initial state and spin-up issues. The 4D-Var ensemble which was used had 10 members and was available for 26 days. This gave 936 realisations of model errors to be used as inputs to the statistical model.

3.2 Model drift covariances

The second model error covariance matrix to be used in the IFS is based on samples of model drift and variations in model drift. As explained above, the aim of the current weak constraints 4D-Var formulation is to capture systematic model error. One visible effect of systematic model error is model drift: forecasts at different ranges do not have the same average characteristics. Seasonal averages of differences between day-5 and day-10 forecasts are for example sometimes used by model developers to diagnose systematic problems in the model.

In the formulation proposed here, the goal is also to capture this systematic error. The definition of the model error term in the cost function shows that **Q** should account for the statistical properties of the variation of model error from one assimilation cycle to the next. The easiest proxy to access for that quantity is the change in model drift from one day to another. Only one realisation of model drift is available for each forecast which does not constitute a proper statistical sample. However, since this is available for every forecast, statistics over a long enough period of time can be used, as was the case for background error in many operational systems until recently with the “NMC” method.

The other main drawback with this method is the fact that forecast differences depend on model error

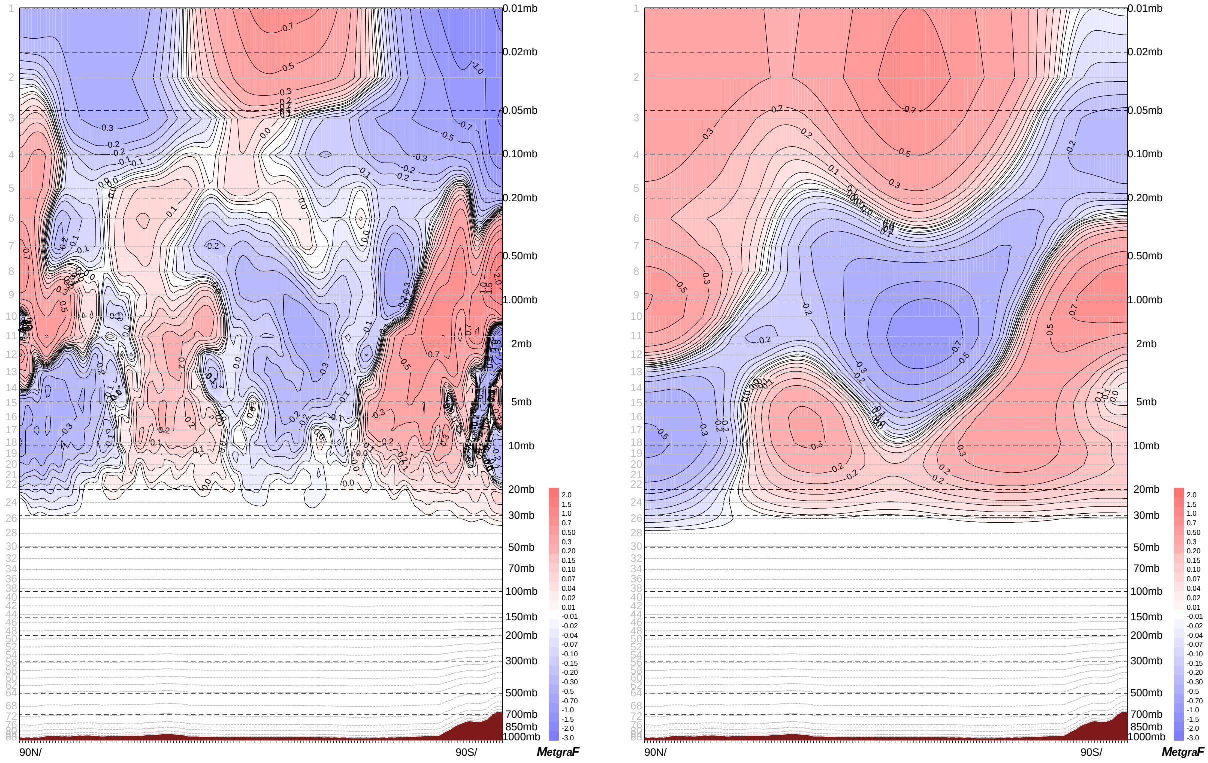

Figure 1: Zonal mean of average temperature model error for the month of August 2010 with the tendency-based model error covariance matrix (left) and the drift-based model error covariance matrix (right).

and initial condition. The assumption being made here is that at long enough ranges, the contribution from model error is larger than the contribution from the initial condition error. In the experiments presented here, the model drift samples were obtained using day-5 minus day-10 forecasts differences over a period of one year from June 2009 to May 2010. This sample is used to generate a covariance matrix using the same statistical model as above.

3.3 Comparison results

Figure 1 shows the impact of the two model error covariance matrices on the model error estimate. It is clear that tendency-based covariances lead to smaller scales structures in the model error field. This is expected as tendencies are instantaneous small scale responses from the model to a given state and lead to shorter correlation length scales than model drift which is an accumulated response over several days.

The model error estimates show some similarities, for example a large scale cooling from the top of the model at the South pole down to around 5hPa above the equator, with warming above and below that. In the Northern hemisphere, the two model error estimates are quite different, especially above 2hPa. Despite these differences, the two systems lead to very similar forecast performances and it is very difficult to determine which covariance matrix is better. A smoother estimate might be preferable and more realistic if longer term systematic model error is sought.

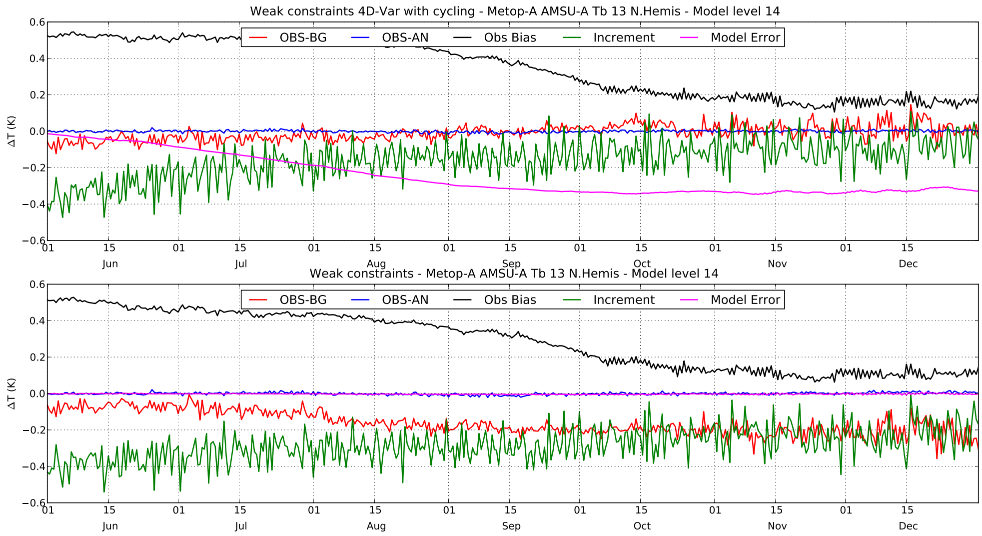

Figure 2: Time-series of mean background and analysis fit to observations, observation bias correction, analysis increment and model error without (top) and with (bottom) model error cycling. The analysis increment and model error statistics are computed at model level 14 where AMSU-A channel 13 has the highest sensitivity.

4 Experimental results

Since there is evidence of model bias in the IFS, especially in the stratosphere, it seems reasonable to address this aspect first. All results presented in this section were obtained with IFS cycle CY37R3 which includes the model bias term described above, but with model bias restricted to be non-zero only in the stratosphere. The experiments were run at T255 with two minimisations at T95 and T159, using the full operational observing system. The results of section 4.1 were obtained with the tendency-based model error covariance, other results were obtained with the drift-based model error covariance matrix.

4.1 Model error cycling

Figure 2 shows a time-series of AMSU-A channel 13 statistics for the Northern hemisphere and analysis increment and model error statistics for the model level where this channel is most sensitive. Without cycling, *i.e.* $\mathbf{q}_b$ is set to zero every cycle, the mean background departure and analysis increment stay negative throughout the period. When model error is cycled, the average background departures stays around zero and the average analysis increment is also reduced, showing that weak constraint 4D-Var can in that case compensate for some of the model systematic error. The mean increment is still not zero even with model error cycling because of the presence of AMSU-A channel 14 which is not bias corrected and to which the level shown has some sensitivity. However, as model error is used only in the stratosphere, the impact on forecast performance is limited.

4.2 Model error estimates

Figure 3 gives an overview of the temperature model error and initial condition increment estimates as zonal mean cross sections for one month (December 2010). In all experiments presented here, model error is estimated within each assimilation cycle, and is used as a first guess and "background" ($\mathbf{q}_b$)

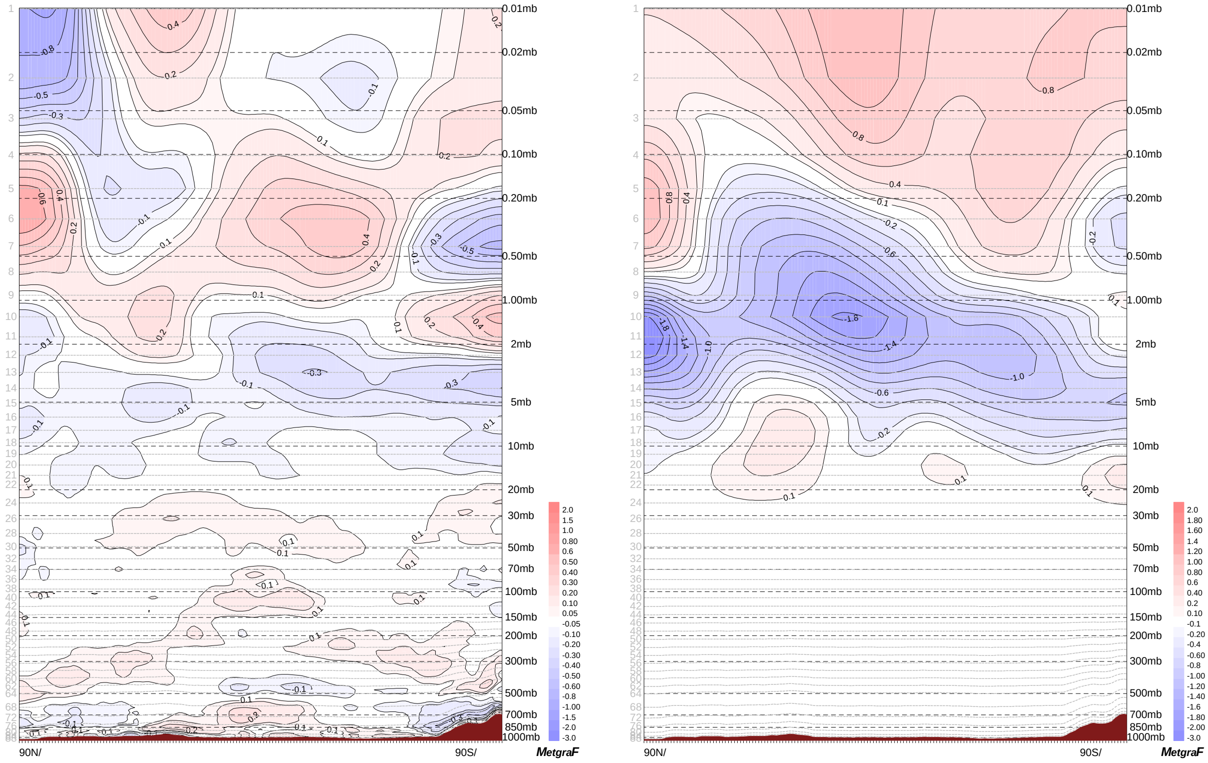

Figure 3: Zonal mean of the average temperature initial condition increment (left, in K) and model error (right, in K/day) for the month of December 2010.

value for the next cycle, but it is not applied as a forcing term during the forecasts. Such experiments have been performed but always lead to poor performance. In principle, if the estimate was an accurate representation of the mean slowly varying model error, applying during the forecast should improve it. However, close examination shows that there is a relatively large day to day variability in the model error estimate which makes it very unlikely that model error determined on a particular day will be meaningful a few days later in the forecast. This most likely explains the poor performance of the forced forecasts.

4.3 Interaction with observation biases

It was shown by Trémolet (2007) that observation biases can affect model error estimates, and stationary differences between the model and the observations are likely to be interpreted as model errors in a weak constraint 4D-Var system. Figure 4 is similar to the top panel of figure 2 but for a different period. It shows an episode of sudden stratospheric warming in the second half of January 2009. The model is not able to produce that event that the observations capture. However, 4D-Var adjusts the observation bias correction (black curve) and initial condition increments (green) but not model error.

As model error can compensate for observation biases, observation bias correction can compensate for model biases. Since there are no additional sources of information to distinguish between sources of biases, additional assumptions will be needed to attribute biases correctly.

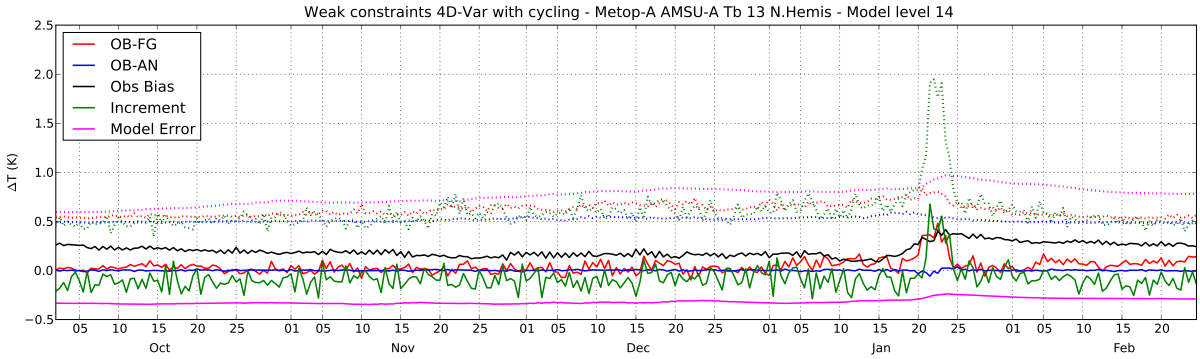

Figure 4: Time-series of mean background and analysis fit to observations, observation bias correction, analysis increment and model error (plain lines represent mean values, dashed lines represent standard deviations).

5 Discussion

All the results obtained with the weak constraint 4D-Var implemented in the IFS show some encouraging signs but also point at weaknesses in the system.

One main area for improvement is the specification of the model error covariance matrix. Two approaches have been tested at ECMWF, based on model tendencies statistics and on model drift statistics. Both of these methods could be improved upon by using the EDA and the EPS to gather more samples of model tendencies and model drift, possibly leading to the use of a flow dependent component in the **Q** matrix.

One on-going direction for research in this area aims at determining a projection of the model error covariance matrix into observation space. This could potentially allow comparison of the model error covariance matrices used at ECMWF with a more rigorous statistical estimation. Unfortunately, this projection into observation space is limited as it requires a fixed observing network which is the case only for a very limited subset of the global observing network. This projection cannot directly provide a model error covariance in model space but can be used for diagnostics. We are not aware of any other attempt to estimate model error statistics although this is clearly an area where much more research is needed.

Another important area for improvement is the formulation of the weak constraint 4D-Var problem. Currently, the assumption of a constant model error forcing is a strong limitation. A proper representation of the random component of model error is also needed and a better distinction between the methods used to estimate the systematic and random components of model error will be important. Only a proper formulation of weak constraints 4D-Var will bring all its expected benefits.

Data assimilation is where the model is systematically confronted with reality through the observations. That makes it the best place to diagnose the model and get estimates of model errors, using weak constraints 4D-Var or other techniques. All sources of model error are diagnosed: convection but also other physical parametrisations, interactions between them, numerical schemes, even bugs... In addition, other sources of error are entangled with model error in a real system: observation errors, errors in the data assimilation system itself, interactions with observation bias correction... All this makes the interpretation of model error estimates difficult and more research will be needed before it can be fully understood and ultimately feed back into improving the model.

References

E. Klinker and P. D. Sardeshmukh. The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.*, 49:608–627, 1992.

M. J. Rodwell and T. Jung. Understanding the local and global impacts of model physics changes: An aerosol example. *Q. J. R. Meteorol. Soc.*, 134:1479–1497, 2008.

M. J. Rodwell and T. N. Palmer. Using numerical weather prediction to assess climate models. *Q. J. R. Meteorol. Soc.*, 133:129–146, 2007.

Y. Trémolet. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.*, 132:2483–2504, 2006.

Y. Trémolet. Model error estimation in 4D-Var. *Q. J. R. Meteorol. Soc.*, 133:1267–1280, 2007.