ECMWF/WWRP Workshop on Model Uncertainty

Working Group 2. "How can we improve the diagnosis of model error?"

Chair: Heini Wernli **Rapporteur**: Mark Rodwell

Group members: Craig Bishop, Neill Bowler, Steven Cavallo, Shuyi Chen, Mike Cullen, Mike Fisher, Jacky Goddard, Heikki Haario, Trond Iversen, Heikki Jarvinen, Christian Keil, Simon Lang, Amos Lawless, Glen Romine, Peter Watson

1. Introduction

The working group discussed the questions in sections 2-5 below. The write-up summarises these discussions and gives some recommendations for future research.

2. How can we better use observations/models to diagnose model error?

Develop and refine Weak Constraint 4DVAR, observation error covariance (R) estimation and innovation/residual based "Desroziers/Todling" diagnostics for estimating the mean and covariance of model errors. Weak constraint 4DVAR provides direct estimates of individual, flow-specific model error realizations, which should be studied. Several techniques for model-error estimation presented during the workshop require weak-constraint 4DVAR and/or the gradient and adjoint of the model and analysis. Hence, the maintenance of advanced adjoint based methods is advised. Toddling estimates of model error covariance require accurate estimates of observation error covariances so accurate techniques for doing this need to be developed (See Craig Bishop's presentation to this workshop).

Consider running idealised experiments with prescribed plausible model error forcing to assess what aspects of the mean and covariance of the distribution of model errors could be recovered from simulated observations that resemble today's observational network and, perhaps, what changes or field campaigns could be recommended to improve its ability to define the mean and covariance of model error. Consider using Intensive Observing Periods from previous field campaigns for estimating model error. We recommend that Field Campaigns make it as easy as possible for their special observations to be used in this way. Experiments with prescribed plausible forcing to assess the quality of estimates of individual model error realizations would also be useful.

Assess the extent to which an ensemble of weak-constraint 4DVAR based error recovery techniques would allow flow-dependent estimates of model error covariance to be derived.

Use high resolution models and coarse graining to improve models of model error (e.g. SPPT, SKEB) and then use observation based model error estimation techniques to better define the variances and correlations of the stochastic fields used in these schemes.

Improve the feedback loop between model error detection and model improvement.

Model parameters are known to contain uncertainties within a given model structure. Existing forecasting systems should be used to diagnose parametric uncertainties. It is established that algorithmic methods can train the model to the desired target to improve deterministic skill and provide density estimates for stochastic schemes based on parameter perturbation. Special attention should be paid to the formulation of the multi-criteria optimization targets. (Laine et al., 2012; Ollinaho et al., 2013, 2014).

3. What length forecast range is necessary to diagnose the main sources of model error? What are the relative roles for assimilation and forecast systems in identifying model error?

There was a consensus that short forecast ranges are very useful to diagnose model error, and hence there is a natural coming-together of data-assimilation and forecasting techniques. However, issues associated with the very first forecast timestep having a different structure to the subsequent timesteps can complicate the identification of model error and the attribution of its systematic component. When forecasts are initialized with "alien" analyses, spin-up issues can further obscure the model error (Klocke and Rodwell, 2014) and so, for such diagnoses, it is important to initialize a forecast from an analysis produced with the same model.

Some coupled processes (associated with the ocean, land-surface, sea-ice etc.) may be too slow to be seen at atmospheric data assimilation timescales but can later lead to large systematic errors. Assessment of the systematic and random aspects of such errors must, therefore, involve longer timescales, with good short-range forecast reliability being an important pre-requisite. There are, however, coupled processes (associated with surface fluxes and upwelling etc.) that are diagnosable at short timescales, and more focus on these aspects would be useful. Understanding bias differences in coupled and uncoupled mode could be a useful approach. Researchers could also use the opportunity of upcoming field experiments to reduce systematic error and improve model uncertainty representation (e.g. in polar regions where mesoscale uncertainty is large at the sea-ice edge).

For regional models, research could focus on model error aspects that evolve independently of the large-scale boundary conditions. The poorer ability to use observational information (with relatively less in-situ data, and difficulties in using remotely-sensed data such as from radar) and the increased degrees of freedom in regional models might make this task more challenging than for global models.

4. **Can we separate errors that are truly random from errors that have complex but systematic dependencies on flow/regimes?**

Model error varies between different flow conditions, depending for example on how well the large-scale flow constrains the small scales. It is valuable to diagnose model errors in these different conditions separately, to give more information about how to improve the model and to produce a more informative estimate of the model error. This has been done by compositing data from locations and periods when a particular regime is in place, for example troughs over the US (Rodwell 2015), and then performing diagnosis of the model error. There are many possible large-scale regimes in different locations where model uncertainty associated with small-scale physics could be examined e.g. MJO phases, European blocks. Tropical cyclones could also be studied. Possible methods are to examine the EDA reliability budget (see talk by Mark Rodwell) or the statistics of analysis increments (see talk by Chiara Piccolo) in each regime. The latter can be compared with the predictions of stochastic physics schemes, or could be applied directly in an ensemble forecast. Selecting points based on the activity of physics schemes may also give useful information – for example, whether the assumption of SPPT that the standard deviation of model error is proportional to the total physics tendency is justified.

One way to diagnose flow-dependent model perturbations which relates to an "error of the day" is to apply an adjoint technique extended to diagnose optimal model tendency perturbations rather than initial state perturbations (*e.g.* Barkmeijer et al., 2003; Iversen et al., 2008). There is code in the IFS for this, which needs to be updated for use. The method can be extended to the non-linear range and to time-variable structures. The method can be used both to diagnose model perturbations for given actual model prediction errors in a pre-defined domain, and to improve the actual forecast.

5. **What are the appropriate metrics for model error: RMSE and bias, ensemble error and spread, reliability, probabilistic scores…?**

The aim of a model error representation is that, when it is included in the model and the model is run in an ensemble, we get a good quality ensemble. In addition also for the use in data assimilation, it should be able to reproduce the correct variance and correlation structure of the model error.

The ensemble quality can be assessed by standard measures. These can be supplemented by noted that verification against a randomly chosen member of the ensemble of analyses is equivalent to verification against the truth (Bowler et al 2015). In order to assess the model error representation within it we need to ensure that the resolution of the ensemble is not compromised. This requires measures of the RMSE and bias of the ensemble mean. We also need to ensure that the ensemble is reliable. It is important to include more complete measures of reliability, such as minimum spanning trees or other multivariate techniques in the verification. If the model is to be used for extended predictions we need to assess the impact on conservation properties of the model. For global models it can be sufficient to use the standard variables for verification. On the other hand, for limited area models it is very important to measure the error of parameters that affect the users most, e.g. reliability of precipitation, 2m temperature and cloud forecasts.

In order to assess the correlation structure of the model error we can again use the minimum spanning tree or other multivariate techniques. Furthermore, to confirm if sufficient state dependency has been achieved one needs to look at case to case variations of variances and non-isotropic correlation structures.

6. References

Barkmeijer, J., T. Iversen and T. N. Palmer, 2003: Forcing singular vectors and other sensitive model structures. *Quart. J. Roy. Meteorol. Soc.*, **129**, 2401–2423, doi: 10.1256/qj.02.126.

Bowler, N. E., M. J. P. Cullen and C. Piccolo, 2015: Verification against perturbed analyses and observations, *Nonlin. Processes Geophys.*, **22**, 403-411, doi: 10.5194/npg-22-403-2015.

Iversen, T., J. Kristiansen, T., Jung and J. Barkmeijer, 2008: Optimal atmospheric forcing perturbations for the cold-ocean warm-land pattern. *Tellus A*, **60**, 528–546, doi: 10.1111/j.1600-0870.2008.00310.x.

Klocke, D. and M. J. Rodwell, 2014: A comparison of two numerical weather prediction methods for diagnosing fast-physics errors in climate models. *Quart. J. Roy. Meteorol. Soc.*, **140**: 517–524, doi: 10.1002/qj.2172.

Laine M., A. Solonen, H. Haario and H. Järvinen, 2012: Ensemble prediction and parameter estimation system: the method. Quart. J. Roy. Meteorol. Soc., **138**, 289-297.

Ollinaho, P., P. Bechtold, M. Leutbecher, M. Laine, A. Solonen, H. Haario and H. Järvinen, 2013: Parameter variations in prediction skill optimization at ECMWF, Nonlin. Processes Geophys., 20, 6, 1001-1010.

Ollinaho, P., H. Järvinen, P. Bauer, M. Laine, P. Bechtold, J. Susiluoto and H. Haario, 2014: Optimization of NWP model closure parameters using total energy norm of forecast error as a target, *Geosci. Model Dev.*, **7**, 1889-1900, doi: 10.5194/gmd-7-1889-2014.

Rodwell, M. J., 2015: Using ensemble data assimilation to diagnose flow-dependent forecast reliability. *ECMWF Newsletter*, **146**, 29-34.