Working group 3: What are and how do we measure the pros and cons of existing approaches?

Reynolds, C., Leutbecher, M., Batte, L., Chen, S., Christensen, H., Klasa, C., Pegion, P., Plant, B. ORCID: https://orcid.org/0000-0001-8808-0022, Raynaud, L., Roberts, N., Sandu, I., Singleton, A., Sommer, M., Swinbank, R., Tennant, W. and Theis, S. (2016) Working group 3: What are and how do we measure the pros and cons of existing approaches? In: ECMWF/WWRP Workshop: Model Uncertainty, 11-15 April 2016, ECMWF, Reading, pp. 13-16. Available at https://centaur.reading.ac.uk/66610/

ECMWF/WWRP Workshop on Model Uncertainty, 11-15 April 2016, ECMWF, Reading, UK.

Summary and recommendations from Working Group 3: What are the pros/cons of existing model uncertainty schemes and how should these be measured?

Carolyn Reynolds (chair), Martin Leutbecher (co-chair), Lauriane Batté, Shuyi Chen, Hannah Christensen, Christina Klasa, Philip Pegion, Bob Plant, Laure Raynaud, Nigel Roberts, Irina Sandu, Andrew Singleton, Matthias Sommer, Richard Swinbank, Warren Tennant, Susanne Theis.

WG3 discussed both the pros and cons of existing schemes as well as metrics to measure relative advantages and disadvantages. We first provide a list of the current operational techniques and their respective advantages and disadvantages that were discussed in the WG. We do not claim that the list is complete, and we note that the pros and cons are neither exhaustive nor quantitative. Nevertheless, it may be useful to note the WG's consensus on the general advantages and disadvantages of the most commonly-used schemes. We then list our recommendations for evaluating model uncertainty schemes. At the end is a short list pertaining to recommendations for further development of methods to represent model uncertainty.

Pros and Cons of Existing Schemes:

1. Stochastically Perturbed Parameterization Tendencies: SPPT is effective in generating ensemble spread, inexpensive, and respects the balance between parameterizations. On the other hand, it is not directly tied to physical processes and violates conservation laws, cannot represent uncertainty when tendencies vanish, and cannot change the vertical distribution of heating, although recent developments such as independent SPPT (iSPPT) can address some of these issues.
2. Backscatter Schemes such as Stochastic Kinetic Energy Backscatter (SKEB) and Stochastic Convective Backscatter: An advantage is that these schemes are designed to represent missing physical processes. However, there is an apparent inconsistency between the scales of forcing that are effective at generating ensemble spread, and the scales of the phenomena for which the schemes are designed to compensate. There are also issues concerning the dissipation calculations. Another potential disadvantage is that the schemes become more expensive and less relevant as resolution is increased.
3. Additive perturbations (increment based methods): These perturbations are obtained using an objective measure of model error from the data assimilation system, and can be effective in generating ensemble spread. However, they are not flow-dependent, are not based on physical understanding, and are a function of the observing network and data assimilation systems.
4. Multi-model/ multi physics techniques: The advantages of these techniques are that each member is physically consistent, and the techniques are pragmatic and can allow for the leveraging of efforts at different institutions. However, the members are

nonexchangeable and will have different biases, necessitating larger reforecast sets for post-processing. Other concerns include nonphysical clustering, discrete sampling, and increased maintenance.

5. Stochastic parameterization methods: Convection schemes such as the Plant-Craig scheme, multi-cloud schemes, and some methods based on eddy diffusivity/ mass flux (EDMF) schemes are advantageous in that they are designed to address specific physical uncertainties. Some of these methods also have the capacity to be naturally adaptive to resolution, which should reduce the need for tuning. However, they are applied at the grid scale and so do not address important upscaling issues, there are potential coding complexities, and certain schemes have been tuned to perform well in certain regions (e.g., the multi-cloud scheme has been developed for the tropics). Cellular Automata (CA) schemes do have a non-local component, can result in convection in new areas, and may help with grey-zone issues. However, it appears somewhat difficult to control CA structures. It was noted that newly developed parameterizations (e.g., for radiation, gravity wave drag) were increasingly including intrinsic stochastic components, but the purpose of these components has often been for cost savings rather than sampling model uncertainty, and the stochastic forcing is uncorrelated in space, which limits impact.
6. Perturbed parameters: These methods have the advantage of being process-related (they should ideally reflect expert opinion on parameter uncertainty). A disadvantage is that they can be relatively costly to develop and maintain as parameterizations are frequently upgraded.
7. Post-processing: Post-processing and calibration can provide substantial benefit in terms of ensemble forecast performance measures and may be used as a benchmark for the development of model uncertainty schemes, provided that reanalyses and reforecasts are available. However, post-processing techniques often do not maintain physical consistency. The consistency may be relevant to generate outputs targeted to applications.

Primary Recommendations for Evaluation Methods:

WG3 discussed ways of measuring benefits and deficiencies of schemes to represent model uncertainty. The outcome of the discussion is a list of recommendations for measures to consider beyond the standard suite of metrics currently used in the verification of ensemble forecasts. Our primary recommendations are listed first, followed by a list of additional, secondary recommendations.

1. WG3 recommends evaluating the impact of stochastic forcing on the model behavior, for instance the impact on the bias or the impact on the frequency of extremes in the model climate. Testing weather models in the extended range and in climate simulations is an efficient way to identify problems with biases, variability, and extreme event frequencies. As summarizing scores can be insensitive to unrealistic extremes in the

predicted distribution, it was recommended to quantify the impact of schemes on model climatology for extreme events. WG3 noted that increases in the RMS errors of single forecasts may arise from stochastic forcing but they can be expected and do not imply that a method is not beneficial in an ensemble forecasting framework.

2. WG3 recommends examining the perturbations that schemes introduce to the model tendencies. This can be seen as a first step towards an objective comparison of model uncertainty representations. Documenting the ensemble variance and structure of the tendency perturbations associated with a model uncertainty representation is expected to help understanding differences between different schemes in the same model as well as differences between the same types of schemes in different models.
3. WG3 noted that variations between the perceived effectiveness of different schemes could be due to different configurations of the schemes (potentially due to tuning) and differences in the initial perturbations for the ensemble forecasts. For these reasons, one should not assume that small impact in one forecast system will imply small impact in other forecast systems.
4. WG3 suggests evaluating the impact of stochastic perturbations with process-based verification. Examples include those used in multi-model evaluations of the MJO[1] and the verification of tropical cyclones.
5. WG3 recommends evaluating the reliability of local (in space and/or time) variations in ensemble spread. It is important to not rely exclusively on the (global or regional) average agreement between ensemble spread and the error of the ensemble mean forecast.
6. WG3 recommends evaluating how model uncertainty schemes impact background error covariance estimates, and model error covariance estimates (for weak-constraint 4D-Var), as this will affect the structure of DA increments.
7. WG3 recommends consideration of spatial verification techniques to enhance the evaluation of meteorological entities with large spatial uncertainty compared to the scale of the entities themselves (e.g. precipitation rates or fog in convective scale ensembles, or frontal rain in medium-range weather forecasts). Upscaling, neighbourhood approaches, and approaches that consider displacement uncertainty are examples.

Additional Recommendations for Evaluation Methods:

8. WG3 noted that case studies and/or regime dependent studies together with subjective verification are also needed. However, one has to be aware of the forecaster's dilemma when interpreting a sample of cases that is conditioned on particular observed events (see http://arxiv.org/abs/1512.09244).

---

[1] For example, see Klingaman, N. P., et al. (2015), Vertical structure and physical processes of the Madden-Julian oscillation: Linking hindcast fidelity to simulated diabatic heating and moistening, J. Geophys. Res. Atmos., 120, 4690–4717, doi:10.1002/2014JD022374.

9. One should assess the impact of model uncertainty applied in one component of the system on the other system components. This is relevant within atmospheric modeling (e.g, the impact of stochastic forcing in one parameterization on other parameterizations) and within the broader context of coupled modeling (e.g., the impact of atmospheric model uncertainty on ocean performance).
10. WG3 noted the potential for ambiguities when specifying sources of model uncertainties with multiple schemes in one system. It was recommended to test methods independently and to use caution as deficiencies in one scheme may be compensated with perturbations from another scheme.

Recommendations for Improvement upon Existing Methods

1. WG3 recommends parameter space exploration research to obtain physically reasonably parameter ranges and correlations. Strong communication between parameterization developers and ensemble developers is encouraged to facilitate effective and realistic perturbed parameter schemes. WG3 also recommends further research into land surface and atmosphere-surface coupling to identify sensitive parameters, as this should lead to improved ensemble forecasts of high-impact near-surface variables.
2. WG3 agreed that more research in characterizing observation errors would be valuable, as this is essential to estimate background error and to verification at early lead times.
3. WG3 saw the need to consider uncertainty in the model dynamics beyond SKEB. The development of schemes may be informed through sensitivity experiments with different resolutions (i.e., coarse-graining studies). WG3 also recommends that one should not assume all model errors originate from sub-grid-scale variability.
4. WG3 noted that there was a need for proxies of model error (a topic under consideration in another working group) as many model uncertainty schemes require the specification of space and time scales for stochastic forcing. An example for obtaining model error proxies is a comparison with very high resolution simulations.