ECMWF/WWRP Workshop

Model Uncertainty

11-15 April 2016

ECMWF
Reading, UK

Content

Session 4 Latest research

Stochastic parameterization: Towards a new view of Weather and Climate Models	Judith Berner
Model uncertainty in global ocean models: Stochastic parametrizations of ocean mixing	Stephan Juricke
Representing model uncertainty for climate forecasts	Antje Weisheimer
On the diagnosis of model error statistics using weak-constraint data assimilation	Neill Bowler
A comparison of the model error and state formulations of weak-constraint 4D-Var	Amos Lawless
Using trajectories of ensemble analyses and tools from weak constraint 4D-Var to represent model uncertainty	Craig Bishop
Model error representation in convection-permitting forecasts and ensemble data assimilation	Glen Romine
Improving the Stochastically Perturbed Parametrisation Tendencies scheme using High-Resolution Model Simulations	Hannah Christensen
Impact of a stochastic deep convection scheme using cellular automata in the meso-scale ensemble prediction system; Harmon-EPS	Lisa Bengtsson
Parameter uncertainty of chaotic systems	Heikki Haario
Inducing tropical cyclones to undergo brownian motion	Daniel Hodyss
Observational based stochastic convection parameterization	Jesse Dorrestijn

Introduction

During 11-15 April 2016, ECMWF hosted the joint ECMWF/WWRP Workshop on Model Uncertainty.

This workshop saw more than 80 international experts meet to discuss the latest developments in diagnosing and characterising model error, and building schemes for simulating model uncertainty in assimilation and prediction systems. The workshop focussed on how to improve the physical basis of the stochastic forcing techniques, which are used to represent the effect of uncertainties in resolved and under-resolved processes in global atmospheric models, convection-permitting models and the longer timescales for land-surface, ocean and sea-ice coupling.

Through a combination of oral presentations, poster presentations and Working Group discussions, the workshop sought to answer the questions:

- What are the fundamental sources of model error?
- How can we improve the diagnosis of model error?
- What are and how do we measure the pros and cons of existing approaches to representing model uncertainty?
- How do we improve the physical basis for model uncertainty schemes?

Presented in these Workshop Proceedings are summaries of the oral presentations; as well as reports from the Working Groups, which outline their discussions and present recommendations for future research directions.

Report from Working Group 1

What are the sources of model error and how can we improve the physical basis of model uncertainty representation?

Group Members

George Craig (Chair), Richard Forbes (co-chair), Saleh Abdalla, Gianpaolo Balsamo, Peter Bechtold, Judith Berner, Roberto Buizza, Alfons Callado Pallares, Pieter De Meutter, Peter Dueben, Inger-Lise Frogner, Normand Gagnon, Dan Hodyss, Darryl Holm, Sarah-Jane Lock, Ekaterina Machulskaya, Husain Najafi, Pirkka Ollinaho, Tobias Selz, Leo Separovic, Aneesh Subramanian, John Tseng, Antje Weisheimer

The aim of Working Group 1 (WG1) was to identify the key issues and recommend priorities for future research directions for ECMWF and the wider research community to understand the sources of uncertainty and improve the representation of uncertainty in models. This summary provides a brief record of some of the main points discussed by the Working Group and the recommendations that came out of the discussion, structured around four questions:

(1) What are the sources of model uncertainty?

(2) What are the characteristics of error growth/scale interactions?

(3) How can we improve the physical basis of model uncertainty representation?

(4) How can we enhance collaboration across the research community?

1) What are the sources of model uncertainty?

- There are various sources of uncertainty in models that can result in model error, arising from spatial and temporal truncation errors, and limitations of our knowledge of physical processes across the "Earth system" (atmosphere, ocean, land surface, sea-ice, atmospheric composition).
- Model error is dominated by the representation of physical processes (e.g. boundary layer turbulence, surface coupling, cloud microphysics, cloud-radiation interaction, aerosols, convection gravity waves, surface drag), but we shouldn't neglect the uncertainty in the dynamics. Convective processes are a prominent source of model uncertainty due to strong non-linearities and upscale growth.
- Parametrization errors can arise from structural uncertainty (incorrect, or partial representation of the equations needed to describe the evolution), parameter uncertainty, and truncation errors. Errors can be thought of as either "systematic" or "random/intrinsic". Systematic errors are often related to particular meteorological regimes (i.e. due to regime-dependent errors in the physics which directly affect the meso/synoptic scales), whereas intrinsic errors can be considered to be due to upscale growth of uncertainty at the small scales, missing degrees of freedom and truncation errors. The former are visible in a deterministic forecast and have scope to be reduced by model improvements, the latter will always require stochastic perturbations in an ensemble. However, it can be very difficult to define and separate these different sources of error in models. Systematic errors may also be due to the non-linear response of the system to random perturbations. Regime-dependent systematic parametrization errors may appear as random error over some space or time scale due to the varying meteorological regimes.

- In practice it is difficult to disentangle different sources of model error (systematic error versus random error, truncation error versus physical process uncertainty, structural error versus parameter uncertainty) and model uncertainty schemes need to represent all these sources of error. Efforts should continue to try to define the different sources so that they can be represented more effectively. This will require different techniques such as coarse-graining studies, and sensitivity experiments to determine the most influencing parameters and terms in the equations. Multi-model/physics ensembles can provide improved spread in some situations – can we learn about structural errors from these?
- Data assimilation provides valuable information on model error in the short range and this should be exploited much more systematically.

Recommendations

1. *WG1 recognises the potential benefit of diagnosing model error from data assimilation, and recommends further work to understand the relationships between the representation of model error in the data assimilation system and the underlying dynamical and physical processes.*
2. *WG1 recommends that sensitivity/coarse-graining studies using convective-scale observations and models should continue to be pursued as they have further potential to inform model uncertainty representation and identify the most important processes.*
3. *WG1 recommends that sensitivity experiments of existing model uncertainty schemes (e.g. SPPT) should continue to be pursued as they have further potential for learning about the representation of model uncertainty (not just a tuning exercise).*
4. *WG1 recognises that multi-model/multi-physics-based ensembles can still add value for model uncertainty representation, particularly in the short-range, and recommends comparing different models to understand/inform how to better represent structural errors in model parametrizations.*

2) What are the characteristics of error growth/scale interactions?

- Representing uncertainty is not just a truncation/parametrization problem – we need to consider how errors propagate through the system.
- Some large scale errors are the result of small-scale errors that have propagated upscale, and some model errors are intrinsically large-scale in nature (e.g. due to regime-dependent systematic errors in parametrizations).
- To what extent do large-scale errors need to be represented at the small scale and propagated through the same processes, or can their large-scale effect be directly simulated as large-scale perturbations (for example, as suggested by the large spatial and temporal decorrelation times in SPPT)?
- Identification whether the errors are from small or large scales can help in targeting the latter where there is larger potential of improvement, compared to the former which may have already hit their intrinsic limit. Possible double counting should be avoided.
- The idea that the -5/3 energy spectrum, as emerged from observations (e.g. Nastrom and Gage), has an important role in getting correct error growth and in responding to stochastic perturbations was discussed. It is of course good to have the correct spectrum but the mechanisms responsible for the -5/3 mesoscale energy spectrum are not fully understood, so it is not clear whether a failure to represent this spectrum is associated with incorrect error growth.

- Is the -5/3 slope universal? Probably not in all regions, e.g. the tropics. The -5/3 spectrum may be a universal property of the system or it may be due to multi-scale interactions. What processes set this slope in the atmosphere – 3D turbulence, gravity waves, convection, orography? What processes set the slope in models – numerical schemes, physical parametrizations and their interaction with the dynamical core? Just because a model has the -5/3 spectra does not necessarily mean it is there for the right reasons.
- Some models represent the -5/3 spectra and some do not. The importance of capturing the correct spectrum could be examined by running two models with and without this spectrum and investigating the error growth from the same stochastic perturbations applied at varying scales.
- There are also scale interactions between land/ocean/atmosphere on different space and time scales which are not well understood and further work is required here.

Recommendations:

5. *WG1 recommends that model experiments are designed and performed to determine how error growth characteristics are captured, using models that do and do not represent the -5/3 spectra, and across different model resolutions (down to convective resolving scales).*
6. *WG1 recommends further analysis of observed atmospheric spectra to determine how universal the -5/3 is or how spectra vary with location, latitude, height, meteorological regime etc...*
7. *WG1 recommends exploring the importance of interactions within and between the uncertainty in various components of the Earth System with different error growth time scales (e.g. importance of resolving mesoscale eddies in ocean models versus stochastic representation of mesoscale eddy processes).*

3) How can we improve the physical basis of model uncertainty representation?

- WG1 discussed what "physically based" actually meant? One interpretation is a model uncertainty representation that is free from tunable parameters, instead based on universal properties that can be defined in some way from observations (e.g. are the dominant synoptic scale spatial patterns of perturbations used in SPPT intrinsic to all models, and if so, why?). An alternative interpretation is a representation of model uncertainty that is close to the relevant phenomena or processes (e.g. stochastically perturbed parameter (SPP) approach or stochastic convection schemes). "Physical consistency" is a different term that could be used. For example, tapering of the SPPT perturbations to zero in the boundary layer in the IFS is done for practical reasons and is not physically consistent with the perturbations in the rest of the column.
- Previous workshops have recommended building representations of uncertainty into the model physics parametrizations (e.g. stochastic convection schemes). We still think this is a priority, but benefits will only be realised if other parts of the model it interacts with are good enough. Model uncertainty is not just a parametrization problem; it also depends on, for example, upscale growth, scale interactions, dynamics and numerics.
- An improved physical consistency will need to address the different sources of model error as directly as possible and will likely consist of a combination of a number of approaches (e.g. representing subgrid-scale uncertainty, physics parameter uncertainty, uncertainty in all the components of the Earth system). We expect there will always be some uncertainty that we don't know how to represent explicitly.

- Stochastic advection (e.g. by a velocity containing a Brownian component with spatial correlations) is an example of how the dynamics and physics can be considered self-consistently. It potentially addresses two aspects of model error below the truncation scale: advective transfer by stochastic flow, and uncertainty and approximations in the physical parametrizations on the sub-gridscale flow.
- In many models there is missing smaller scale variability in the ensemble of near-surface parameters, which are important for forecast users (e.g. 2m temperature). Perturbations to soil moisture could be explored, or parameters in the land surface model, such as coupling coefficients or soil characteristics. Surface model perturbations could address the fast-coupling processes first, which should be climate neutral, but other perturbations may also be required to represent longer timescale uncertainties.

Recommendations

8. *WG1 recommends to continue working on improving the physical basis and physical consistency of model uncertainty representation, but it needs to be considered in the context of the whole ensemble prediction system and on improving our understanding of all the sources of model uncertainty, such as physics, dynamics, numerics and multi-scale interactions.*
9. *WG1 recommends investigation of stochastic advection processes to represent the advective transfer by stochastic flow below the truncation scale in models.*
10. *WG1 recommends a more concerted effort to improve the ensemble spread of near-surface fields, which are important for forecast users.*

4) How can we enhance collaboration across the research community?

- WG1 discussed how research in the area of model uncertainty could be enhanced by increased collaboration between the NWP community and the academic community.
- Specifically for ECMWF, links with the academic community are good, for example through the OpenIFS initiative, the ERA reanalysis projects or TIGGE datasets. These play a very important role in stimulating research. Links could be strengthened to enhance collaboration for mutual benefit, realising that this takes investment of time on both sides.
- Personal contacts are very important to facilitate collaboration, either through meetings and workshops, scientific visits (in both directions), joint research projects or PhD students.
- Improved representation of model uncertainty needs to be explored in an ensemble context, but it is difficult to run the ensemble system outside of an NWP centre and this needs to be made easier to encourage research.
- Model uncertainty is more than a parameterization problem – it includes dynamical meteorology, physical processes, numerics, and mathematics including stochastic methods. It is therefore a topic that would benefit from a range of ideas from different disciplines. The research community should be exploring alternative well-founded approaches to representing model uncertainty.

Recommendations

11. *WG1 recommends that ECMWF continues with and enhances collaboration with external researchers.*
12. *WG1 recommends that ECMWF consider how to facilitate access to the ensemble prediction system (ENS) for external researchers, so that modifications can be made without intensive ECMWF staff support and so that evaluation can be done more easily outside ECMWF or within a Special Project.*
13. *WG1 recommends that WWRP/WCRP and other organisations include model uncertainty as a topic in future meetings, to gain expert input, to focus interest and foster collaboration.*

Report from 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 modelerror?**

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.

Report from Working Group 3

What are the pros/cons of existing model uncertainty schemes and how should these be measured?

Group Members

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

non-exchangeable 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.

Session 1
Fundamental ideas

Weather prediction in a world of uncertainties: should ensembles simulate the effect of model approximations?

Roberto Buizza

ECMWF

This main topic of this workshop is the simulation of model uncertainties in ensembles designed to provide an estimate of the probability distribution function of analyses and forecast states. This is the context within which I will discuss the question posed above in this short communication.

Ensembles have proven, so far, to be the most effective way to provide a range of possible forecasts, thus complementing information about the most likely state with a confidence level. Ensembles, if accurate and reliable, provide more consistent (in time) and valuable information than single forecasts. To achieve greatest accuracy and reliability, the operational ensembles have been designed to simulate all the 'most relevant sources' of forecast error, which can be classified broadly as linked to initial condition (ICs) and to model uncertainties. The ICs' ones are due mainly to observations not being geographically uniform and being affected by measurement and representativeness errors, and to approximations and simplifications used in data assimilation. The model ones are linked to the fact that the equation of motion of the atmospheric flow are solved on a finite, discrete grid and include only an approximate description of the real physical processes.

In the early days on ensemble prediction (1980s and early 1990s), attention focused mainly on the simulation of ICs' uncertainties. In 1995, the Canadian global, medium-range ensemble was the first to include model uncertainties (*Houtekamer et al*, 1996, MWR 124). At ECMWF, the first stochastic scheme designed to simulate model uncertainties was implemented in 1999 (*Buizza et al*, 1999, QJRMS 125). Results from these two centres indicated that simulating model uncertainties was beneficial and improved accuracy and reliability. Following their examples, most of the operational ensembles have included model uncertainty schemes. Today, at ECMWF, two stochastic schemes are used to simulate model uncertainties: the Stochastically Perturbed Parameterized Tendencies (SPPT), an improved version of the original scheme with perturbations with up to 3 different scales, and the Stochastic Kinetic Energy Backscatter (SKEB) schemes (*Palmer et al*, 2007, ECMWF TM 540).

Following the Canadian example, since about 10 years ago ECMWF has been using ensembles also to estimate analyses' uncertainties, both for the atmosphere (say the wave-land-atmosphere) and the ocean. Considering the atmosphere, since 2008 an Ensemble of Data Assimilations (EDA; *Buizza et al,* 2008, QJRMS 134) has been used to give a measure of analysis' uncertainties, to provide flow- dependent background-error statistics to the ECMWF data assimilation systems, and to initialize the medium-range/monthly ensemble (ENS). Considering the ocean, an ensemble was used to produce the ocean analysis version 3, and is currently used to produce the operational Ocean Re-Analysis version 4 (ORAS4; *Balmaseda et al*, 2013: QJRMS, 139), which includes 5 members, generated perturbing the surface wind stresses. Also the ORAS4 ensemble members are used to initialize the ENS forecasts, since each of them is based on a coupled ocean-atmosphere model.

Ideally, the analysis and forecast ensembles should be consistent and have the same characteristics, to avoid initialization shocks and to initialize better all scales: the same number of members, the same, coupled model, with each forecast starting from one analysis, and with both ensembles using the same method to simulate model uncertainties. Full consistency would also allow diagnostics based on the analysis' ensemble to give us indications on how to improve the forecast ensemble, and vice-versa. We have not yet achieved full consistency, but we have been working hard to make two of these ensembles as consistent as possible (see Table A). Considering the atmosphere component, since March 2016 EDA and ENS use the same model version and horizontal resolution, albeit a different number of vertical levels. In terms of model uncertainties, the EDA uses a 1-time- scale version of the SPPT scheme, while ENS uses a 3-time-scale version of SPPT and SKEB. Furthermore, the EDA runs with a 12-hour delay and provides ENS only with a set of 25 perturbations (instead of 51 full model state), which are combined with the unperturbed high- resolution analysis and the singular vectors to generate the 51 ENS initial conditions. For the ocean component, both ORAS4 and ENS use the same version of the NEMO model with the same resolution. Ocean model uncertainties are not simulated in either ensemble. Finally, there are only 5 ocean analyses that are used to initialize the 51, coupled ENS forecasts.

Operational suites		Sources of uncertainty		
Type	**Hor. Resol. – Vert. levels – Fc length (days)**	**Obs**	**ICs**	**Model**
HRES	$T_{CO}1279$ (~9 km) - L137 – (0-10d)	--	--	--
H4DV	$T_{CO}1279$ (inner loops $T_{CO}255/319/399$) - L137	--	--	--
EDA	25 members: $T_{CO}639$ (~18km) - L137	δo	--	SPPT(1L)
ENS	51 members: $T_{CO}639$ (~18km) - L91 - (0-15d) $T_{CO}319$ (~36km) - L91 - (15-46d) - Ocean: NEMO ORCA100z42	-- --	$EDA^{25}+SVs^{50*Na}$ $ORAS4^{5}$	SPPT(3L) + SKEB --
S4	51 members: T_L255 (~80km) L91 - Ocean: NEMO ORCA100z42	-- --	SVs $ORAS4^{5}$	SPPT(3L) + SKEB --
ORAS4	5 members: NEMO ORCA 1 degree and 42 layers – Run with perturbed forcing fields			

Table A. Key characteristics of the ECMWF operational suites: the high-resolution forecast (HRES) and analysis (H4DV), the Ensemble of Data Assimilations (EDA), the medium-range/monthly (ENS) and the seasonal system- 4 (S4) ensembles, and the ocean analysis ensemble (ORAS4). For the wave-land-atmosphere component (the Integrated Forecasting System, IFS), TcoNNN indicates a spectral-triangular truncation NNN with a cubic- octahedral grid; Lxx is the number of vertical levels (all suites have the top of the atmosphere at 0.01 hPa).
Three sources of forecast error are simulated, linked to observations' errors (simulated in the EDA by perturbing the observations), initial-conditions (simulated both in ENS and S4 with two different methods) and model uncertainties. ORAS4, the ocean data assimilation, includes 5 members, which are used to initialize ENS and S4.

As part of this workshop, we will be discussing how to progress in the simulation of model uncertainties. It is worth recollecting few key recommendations that were made at three workshops held at ECMWF in 2005 on 'The representation of sub-grid scales', in 2007 on 'Ensemble Prediction', and in 2011 on 'Model uncertainty'. On diagnostic and evaluation, it was recommended to develop a methodology to diagnose the spectral energy transfer, to use coarse-graining strategies (with a factor of 10 difference in resolution) to determine the statistics that an effective stochastic scheme should generate, and to use initial tendencies and analysis increments to determine model error statistics. On the physical basis of model uncertainty simulation schemes, it was suggested to explore the physical basis of the stochastic schemes, to develop physical parameterisations that include explicitly model uncertainty estimations, and to apply 'falsification concepts' (does the model error scheme invalidate physical constraints?) in the scientific work.

I think that the recommendations listed above are still valid, and I would like to conclude by suggesting that we add 'consistency' between the analysis and forecast ensembles, as another goal to achieve. At ECMWF, last we have recently coupled the NEMO ocean model in ENS from day 0 because we have shown that this coupling improves the accuracy and reliability of our global, medium-range/monthly ensemble forecasts. Going back to the question that I posed at the start of this communication, my answer is affirmative: we should simulate all relevant sources of model uncertainties. Furthermore, I suggest that we aim to achieve full consistency and develop an Integrated Coupled Analysis and Forecast Ensemble (I-CAFÉ) that includes the same model uncertainty scheme(s) in both the coupled (ocean-wave-land-atmosphere to start with) analysis and the forecast elements, with forecasts' initial conditions given by the coupled ensemble of analyses.

Physically-based stochastic parameterisation

George C. Craig

Meteorological Institute, LMU Munich

Parameterisation of unresolved variability in the atmosphere leads to uncertainty in the resolved state of the atmosphere that can often be represented stochastically. Since most parameterisations are based on physical understanding of a small-scale process, that understanding can be used to describe the stochastic variability. A list of criteria that a physically-based stochastic parameterisation should satisfy will be presented, and two examples of stochastic parameterisation schemes will be discussed: a deep convection scheme (Plant and Craig 2008) and a representation of boundary layer variability (Kober and Craig, in review) for convection permitting models. Some comments will be made regarding upscale error growth and role of stochastic perturbations in the predictability of weather.

Stochastic parametrisation models for GFD

Darryl D Holm

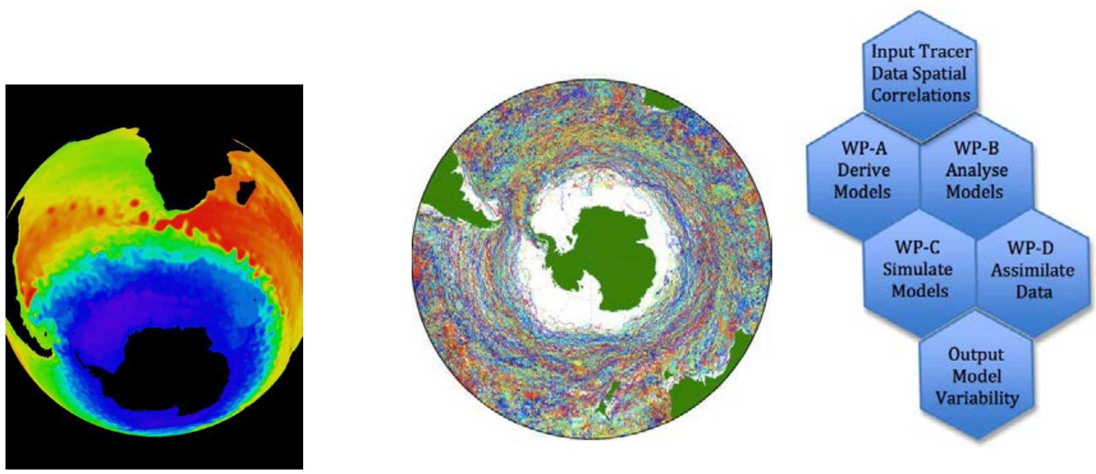
Imperial College London, Mathematics Department

In next-generation weather and climate models, stochastic parameterisation should be an important element in providing reliable estimates of model uncertainty. A fundamental conclusion of Berner, Jung & Palmer [2012] is that

> *"a posteriori* addition of stochasticity to an already tuned model is simply not viable [satisfactory]. This in turn suggests that stochasticity must be incorporated at a very basic level within the design of physical process parameterisations and improvements to the dynamical core."

This talk responded to the workshop's challenge of "How do we improve the physical basis for model uncertainty schemes?" It proposed an *a priori* introduction of stochasticity for GFD models at various levels of approximation, by introducing the methodology of Holm [2015] as a potential framework for quantifying model *transport error*. In turn, the stochastic representation of model transport error would introduce stochasticity into the parameterisations of subgrid scale processes.

This methodology introduces Stochasticity into Partial Differential Equations (SPDEs) for the model, via Variational Principles (SVPs), with corresponding implications for Numerical Modelling, Stochastic Data Analysis, and Geophysical Fluid Dynamics (SGFD). The motivation for introducing stochasticity was illustrated by comparing the relative resolutions of numerical simulations and satellite data for tracers on the surface of the ocean; and the methodology was sketched as a series of interconnected hexagons in the following Figure. The left and centre panels of the Figure illustrate the difference in scales between the numerical resolution and the satellite observations for this problem, e.g., for estimating the spread of floating tracers such as plastic containers (or, "rubber duckies") in the Southern Ocean. The rightmost panel in the Figure shows the closely integrated tasks in formulating the methodology for stochastic estimation of model transport error.



In this methodology, the transport stochasticity is introduced via the correlation eigenfunctions for the advection data being analysed, by multiplying each eigenvector of the correlation matrix for the tracer data (called an empirical orthogonal function, or EOF) with a stochastic amplitude in the Stratonovich sense, and then taking the sum over these stochastic products as the deviation from the drift velocity. The drift velocity itself is then obtained via the known Hamilton's variational principal for deterministic fluid dynamics, but with variations of the velocity and advected quantities which are constrained to satisfy the stochastic EOF approximation of the satellite tracer data. This methodology of stochastically constrained variational principles is complementary to the customary practice in weather forecasting in which data is assimilated using variational principles. However, this proposed methodology is to be used in formulating the model for the dynamical core, rather than assimilating the observed data. The task of the methodology is to learn from stochastic assimilation of data (tracers) the spatial correlation features of the observed advected quantities. These quantities are needed as input into a constrained variational principle to derive the stochastic fluid motion equations, whose transport will predict statistics such as the variability of the advected data which, by constrauction, will be consistent with the observations.

The talk outlined this methodology, then illustrated it by deriving several new stochastic GFD models for predicting the evolution of climate and weather variability, based on observations of tracer data. The new feature of these potential dynamical core motion equations is that they contain stochastic perturbations which multiply both the solution velocity *and* its spatial gradient. Remarkably, these stochastic GFD models still preserve fundamental fluid properties such as Kelvin's circulation theorem and PV conservation. Indeed, as illustrated by the Kelvin circulation theorem for three-dimensional incompressible stochastic fluid motion, these fundamental mathematical structures in fluid dynamics retain their deterministic forms. However, their transport velocities are augmented by advection along the stochastic Lagrangian particle paths obtained from the spatial correlations of the tracer data. As a mathematical bonus, the equivalent Ito forms of these Stratonovich equations contain symmetric, second-order, derivative operators which tend to regularize the solutions of the new stochastic GFD equations, without any additional viscosity. This is apparently because the stochasticity in these new motion equations multiplies the gradients of the solutions, so its effects are enhanced in the vicinity of strong gradients during the evolution of the flow.

The areas of relevance of the new approach in matters of potential interest for ECMWF are:
(A) New stochastic parameterisation models for GFD derived using variational methods;
(B) Mathematical analysis of the stochastic transport equations in these models;
(C) Development of numerical methodology for stochastic GFD;
(D) Stochastic data assimilation using nonlinear particle filtering.

Key references:

J. Berner, T. Jung and T. N. Palmer [2012] Systematic Model Error: The Impact of Increased Horizontal Resolution versus Improved Stochastic and Deterministic Parameterizations, Journal of Climate, 25: 4946--4962.

D. D. Holm, [2015] Variational Principles for Stochastic Fluid Dynamics, Proc Roy Soc A, 471: 20140963.

Session 2
Diagnosing model error

Mesoscale convective systems as a source of model error	Glenn Shutts
A weather-system perspective on forecast errors	Heini Wernli
Resolved and parmetrized energy cascades in the IFS	Sylvie Malardel
Diagnosing and representing model error in 4DVar	Katherine Howes
Estimate of model error covariance Q for weak-constraint 4D-Var	Jacky Goddard
Evaluation of model error using data assimilation	Chiara Piccolo
Diagnosing systematic numerical weather prediction model bias over the Antarctic from short-term forecast tendencies	Steven Cavallo
Using forecast temporal variabiliy to evaluate model behaviour	Carolyn Reynolds
Using ensemble data assimilation to diagnose model uncertainty	Mark Rodwell

Mesoscale convective systems as a source of model error

Glenn Shutts

Met Office, UK

In the search for the dominant sources of systematic and random model error, high model resolution datasets and satellite imagery are highlighted as particularly useful tools. With the recent introduction of simulated infra-red (IR) radiance to global operational model forecast diagnostics, direct comparison with the corresponding satellite product gives a new way of assessing model error. Typically, the simulated IR over the North Atlantic ocean and Europe shows impressive agreement with actual imagery, apart from textual differences that reflect the higher resolution of the satellite data. However, in spring and early summer in particular, the extent, brightness temperature and location of upper cloud over the United States and the Caribbean often exhibit large errors due to the explosive growth of mesoscale convective systems (MCS) for which the assumptions of convection parametrization schemes are not strictly valid. Furthermore, these major convective events often occur in baroclinic environments close to jetstreams and their effect on the meso- to synoptic scale potential vorticity field can influence downstream Atlantic cyclogenesis.

Simulated IR images from 2.2 km forecasts with the Met Office Unified Model (UM), in which convection is explicitly represented, appear to show much more realistic representations of MCS cloud shields although the MCSs are often too intense and sometimes triggered spuriously.

This presentation shows examples of simulated IR versus actual brightness temperature and for one case that was well-simulated by the 2.2 km UM, model winds and potential vorticity are computed. Divergent winds within the cloud shield carry low potential vorticity away from the updraught cores and attempts are made to compute the mean absolute vorticity for different specifications of the cloud shield in terms of brightness temperature. In addition to these quasi-operational high resolution forecasts, idealized simulations of mesoscale convective events are made with the Met Office Large Eddy Model with a view to better understanding their impact on the mesoscale vorticity field.

The ultimate goal of these studies is to quantify the likely random model vorticity and divergence errors associated with MCSs and relate these to, and improve, the formulation of the Stochastic Convective Backscatter scheme (Shutts, 2015).

A weather-system perspective on forecast errors

Heini Wernli

Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland

This presentation investigates the linkage between forecast errors and specific weather systems and contributes to addressing the following questions: (i) How well can NWP models represent and predict certain weather systems (e.g., the track and intensity of a cyclone or the lifecycle of a Rossby wave?), and (ii) Are there specific weather systems involved in situations where fore- casts errors are large?

In a first part (slides 3-12), a brief overview is given on previous studies quantifying the quality of forecasts of specific weather systems, i.e., tropical cyclones, extratropical cyclones, and Rossby waves (troughs and ridges). The results indicate a general improvement of forecasts during the last decades, the occurrence of cases with still large forecast errors, and a systematic tendencies of medium-range NWP forecasts to underestimate Rossby wave amplitudes, which is likely due to an underrepresentation of diabatic modification and transport of air from the lower troposphere into upper tropospheric ridges (Gray et al. 2014).

The presentation then focuses specifically on so-called warm conveyor belts (WCBs) – coherent airstreams in extratropical cyclones, which produce intense precipitation, latent heating, and lead to a net transport of low potential vorticity (PV) air into upper-level ridges. WCBs can be identified objectively by calculating air parcel trajectories and selecting those that ascend by more than 600 hPa in 48 h in the vicinity of a cyclone (e.g., Joos and Wernli 2012). Compared to climatology, the outflow regions near the jet stream of these WCBs constitute strong negative PV anomalies (Madonna et al. 2014), which can significantly affect the downstream flow evolu- tion and, in certain cases, contribute to the formation of blocking (Pfahl et al. 2015). A specific PAL verification technique is then briefly introduced to quantify three aspects of the quality of WCB forecast: A, the amplitude of the WCB (number of strongly ascending trajectories); L, the location of the WCB outflow; and P, the amplitude of the associated upper-level negative PV anomaly (Madonna et al. 2015). It is shown that (i) all three components of WCB forecast errors increase with forecast lead time (slides 18,19), (ii) WCB forecasts of the high-resolution IFS im- proved over the last 15 years (slide 20), (iii) in today's forecast system no systematic over- or underprediction of WCB intensity occurs, and (iv) poor forecasts with a low anomaly correlation coefficients (ACC) are also associated with high values of PAL (red circles on slide 19). Finally, first results are shown from an ongoing master thesis project at ETH, in which WCBs are inves- tigated in ECMWF ensemble forecasts. The Brier skill score for the occurrence of WCBs over the North Atlantic indicates fairly high values for 2-day forecasts (BSS > 0.5) and clearly re- duced values for 5-day forecasts in particular in a band near 45ºN (BSS < 0.3, slide 23).

The final part of the presentation addresses the underlying meteorology of selected fore- cast bust events. Martinez-Alvarado et al. (2016) pointed to the importance of WCBs for correct- ly forecasting a high-amplitude Rossby wave evolution. Rodwell et al. (2013) also emphasized diabatic processes, in this case, MCSs over the eastern US, for European forecast busts (see also presentation at the workshop by Glenn Shutts on forecast errors induced by MCSs). Slides 27-50 then show preliminary results from an investigation of a forecast bust in October 2013. The ACC over Europe dropped below 0.2 at forecast day 5, leading to a too zonal flow in the forecast instead of a strong northerly flow in the analysis, which brought the first cold spell to Switzerland in 2013. Reasons for this poor forecast were (i) a slight mismatch in representation of a WCB over the eastern North Pacific, (ii) subsequently a missed reabsorption of an upper-level PV cut- off over North America (slide 41), and (iii) resulting from this a too weak Rossby wave amplifi- cation over the North Atlantic. Very interestingly, in this case the question whether a pre-existing PV cutoff is reabsorbed or not (a strongly non-linear process!) plays a crucial role for the down- stream flow evolution. Systematically, the 10 best (worst) ensemble forecasts did (not) reabsorb the cutoff (slides 48,49) and this went along with a systematic shift of the upstream WCB out- flow (slide 50).

In summary, this presentation tried to emphasize the following aspects:

1) It is meaningful to look at forecast errors from a weather system perspective. This requires the development of specific algorithms and metrics, which as a drawback involve some subjective decisions (e.g., what field should be taken to identify a cyclone?) and thresholds (e.g., what is the minimum lifetime of a cyclone?).
2) Research during the last years produced several promising results and emphasized the key role of the interaction of moist diabatic processes and the larger-scale flow evolution (e.g., role of latent heating in WCBs for evolution of cyclones and upper-level Rossby waves). It is likely that in certain cases deficiencies in the model physics negatively impact forecast quality, typically on the medium-range and downstream of the main diabatic activity.
3) Much needs to be done to more systematically investigate this pathway of research and to specifically identify critical aspects of model physics and its large-scale flow interaction.

Acknowledgements: Many thanks to my colleagues, in particular Maxi Boettcher, Christian Grams, Hanna Joos, and Erica Madonna for their input to this presentation.

References

Gray, S. L., C. M. Dunning, J. Methven, G. Masato, and J. M. Chagnon, 2014. Systematic model forecast error in Rossby wave structure. Geophys. Res. Lett., 41, 2979-2987.

Joos, H., and H. Wernli, 2012. Influence of microphysical processes on the potential vorticity development in a warm conveyor belt: a case study with the limited area model COSMO. Quart. J. Roy. Meteorol. Soc., 138, 407-418.

Madonna, E., H. Wernli, H. Joos, and O. Martius, 2014. Warm conveyor belts in the ERA-Interim data set (1979-2010). Part I: Climatology and potential vorticity evolution. J. Climate, 27, 3-26.

Madonna, E., M. Boettcher, C. M. Grams, H. Joos, O. Martius, and H. Wernli, 2015. Verification of North Atlantic warm conveyor belt outflows in ECMWF forecasts. Quart. J. Roy. Meteorol. Soc., 141, 1333-1344.

Martinez-Alvarado, O., E. Madonna, S. L. Gray, and H. Joos, 2016. A route to systematic error in forecasts of Rossby waves. Quart. J. Roy. Meteorol. Soc., 142, 196-210.

Pfahl, S., C. Schwierz, M. Croci-Maspoli, C. M. Grams, and H. Wernli, 2015. Importance of latent heat release in ascending air streams for atmospheric blocking. Nature Geosci., 8, 610-615.

Rodwell, M. J., and Coauthors, 2013. Characteristics of occasional poor medium-range weather forecasts for Europe. Bull. Amer. Meteor. Soc., 94, 1393-1405.

Resolved and parametrised energy cascades in the IFS

Sylvie Malardel and Nils Wedi

ECMWF

Spectral energy budgets following the method proposed by Augier and Lindborg (2013) have been used to illustrate how physical parametrisations influence the energy spectra and the non-linear energy transfer across scales in the IFS.

Simulations with increasing complexity show that the surface drag and the vertical subgrid mixing of momentum in the boundary layer have a strong control on the non-linear energy transfers of both kinetic energy and available potential energy and that they influence the shape of the energy spectra.

Spectral analyses of the tendencies issued from the parametrisations show that the physical parametrisations act at all scales. Simulations with explicit convection also suggest that the convection parametrisation disable natural energy transfers across scales and replace them by direct and adjustable forcing at all scales. By comparing the spectral diagnostics for model simulations with different complexity and by comparing different modelling choices, an attempt is made to assess model error behaviour.

Malardel, S., and N. P. Wedi (2016), How does subgrid-scale parametrisation influence nonlinear spectral energy fluxes in global NWP models?, J. Geophys. Res. Atmos., 121, doi:10.1002/2015JD023970

Diagnosing and representing model error in 4DVar

Katherine Howes, Dr Amos Lawless, Dr Alison Fowler

University of Reading

Four dimensional variational data assimilation (4DVar) can be used to obtain the best estimate of the initial conditions of a weather forecast model, namely the analysis. Our work is focused on improving the analysis by allowing for the fact that the model contains error, without requiring prior knowledge about the model error statistics.

The 4DVar method developed acknowledges the presence of random error in the model at each time step, by replacing the observation error covariance matrix with an error covariance matrix that includes both observation error statistics and model error statistics. A method for estimating this matrix is presented. In summary this combined observation error and model error covariance matrix is estimated with 'Dezrosiers-type diagnostics' that account for the presence of random error in a model.

We present analytical results for an erroneous scalar model which show a decrease in the variance of the error in the analysis when using our new method. We show that the improvement the method can make to the accuracy of the analysis is dependent on both the size of the model error and on the ratio between the observation error variance and background error variance. We then further demonstrate numerically that the new method also works to reduce the analysis error covariance when using a non-linear chaotic system with random error present. We discuss the fact that an improved analysis will not necessarily provide a better forecast.

Estimate of model error covariance Q for weak-constraint 4D-Var

Jacky Goddard and Mike Fisher - ECMWF

Currently the operational implementation of 4D-Var at ECMWF uses strong-constraint 4D-Var [4][5][7]. Strong-constraint 4D-Var relies on the assumption that the numerical model's representation of the evolution of atmospheric flow is perfect, or at least that the model errors are small enough to be neglected compared to other errors in the system [3]. Errors in observations and background state are accounted for using the **R** observation and the **B** background error covariance matrices. As other aspects of data assimilation processes have advanced, the validity of this perfect model assumption becomes more questionable and limits the length of the analysis window to roughly 12 hours. Weak-constraint 4D-Var relaxes the perfect model assumption by explicitly representing model error as part of the 4D-Var control variable. The model is now only a weak constraint on the system. However, a model error covariance matrix is required. Here, a new model error covariance matrix based on statistics from parametrised model error schemes is proposed for use in the short forecast.

Model Error Formulation

Model error contains both random and systematic (or even constant) components. To simplify the problem we consider the model error to be constant by intervals. For the work presented here, the interval we have chosen is one constant forcing for the whole 12 hour assimilation window, at the other extreme we could (in principle: it is not yet technically possible in the IFS) have chosen to have the interval as short as a model time step; this would be the full 4D problem.
The 4d-Var cost function we are considering is:

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

where η_b is the mean model error, x_k is the state at time k with $x_k = \mathcal{M}(x_{k-1}) + \eta$ representing the state at time k resulting from the forced model integration from time $t = 0$ to $t = k$, η represents the instantaneous model error. Observations and model error are assumed uncorrelated in time.[6]

Calculation of model error covariance matrix (Q)

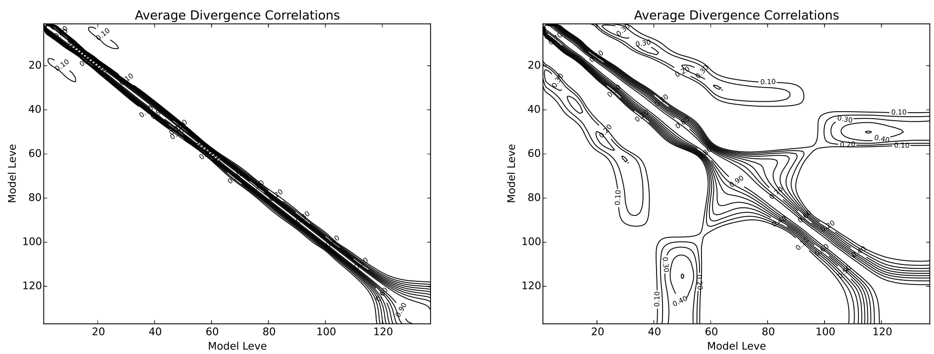
In order to calculate the new Q matrix, statistics are generated from special runs of the Ensemble Prediction System (EPS) in which initial perturbations are removed. In these runs, members diverge from each other due to their different realisations of parametrised model error (SKEB and SPPT). The differences between members after 12 hours of model integration give an estimate of the integrated effect of model error over 12 hours; from which statistics appropriate for use in 4D-Var can be calculated. These statistics are used to construct a covariance model similar to that described by Derber and Bouttier [1] for background error covariances. (Note, however, that the model error covariance matrix we have constructed does not include a balance operator: model errors for different variables are assumed to be uncorrelated.) This method of generating model error covariance statistics provides greater consistency between the approaches taken to representing model error in the 4D-Var and EPS systems than previous methods.

EPS experiment description:

- 50 member ensemble + control;
- $\mathrm{T_L}$ 399 resolution;
- 12 hour forecast;
- cycle 40R3;
- 20 days of forecasts (2013083100 - 2013091900);
- identical initial conditions (ensemble members are not perturbed).

Initial experimentation using the new $\mathbf{Q}$ matrix suggested that the implied variances of model error were too large. Weak-constraint analyses using the matrix were found to have very small initial increments as 4d-Var found it less costly to nudge the state towards the solution via a model-error correction than to correct the initial state. In order to select a reasonable magnitude of the $\mathbf{Q}$ matrix we looked at the minimisation of the cost function for a range of different values of multiplicative factors between 0 and 1 and choose a value for which the model error $\mathbf{Q}$ term has an influence but does not dominate over the other error terms.

For the multiplicative factor value chosen ($\alpha = 0.2$), a 4D-Var weak-constraint assimilation experiment was run. The 4D-Var model error estimates η from this experiment were then used to calculate a covariance matrix that could be compared with the $\mathbf{Q}$ matrix used in the assimilation. The experiment was run for 90 days with 12-hour assimilation windows starting at 0900 and 2100; both times were used for calculating the model error covariance estimates.



(a) 12 hour stochastic

(b) Weak constraint

Figure 1: Divergence model error average vertical correlations for 12 hour stochastic model error and weak-constraint 4D-Var model error estimate. Contour interval is 0.1 for both figures.

In fig. 1 we see the average vertical correlations of estimated model error for divergence. The 4D-Var weak-constraint model error covariance estimate (right panel) does not retain the same structure as the stochastic model error covariance, $\mathbf{Q}$ (left panel). In particular, we see some unexpected correlations between levels that are far apart.
By looking at the geographical location of these correlations we saw a clear pattern over North America and Europe corresponding to areas with a high number of aircraft observations. This suggests that 4D-Var is misinterpreting aircraft observation error or bias as model error. In order to avoid this interaction with aircraft observations, subsequent experiments restricted the effect of model error to be active only above 100hPa.
A CY41R2 $T_{co}1279$ experiment was run. Forecast skill scores were verified against own analysis and also against GPS Radio Occultation (GPSRO) observations. The verification against own analysis in the northern hemisphere (fig. 2) showed a significant reduction in RMS error at 100hPa. GPSRO verification in the stratosphere showed a change in the bias structure throughout the forecast. In the northern hemisphere bias is slightly improved at all levels, in the tropics it is largely unchanged and in the southern hemisphere the results are mixed (but the differences from the control in this region are very small).

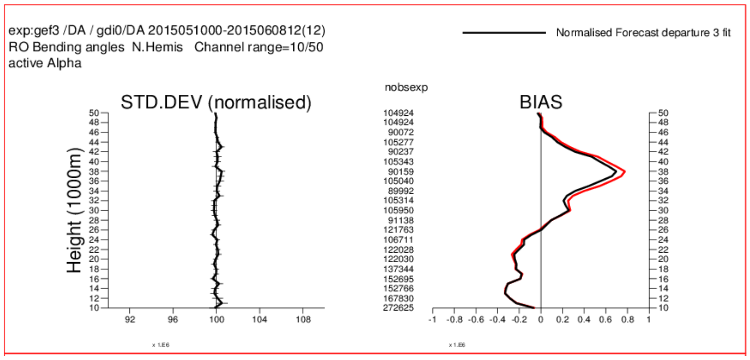


Figure 2: GPSRO verification northern hemisphere

We hope to introduce this configuration of weak-constraint 4D-Var with model error forcing above 100hPa into CY43R1. Understanding and reducing aliasing of observation error is critical to plans to extend the model error representation to all model levels, and will require improvements to the representation of systematic observation error (in particular biases in aircraft data). Finally, we plan to extend our research to encompass the random component of model error once the technical facility to represent it in 4d-Var exists. For this, we rely on ongoing developments within the OOPS project.

Bibliography

[1] Derber J., Bouttier F., *A reformulation of the background error covariance in the ECMWF global data assimilation system*, Tellus, 51A, 195-221, 1999.

[2] Elías Valur Hólm, Yannick Trémolet, *Investigation leading to turning off model error for L137 in CY38R2*, Research Department Memorandum, RD 13-422, October 29, 2013.

[3] Yannick Trémolet, *Model error estimation in 4D-Var*, Research Department Memorandum, RD 07-520, March, 2007.

[4] Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A. (2000), *The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics.* Q.J.R. Meteorol. Soc., 126: 11431170., doi: 10.1002/qj.49712656415.

[5] Mahfouf, J.-F. and Rabier, F. (2000), *The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics.* Q.J.R. Meteorol. Soc., 126: 11711190., doi: 10.1002/qj.49712656416.

[6] Mike Fisher, Yannick Trémolet, Harri Auvinen, David Tan and Paul Poli, *Weak-Constraint and Long-Window 4D-Var.* Research Department Memorandum, RD-655, December 2011.

[7] Klinker E, Rabier F, Kelly G, Mahfouf JF. (2000) *The ECMWF operational implementation of fourdimensional variational assimilation. III: Experimental results and diagnostics with operational configuration.* Q.J.R. Meteorol. Soc., Apr 1;126(564):1191-215.

Evaluation of model error using data assimilation

Chiara Piccolo and Mike Cullen
Met Office, Fitzroy Road, Exeter, EX1 3PB, UK.

May 6, 2016

The principles of initiating an ensemble forecst (EPS) with ensemble data assimilation (EnDA) are reviewed. This allows an estimate of initial uncertainty consistent with the uncertainties due to the model and the available present and past observations. Maximum resolution of the EPS is achieved by using the best available and affordable deterministic model. Achieving reliability then requires an estimate of the errors in the deterministic model. The true state in model space is filtered to the model resolution. This means that the true evolution is stochastic, as it depends on information that is not present in the initial state. The error in a deterministic model is therefore also stochastic.

If the statistics of the model error are known, then a reliable forecast ensemble can be generated given a reliable analysis ensemble. In particular, a reliable prior ensemble can be generated for the next analysis cycle. If the statistics of the observation errors are also known, and represented by perturbed observations, then an analysis ensemble performed by updating a randomly chosen prior ensemble member using a random draw from the perturbed observations will also be reliable. This is because the true state is statistically indistinguishable from a random member of the prior ensemble, and the true state mapped to observation space is statistically indistinguishable from a randomly chosen set of perturbed observations. Thus no update is performed at the true state, and so the reliability of the analysis ensemble is assured whatever method of analysis update is used, and whether or not the statistics are Gaussian.

Since the model error is inherently unknowable *a priori* because it depends on unknown information, the statistics of model error can only be estimated from observations. Data assimilation provides a way of doing this which allows all observations to be used while properly allowing for observation error. Ideally this should take the form of a reanalysis. The weak constraint 4dVar method is designed to estimate the forcing term with the minimum variance which, when included in the model, allows the model to fit the observations to within observation error over an extended period. We can infer the statistics of the necessary forcing term by performing cycled weak constraint 4dVar with no background increments. This can only give the statistics of the model error over a sufficiently long period for the data assimilation to be fully spun up. It requires a prior estimate of the model error statistics, which should ideally be bootstrapped. If the forcing terms estimated from the assimilation can be regarded as a random draw from an archive of such increments, then the reanalysis trajectory will be staistically indistinguishable from a model trajectory forced with randomly chosen increments from the archive.

This idea is tested using the Met Office Unified Model with 40km horizontal resolution and 70 levels. An archive of model error forcing terms is generated using weak constraint 4dVar with no background term. An ensemble data assimilation and forecast system is then run with 10 members, perturbed observations, and strong constraint 4dVar. Randomly chosen model error forcing terms from the archive are added to the model trajectories. 6 hour forecasts from the system are then verified against randomly chosen members of the analysis ensemble. This is equivalent to verifying against the truth if the analysis is properly set up. The spread-skill relation is satisfied to within sampling error.

Results are presented for 6 day forecasts, which are found to be reliable based on the spread-skill relation. They are also presented for 10 year AMIP simulations verified against ERA-Interim analyses. These show large improvements over the control, primarily because the systematic errors are removed by the forcing terms. Some of the remaining errors are because our simulations should reproduce a Met Office reanalysis, which will not be the same as ERA-Interim due to differences in the two assimilation systems.

Additional results are presented which show that our system, when used only in forecast mode, outperforms the Met Office operational EPS. This is because the model error forcing is significant in all regions, while the stochastic physics used in the operational EPS is mostly restricted to the storm tracks. We also illustrate that the use of weak constraint 4dVar to estimate the model error forcing is important. Analysis increments calculated on the assumption that increments are only added every 6 hours are different in character, typically smaller and on smaller scales.

C. Piccolo and M.J.P. Cullen (2016) Ensemble Data Assimilation Using a Unified Representation of Model Error. *Mon. Weather Rev.,***144***, 213-224* .

Diagnosing systematic numerical weather prediction model bias over the Antarctic from short-term forecast tendencies

Steven M. Cavallo

University of Oklahoma School of Meteorology, Norman, OK

The Antarctic Mesoscale Prediction System (AMPS) is a derivative of the Advanced Research Weather Research and Forecasting (ARW-WRF) limited area model (LAM), with modifications in physics parameterizations aimed to improve polar prediction (Powers et al. 2003). AMPS forecasts are a single deterministic realization initialized using a 3DVAR method employed from a Global Forecasting System (GFS) analysis. Data assimilation is not continuously cycled and AMPS forecasts are initialized from a non-native analysis (a different model is used in the data assimilation and the forecast). During September-December 2010, the Concordiasi intensive observing period (IOP) occurred over the Antarctic and parts of the Southern Ocean, where unique vertical atmospheric profiles were collected from dropsondes that were deployed from driftsondes within the Southern Hemisphere polar vortex (Rabier et al. 2010). This provided an opportunity to examine systematic model bias in the AMPS LAM, where horizontal grid spacing was 45-km at the time of the IOP. This study uses short-term forecast tendencies to formally diagnose systematic model error for the period 21 September - 30 September 2010. The hypothesis of this study is that the sources of model bias can be diagnosed to the precise physical parameterization and locations using short-term forecast tendencies in a method referred to here as the mean initial tendency analysis (MITA) increment method (e.g., Klinker and Sardeshmukh 1992; Rodwell and Palmer 2007).

To minimize initial condition error so that it is statistically distinguishable from model error, it is best if the same identically configured model used to create the analysis is used for forecasts (e.g., Rodwell and Palmer 2007; Klocke and Rodwell 2014). In these cases, the analysis is referred to as a 'native' analysis since the same model is used in both the data assimilation and forecasts (e.g., Klocke and Rodwell 2014). The departure of the short- term forecasts from the observed atmospheric state in the early time steps allows for the potential identification of process-level errors, and this growth of errors is often referred to as model 'spin-up' (Rodwell and Palmer 2007). Thus, the AMPS numerical weather model and an Ensemble Adjustment Kalman Filter (EAKF) within the Data Assimilation Research Testbed (DART) framework (Anderson et al. 2009) are used here to create a fully cycled atmospheric ensemble data assimilation system, which is hereafter referred to as 'A-DART' for brevity.

A one-month control ensemble analysis is created for which DART is identically configured as in Cavallo et al. (2013) assimilating the following conventional observations: Radiosondes, marine buoys, geostationary satellite atmospheric motion vectors (AMVs), METAR, Aircraft Communications Addressing and Reporting System data (ACARS), Automatic Weather Stations (AWS), and Global Positioning System (GPS) data. Cavallo et al. (2016) previously implemented the MITA increment method over a domain in the tropical North Atlantic Ocean with 36-km horizontal grid spacing using the conventional observations listed above with the ARW-WRF LAM model. A systematic model bias was diagnosed from the planetary boundary layer parameterization scheme, and was found to originate from erroneously specified sea surface temperatures (SSTs) over the North Atlantic Ocean. In addition, a systematic warm temperature bias in the free-troposphere was found to originate from the convective parameterization.

A large-scale upper-level wind bias is immediately evident in the A-DART control. This bias is evident from the analysis increments of wind and is greatest in the 45˚S-60˚S latitude range. Comparison of 6-h forecasts to Antarctic radiosonde profiles reveal a strong warm temperature bias in upper-levels, everywhere above 300 hPa. In addition to A-DART, this bias is apparent in forecasts from both the stand-alone AMPS and the GFS models. Given that ARW-WRF and AMPS physical parameterizations do not use a time-varying ozone profile, and that this time period exactly coincides with the annual depletion of Antarctic ozone, it is first hypothesized that a large-scale upper-level circulation bias is present due to too much shortwave heating over the Antarctic continent as a result of erroneously high ozone concentrations. Two experiments are then devised to test this hypothesis: (1) Implement time-varying, latitude-dependent ozone profiles in the physics, and (2) assimilate AMV data from polar-orbiting satellites. Both experiments result in only modest reductions in model bias. Therefore, the MITA increment method is then applied to determine where additional model bias originates.

While the MITA increment method has been more commonly used in global models, Cavallo et al. (2016) is the only study where it has been applied to a limited area mesoscale numerical weather prediction model. They found that analyzing forecast tendencies over shorter time intervals can be successful as long as the forecast model has appropriately spun-up to represent the mean analysis increment over the 6-h data assimilation cycling period, which in their study was sufficient after about 30-minutes. In the present study, this was found to be sufficient for forecast tendencies beginning at 1 hour, or 25 model time steps.

Statistically significant cold temperature biases are found in the boundary layer from 0-4 km above ground level (AGL) (700-1000 hPa), and in upper-levels from 10-20 km AGL (20- 300 hPa). This cold bias is in contrast to the warm bias seen from radiosonde comparisons, however, it is noted that radiosonde locations are predominantly located around 60˚S latitude in the Antarctic. A decomposition of the forecast tendency components reveals that the upper-level bias derives from the dynamics and longwave radiation tendencies equatorward of 60˚S latitude near the locations of geostationary satellite AMV observations. To test whether the geostationary AMV observations are biased, the case is re-cycled with Atmospheric Infrared

Sounder (AIRS) satellite retrievals and results in substantially reduced bias. Given the number of AIRS retrievals is much greater than the number of AMV observations, it is concluded that the geostationary AMVs exhibit a possible wind bias, where the wind from AMVs is too strong. Regarding the boundary layer bias, A-DART is again re-cycled to test whether well-documented cloud-phase errors in polar regions (e.g., Sandvik et al. 2007) contribute to a cold mid-tropospheric temperature bias. Model bias is reduced approximately by half, and the reduction in bias occurs primarily in the storm-track region over the Southern Ocean and over sea ice areas.

References

Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: The data assimilation research testbed: A community data assimilation facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296.

Cavallo, S. M., J. Berner, and C. Snyder, 2016: Diagnosing model error from time–averaged tendencies in the weather research and forecasting model. *Mon. Wea. Rev.*, **144 (2)**.

Cavallo, S. M., R. T. Torn, C. Snyder, C. Davis, W. Wang, and J. Done, 2013: Evaluation of the Advanced Hurricane WRF data assimilation system for the 2009 Atlantic hurricane season. *Mon. Wea. Rev.*, **141 (2)**, 523–541.

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

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. Meteor. Soc.*, **140 (679)**, 517–524.

Powers, J. G., A. J. Monaghan, A. M. Cayette, D. H. Bromwich, Y. H. Kuo, and K. W. Manning, 2003: Real-time mesoscale modeling over Antarctica: The Antarctic Mesoscale Prediction System. *Bull. Amer. Meteor. Soc.*, **84**, 1533–1545, doi:0.1175/ BAMS-84-11-1533.

Rabier, F., et al., 2010: The CONCORDIASI project in Antarctica. *Bull. Amer. Meteor. Soc.*, **91 (1)**, 69–86.

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

Sandvik, A., M. Biryulina, N. G. Kvamstø, and K. Stamnes, 2007: Observed and simulated microphysical composition of Arctic clouds: Data properties and model validation. *J. Geophys. Res.*, **112 (D05205)**, doi:10.1029/2006JD007351.

Using Forecast Temporal Variability to Evaluate Model Behavior

Carolyn. A. Reynolds, Elizabeth A. Satterfield, and Craig H. Bishop

Naval Research Laboratory, Monterey, CA, USA

We use simple diagnostics to quantify the temporal variability in analyses, a_i, and forecasts, f , for increasing time differences from i=1, 10 days. In addition, we introduce a diagnostic that reflects the day-to-day variability of the forecasts, d_i, for increasing forecast lead time i. In a perfect system, we expect $a_i > f$, and $a_1 > d_i$ due to the presence of uncorrelated analysis errors. We apply these diagnostics to control and perturbed ensemble initial states and forecasts from the NCEP and CMC global ensemble forecasting systems to demonstrate the utility of the diagnostics in quantifying aspects of forecast performance related to temporal variability. We relate the results to ensemble design and, in the case of CMC, a system upgrade.

While $a_i > f_i$ and $a_1 > d_i$ for most NCEP fields, which is expected in a perfect system, $a_i < f_i$ and $a_1 < d_i$ for several CMC fields, indicating that the CMC system may have excessive temporal variability as compared to the analyses. This is probably due to the excessive smoothing of the CMC analyses through the application of a digital filter (since replaced by 4DIAU in the deterministic global system as discussed in Buehner et al. 2015, and under development in the ensemble system). Trends in d_i illustrate how both the control and perturbed NCEP forecasts show a small but steady decrease in day-to-day temporal variability with increasing forecast time. In contrast, the CMC control forecasts show increasing temporal variability for temperature and humidity during the first few days, illustrating the spin-up of the system after the initial excessive digital filter smoothing.

The diagnostics also clearly reflect the upgrade in the CMC system on 13 February 2013. Before the upgrade, f_i was greater than a_i in the tropics for the CMC perturbed ensemble member for height, winds and temperature, which is not expected in a perfect system. After the upgrade, f_i was less than a_i in the tropics for the perturbed member. The trends in d_i for the perturbed member also change, remaining fairly constant or increasing before the upgrade, and decreasing after the upgrade. These differences are consistent with changes made to the stochastic physics perturbations in order to reduce excessive precipitation (Gagnon et al. 2013).

An advantage of these diagnostics is the ability to assess forecast temporal variability on different time scales without the need for very long forecast integrations. For example, the locations of the maxima in height field variability shift or extend from the North Atlantic and North Pacific jet regions for i=1 downstream to northern Europe and the eastern North Pacific for i=10. These shifts are consistent with patterns found in temporal filtering diagnostics of analyses time series that differentiate between regions of synoptic variability and blocking (e.g., Blackmon et al. 1977; Lau and Nath 1987; Cai and Van Den Dool 1991) using low-pass (>10 d) and band-pass (7-90 and 8-64 d) filters that could not be applied to the 10-d forecast integrations considered here.

Diagnostics measuring temporal variability are complementary to other diagnostics, such as those that focus on time-mean quantities or model bias (e.g., Klocke and Rodwell, 2014), spatial scale separation techniques (e.g., Harris et al. 2001), and techniques to quantify differences between forecast fields and reality as represented on the scales resolved by the data assimilation and forecast systems (e.g., Peña and Toth 2014). Using diagnostics to assess the accuracy of both temporal and spatial variability will become increasingly important as stochastic techniques to account for model uncertainty proliferate in ensemble forecasting systems, as both spatial and temporal correlations are often parameters in these schemes that need to be tuned. Potential future work includes consideration of other forecast systems, as well as an extension to a comparison with observations.

This work has been published in Reynolds, C. A., E. A. Satterfield, and C. H. Bishop, 2015: Using forecast temporal variability to evaluate model behavior. *Mon. Wea. Rev.*, **143**, 4785- 4804.

Using Ensemble Data Assimilation to Diagnose Model Uncertainty

Mark J. Rodwell[a], Simon T.K. Lang[1], N. Bruce Ingleby[a], Niels Bormann[a], Elias Hólm[a], Florence Rabier[a], David S. Richardson[a], Munehiko Yamaguchi[b]

[a]ECMWF, UK, [b]MRI, Japan

True model uncertainty is dependent on the instantaneous and local state of the system. To assess our representation of model uncertainty it is beneficial, therefore, to focus on the ensemble distribution at very short leadtimes. The traditional 'spread-error relationship' is not suitable for this assessment because, at short leadtimes, uncertainties in our knowledge of the truth cannot be ignored in the estimation of the error. Here, therefore, we develop and use a consistency relationship that adopts the ideas and terminology of ensemble data assimilation. In particular, this relationship decomposes the mean-squared departure of the ensemble-mean background forecast (relative to each observation) into a squared-bias term, an ensemble variance term and an observation uncertainty term. Any imbalance (or residual) in this budget highlights a deficiency of background reliability, and so the budget is known as the 'EDA reliability budget'. Results will be shown that demonstrate that the residual term is sensitive to the local and flow-dependent representation of model uncertainty. (It can also be sensitive to the modelling of observation error, but this can often be estimated by other means). The hope is, therefore, that this budget will facilitate future improvements in the representation of state-dependent model uncertainty.

Session 3
Operational schemes

Stochastic representations of model uncertainties in the IFS

Martin Leutbecher, Pirkka Ollinaho, Sarah-Jane Lock, Simon Lang, Peter Bechtold, Anton Beljaars, Alessio Bozzo, Richard Forbes, Thomas Haiden, Robin Hogan and Irina Sandu

ECMWF

The operational ensemble forecasts at ECMWF use the stochastic schemes SPPT and SKEB to represent model uncertainties. The talk describes the configuration of these schemes at ECMWF and shows the impact of the schemes on ensemble spread and probabilistic skill. Relative to an ensemble forecast with initial perturbations only, SPPT increases the ensemble spread considerably up to about 3 weeks in the extra-tropics and beyond 4 weeks in the tropics. The additional spread generated by SKEB is quite moderate. The representation of model uncertainties with SPPT+SKEB leads to statistically significant reductions of the continuous ranked probability score and even more pronounced reductions of the logarithmic score.

While SPPT is efficient in generating ensemble spread, it is recognised that its current formulation lacks physical consistency in several ways: (i) there are no flux perturbations at the top of the atmosphere and the surface that are physically consistent with the tendency perturbation in the atmospheric column; (ii) SPPT does not conserve water; (iii) SPPT includes ad-hoc elements like tapering in the boundary layer or stratosphere; (iv) SPPT is unable to represent multi-variate aspects of uncertainties, for instance it cannot alter the shape of the heating profile due to convection.

Progress towards the development of a new model uncertainty representation at the process-level is also reported. A stochastic scheme embedded within the IFS physics has been developed that introduces local stochastic perturbations of parameters and variables. The new scheme is referred to as the Stochastically Perturbed Parametrisation scheme (SPP). Through its formulation it maintains physical consistency in the perturbations and addresses the points (i)-(iv) mentioned above. SPP targets uncertainties that are known to matter based on the experience of the scientists working on the parameterisation of individual processes. SPP, like SPPT, converges to the deterministic IFS physics in the limit of vanishing variance. The current version of SPP can sample distributions for up to 20 different parameters and variables in the parameterisations of (a) turbulent diffusion and subgrid orographic drag, (b) radiation, (c) cloud and large-scale precipitation, and (d) convection. The development started from distributions with variances proposed by the scientists working on the parameterisations. Sensitivity experiments with modified variances informed decisions on adjusting the initial variance estimates. Among the tested variances, the best candidate configuration was selected based on increases in ensemble spread and more importantly the reduction of ensemble mean RMS error.

The different parameters and variables are sampled in SPP using independent random patterns with prescribed time and spatial decorrelation scales. The sensitivity to the decorrelation scales was tested. Among the scales tested, a configuration with decorrelation scales of 2000 km and 72 h resulted in the most skilful medium-range predictions. Both smaller scales (500 km and 6 h) as well as infinite scales (globally fixed perturbations) resulted in lower ensemble spread and also reduced probabilistic skill.

In order to better understand the different characteristics of SPPT and SPP, the tendency perturbations due to the two schemes have been compared. As expected, SPP generates considerable perturbations in the lowest model level in contrast to SPPT. In the free troposphere, the tendency perturbations of SPP appear to be more confined to localised regions than those of SPPT. Looking at area-averages, SPP generates about the same (less) variance in the tendencies perturbations than SPPT in the free troposphere in the tropics (in the extratropics) in the first hours of the forecast. However, at longer lead times, SPPT generates more variance in the tendencies everywhere except close to the surface.

The impact of SPP and SPPT on ensemble forecasts has been examined up to a lead time of 32 days. Compared to an experiment with initial perturbations only, both schemes significantly increase spread. The additional spread generated by SPP ranges between about 0.6 and 1.1 of the additional spread generated by SPPT depending on variable and region. SPP also leads to more skilful ensemble forecasts compared to the experiment with initial perturbations only. The reductions in CRPS due to SPP range between about 0.5 and 0.9 of the reductions in CRPS obtained with SPPT.

As part of the development of SPP, its impact on the model climate has been evaluated as well. Based on four 13 month integrations RMS errors of annual mean fields have been compared for runs with the unperturbed IFS model, with SPPT and with SPP. Relative to the run with the unperturbed model, the run with SPP consistently reduces RMS errors of the annual mean of a range of fields from tropical winds, precipitation, total column water vapour to top-of-the-atmosphere thermal radiation. SPPT also results in improvements of the climate but in a less consistent way. For instance, it clearly degrades total column water vapour. This is believed to be caused by its lack of humidity conservation.

The evaluation of SPP in the Ensemble of Data Assimilations (EDA) is ongoing. Like PPT, using SPP results in considerable additional spread in EDA analyses and EDA short-range forecasts. Preliminary results show that the additional spread introduced by SPP does not decrease towards the surface in the boundary layer as is the case with SPPT. In the free troposphere, the spread increase due to SPPT and due to SPP are of a similar order of magnitude.

Future extensions to the SPP scheme are envisaged that would address further uncertainties in (i) the vertical mixing above the boundary layer, (ii) the thermodynamic coupling between surface and atmosphere and (iii) trace gas sources. Future progress will also rely on process-oriented diagnostics of ensemble forecasts with the stochastic representation of model uncertainties.

Model error representation in Météo-France ensemble NWP systems

Laure Raynaud, M. Boisserie, F. Bouttier, L. Descamps, C. Labadie, Y. Michel

Météo-France – CNRM/GMAP/RECYF

A variety of ensemble systems are currently operationally running or under development at Météo-France. At the global scale, operational ensemble data assimilation (EDA) and ensemble prediction systems (EPS) are based on the Arpège model, while at the convective scale EDA and EPS are being developed based on the non-hydrostatic limited- area Arome-France model. Experiments with these different systems have shown that the representation of model errors is a key aspect, hence appropriate strategies have been implemented in order to account for their contribution. We provide in the following paragraphs a short description of the methods used in each system.

The Arpège EDA is an ensemble of perturbed 4D-Var assimilations primarily developed to compute flow-dependent background-error covariances for the deterministic 4D-Var assimilation. A methodology has been proposed to estimate model-error global contribution from diagnostics relative to the minimum of the variational cost function (Raynaud *et al.*, 2012). This information is then used to implement a flow-dependent adaptive multiplicative inflation of background perturbations after each forecast step. This leads to an increase of the ensemble spread by roughly a factor of 2, which improves the consistency of ensemble variance estimates with innovation-based estimates. Positive impacts of inflated ensemble covariance estimates on the analysis and forecast scores are also observed.

Perturbed states from the global EDA also provide initial conditions for the Arpège EPS (Descamps *et al.*, 2015), which is currently running on a stretched grid with a 10km resolution over France. Model error in the EPS is accounted for with the multiphysics approach, which is considered to provide a valuable flow-dependent sampling of the uncertainty in the physical parametrizations. It is based on ten different physical parametrization sets, including the Arpège deterministic physical package, designed from different schemes for turbulence, shallow convection, deep convection and for the computation of oceanic fluxes. Scores indicate that the multiple parametrization set increases the ensemble spread, especially for 850hPa temperature, and also slightly improves the ensemble resolution.

To complement the global systems, convective-scale ensembles for both assimilation and forecasts are currently under development at Météo-France. The near-operational Arome-EPS is running with a 2.5km horizontal resolution, and model error is represented with the SPPT scheme (Bouttier *et al.*, 2012), which is a limited-area version of the ECMWF scheme. As expected, SPPT enhances ensemble spread, especially in the lower troposphere, and it generally improves the ensemble performance for the prediction of surface weather variables, as measured by the spread-skill relationship and various other probabilistic scores. The SPPT scheme is also shown to have a significant impact on rain forecasts, although it does not directly perturb condensed water species. On the other hand, the SPPT is shown to produce a drying of the lower atmosphere that should be further investigated.

Finally, preliminary versions of the Arome EDA, running with a 4km horizontal resolution, combines this SPPT scheme with a multiplicative inflation of background states to represent model error contribution. The impact of this EDA setting is currently being examined regarding both the initialization of the Arome-EPS and the estimated ensemble background-error covariances then used in the deterministic Arome analysis scheme.

The representation of model-error in the ensemble systems of Météo-France is an active area of research. While the currently used strategies have significant positive impacts on forecast scores there is still room for improvement. Future works will examine the application of the SPPT scheme in the Arpège EPS, as a potential replacement for the multiphysics approach which is typically difficult to maintain in an operational environment.

Going from the global SPPT to the independent SPPT scheme (iSPPT) in the Arome-EPS will also be considered, based on the preliminary promising results reported with this scheme. More generally, future developments regarding model error will focus on more physically-based strategies.

Stochastic Parameterization Development in the NOAA/NCEP Global Forecast System

Philip Pegion[1,2], Gary Bates[1,2], Maria Gehne[1,2], Thomas Hamill[1], Walter Kolczynski[3,4], Jeffrey Whitaker[1], and Yuejian Zhu[3]

1-NOAA/Earth Science Research Lab, Physics Sciences Division. Boulder, CO

2- University of Colorado. CIRES Boulder, CO

3- NOAA/NCEP/Environmental Modeling Center. College Park, MD

4- I.M. Systems Group. College Park, MD

The current operational Global Forecast System (GFS) run at the National Centers for Environmental Prediction (NCEP) uses very different methods of treating model uncertainty for different applications. The short range ensemble that is used for data assimilation (9 hour forecasts) incorporates four stochastic parameterization schemes working in concert during the 9-hour forecast. The medium range ensemble (out to day-16) currently uses an old method developed at NCEP that applies random perturbations once every 6-hours and requires all of the ensemble members to be able to communicate with each other during the forecast, which limits the number of ensemble members that can be run.

There are ongoing projects to put this suite of stochastic parameterizations into the medium range forecast, and testing shows a well calibrated ensemble system (ensemble spread matches ensemble mean error) for upper-air fields such as 850 hPa Temperature and winds, but there are still deficiencies in ensemble spread in sensible weather elements, such as precipitation, and 2 metre temperature. To remedy this, new methods are currently being tested to address model error that is related to the land and sea-surfaces interface with the atmospheric model.

During the development and testing of the stochastic parameterization schemes, we have found that the Stochastically Perturbed Parameterization Tendencies Scheme (SPPT) produces a positive bias in precipitation, which was opposite to what we expected from the literature, and modifications to address this issue will be presented. Plans of integrating stochastic parameterizations at the process level of the new physicals being developed for the Next Generation Global Predication System (NGGPS) will also be presented.

Model Error Representation in the Canadian Ensemble Prediction Systems

Leo Separovic[1], Martin Charron[1], Amin Erfani[2], Normand Gagnon[2], Ayrton Zadra[1] and Paul Vaillancourt[1]

[1]Recherche en Prévision Numérique Atmosphérique, Meteorological Research Division, Environment and Climate Change Canada, Dorval, Quebec, Canada

[2]Meteorological Service of Canada, Environment and Climate Change Canada, Dorval, Quebec, Canada

State-of-the-art ensemble prediction systems typically use deterministic physical parameterisations (single or multiple choices) and ad hoc techniques for sampling uncertainties originating from the subgrid-scale processes, truncation and diffusion. Random perturbations of poorly constrained physics parameters, although having potential to improve the skill of ensemble prediction systems, may lead to under- or over-dispersive forecasts. Most operational centres thus resort to adding perturbations directly to the physics tendencies and applying the stochastic kinetic energy backscatter algorithm. While these ad hoc methods are relatively simple to apply they are rather unsatisfactory from a more fundamental perspective. In the long term, development of inherently stochastic physics appears to be a more appropriate approach to represent model errors originating from the unresolved-scale processes.

In this presentation we will discuss the current methodological approach at Recherche en Prévision Numérique Atmosphérique (RPN-A), as well as the most recent developments aimed at stochastic parameterisations of physical processes and their application in the ensemble prediction. Currently, we are working on stochastic parameterisation of convection based on the Plant-Craig (PC) approach applied to the Bechtold convection scheme. The Bechtold scheme was modified in order to allow for multiple random plumes of different characteristics, such as cloud lifetime and radius at the cloud base, whereas, the plumes' statistical properties are derived from the PC theory. In addition to convection, we are also investigating ways to introduce stochasticity in a TKE-based boundary layer scheme.

Representing model error in the Met Office convection permitting ensemble prediction system

Anne McCabe, Richard Swinbank, Warren Tennant and Adrian Lock

Met Office, UK

Recently, many weather centres have developed ensembles at the convective-scale to provide detailed forecasts over regions of particular interest. While these high resolution models are able to provide very realistic looking forecasts, there continue to be cases where the observed weather has not been captured by the ensemble, and a common complaint of forecasters is that the ensemble is over-confident, with the members too similar. This lack of spread, typical of ensemble prediction systems, is at least partly attributed to uncertainty arising from the model itself.

At the Met Office, we routinely run a limited-domain convection-resolving ensemble prediction system (EPS) over the British Isles, known as MOGREPS-UK. The ensemble consists of twelve members (one control and eleven perturbed) and is run out to T+36, every 6 hours, four times a day. It is widely recognises that a successful EPS should represent all forms of uncertainty in the forecast, namely, uncertainty arising from (i) the initial conditions, (ii) the boundary conditions, and (iii) the model physics (Buizza *et al.* 1999). MOGREPS-UK is nested in the Met Office global ensemble, with initial and boundary conditions directly downscaled from the corresponding global ensemble members (since March 2016, each ensemble member is initialised by re-centering over the UKV analysis, as described in Tennant, 2015). The question then remains of how best to represent model uncertainty in MOGREPS-UK.

When it comes to representing model uncertainty at the convective scale, a natural starting point it to use one of the established stochastic physics schemes used at global and meso-scales. It is not as straightforward as simply choosing one such scheme and implementing it, as different physical processes dominate at the convective scale and any such scheme will first need to be adapted to the relevant small-scale processes.
Determining the nature of the model uncertainty is also not straight forward – ideally, we would have a full evaluation of the type and extent of model uncertainty and could match our choice of stochastic physics scheme appropriately. In the absence of such an evaluation, we rely on the experience of forecasters and parametrization modellers to identify key areas in the model that are either inherently uncertain or known to be inadequately represented.

One stochastic physics approach used in the Met Office global EPS is the Random Parameter (RP) scheme (Bowler *et al.* 2008). The RP scheme perturbs a set of parameters from relevant parametrizations and varies them stochastically throughout the forecast. Like the Stochastically Perturbed Physics Tendency (SPPT) schemes (*e.g.* Buizza *et al.* 1999; Charron *et al.* 2010), the RP scheme is used to represent the knowledge uncertainty in the physics parametrizations. The RP scheme has the advantage over SPPT that it targets known areas of uncertainty within the parametrizations, it produces physically realistic tendencies, is conceptually simple and cheap to implement. The limitations of the scheme are that it needs regular updates as physics parametrizations are changed and developed, and the choice of parameters and their ranges can be subjective. Historically, the RP scheme has shown only a modest impact on spread and skill and there is a question over the most appropriate way to evolve the parameters through time. We chose this approach over the SPPT scheme as, while the SPPT scheme has the potential to increase the spread of the ensemble, it is applied in a general way to the atmospheric variables, and not linked to any particular physical processes or physical understanding.

To make the RP scheme suitable for the convective scale of MOGREPS-UK, we have revised the RP algorithm to make it easily adaptable to different spatial and temporal resolutions. In the original RP scheme as used in the global ensemble, the parameters are updated every 3 hours with shocks up to a third of the parameter range.
For MOGREPS-UK, we use the revised algorithm and the parameters are updated more frequently (every 5 minutes) with smaller perturbations so that the parameters take a smoother, more slowly varying path throughout the forecast. We apply the revised algorithm to a set of parameters in MOGREPS-UK chosen to represent uncertainty in the parametrizations relevant to the convective-scale UK forecast. These parameters are from the boundary layer and micro-physics parametrizations and cover processes including cloud formation, rain rate, turbulent mixing, entrainment at the boundary layer top and near-surface droplet settling.

Fog forecasting is of particular interest to the UK forecast - accurately predicting the timing, location and extent of fog can be challenging, and the implications of getting the forecast wrong can cause major problems, particularly in the aviation industry. One of the reasons that fog is so difficult to forecast, is that it depends on local scales that may be inherently uncertain and poorly observed. The forecast of fog is affected by many of the parameters used in the revised RP scheme. Two of the new parameters have been chosen to explicitly address the uncertainty in fog formation. The first is related to droplet settling and fog dissipation and is a parameter to which fog formation in the model is known to be particularly sensitive (Wilkinson *et al.* 2013). The second is related to the contribution of wind shear to entrainment at the BL top and addresses a known issue in the model where fog is erroneously lifted into stratocumulus (see discussion in Price *et al.* 2015).

To assess the impact of the revised RP scheme on MOGREPS-UK, we consider fog case studies and the objective verification statistics of two separate month-long trial periods covering winter and summer. In each case, a reference ensemble (straight downscaler, no stochastic physics) was compared with the RP ensemble (as the reference ensemble plus the revised RP scheme). The case study results show that the revised RP scheme in MOGREPS-UK increases the variability in the forecast of fog while producing physically realistic forecasts. This increase in variability results in a reduction of over-confident probabilities of fog and therefore a more useful probabilistic forecast. The case studies also show the encouraging results that the revised RP scheme enables the ensemble to capture observed fog events otherwise missed by the forecast.

For both the winter and summer trials, the RP ensemble shows a small increase in spread for surface temperature and 10m wind compared with the reference ensemble. Probabilistic scores show an overall improvement in ensemble skill for visibility and surface temperature. Positive results were also seen for cloud base height and fog, however statistical tests (using the non-parametrix Wilcoxon test for paired data as described in Hamill 1999) indicate that these last two results are not statistically significant – we hypothesise that this is because there are too few fog and low cloud cases in a month and that a better test would be to run the ensemble over a series (30+days) of interesting fog case studies.

Overall, we have found the revised RP scheme to have a positive impact on MOGREPS-UK with particular benefit to fog forecasting. The scheme has been running operationally in MOGREPS-UK since March 2016. Currently, work is undergoing at the Met Office to extend the RP scheme to parameters in the land-surface scheme, and Warren Tennant (Met Office) is trialling a spatially varying version of the RP scheme in the global ensemble. In parallel to the work described here, Adrian Lock (Met Office) has developed a scheme to stochastically perturb potential temperature in the lower part of the boundary layer in conditionally unstable regimes. This scheme has a positive effect on the initiation of convection and is currently being run operationally at the Met Office in both the single high resolution model (UKV) and MOGREPS-UK. Refinements to the BL perturbations are also underway and there are plans to apply these perturbations to other variables in the boundary layer (Adrian Lock and Carol Haliwell Met Office, and Peter Clark, University of Reading).

References

Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE. 2008. The mogreps short-range ensemble prediction system. Quarterly Journal Royal Meteorological Society 134: 703–722

Buizza R, Miller M, Palmer TN. 1999. Stochastic representation of model uncertainties in the ecmwf ensemble prediction system. Quarterly Journal of the Royal Meteorological Society 125: 2887–2908.

Charron M, Pellerin G, Spacek L, Houtekamer P, Gagnon N, Mitchell HL, Michelin L. 2010. Toward random sampling of model error in the Canadian ensemble prediction system. Monthly Weather Review 138: 1877 – 1901.

Hamill TM. 1999. Hypothesis tests for evaluating numerical precipitation forecasts. Weather and Forecasting 14: 155–167.

Price J, Porson A, Lock A. 2015. An observational case study of persistent fog and comparison with an ensemble forecast model. Boundary-Layer Meteorology 155: 301–327.

Tennant W. 2015. Improving initial condition preturbations for MOGREPS-UK. Quarterly Journal of the Royal Meteorological Society 141: 2324–2336. Doi:10.1002/qj.2524.

Wilkinson JM, Porson ANF, Bornemann FJ, Weeks M, Field PR, Lock AP. 2013. Improved microphysical parametrization of drizzle and fog for operational forecasting using the met office unified model. Quarterly Journal Royal Meteorological Society 139: 488–500. Doi:10.1002/qj.1975.

Model Uncertainty Representation in COSMO-DE-EPS

Susanne Theis, Michael Buchhold, Christoph Gebhardt, Regina Kohlhepp, Andreas Röpnack

Deutscher Wetterdienst, Offenbach, Germany

The convection-permitting ensemble prediction system COSMO-DE-EPS has been running operationally at the German Weather Service (DWD) since 2012. It consists of 20 ensemble members with a grid spacing of 2.8 km. Ensemble forecasts are started 8 times per day with a lead time of 27 hours (03UTC run: 45 hours). The model domain comprises Germany and neighbouring areas. The ensemble represents forecast uncertainties by variations in initial conditions, lateral boundary conditions and model physics (Gebhardt et al., 2011; Peralta et al, 2012).

In the operational version of COSMO-DE-EPS, model uncertainties are represented by a multi-parameter approach. The method is closely connected to the advice of parametrization experts and it has the advantage of fast implementation. It is targeted to user-specific aspects of the forecast (e.g. precipitation, 2m-temperature). As a starting point, ensemble experts and parameterization experts come together, discuss forecast uncertainties and pre-select promising candidates for parameter perturbations (e.g. entrainment rate of shallow convection, critical value of normalized oversaturation in the microphysics scheme, asymptotic mixing length in the turbulence scheme). The discussion already refers to the evaluation criteria described below.

The current implementation is fairly simple. Some ensemble members run with the default value of the parameter and some run with an alternative value. These values are constant within the model domain and during forecast integration (0-27 hrs / 0-45 hrs). For some parameters, there is only one alternative value. For other parameters, there are two alternative values, but their differences to the default are not necessarily symmetric. Today, the operational COSMO-DE-EPS perturbs 5 parameters (5 default values and 7 alternative values).

Before operational implementation, the parameter perturbations have been evaluated with regard to the following criteria: (1) ensemble spread and (2) ensemble quality, including the criterion of similar quality and bias in each individual member (Gebhardt et al., 2011).

(1) In terms of ensemble spread, case studies indicate whether the multi-parameter approach is able to capture events that would have been missed otherwise. Further evidence is added by statistical analysis of many cases. In the statistical context, it can be crucial to diagnose spread in a regime-dependent way (Keil at al., 2013). It is also interesting to estimate "spread in location", detecting to which extent the alternative scenarios cause precipitation to occur at locations different from the defaultscenario.

(2) In terms of ensemble quality, each member is evaluated individually and also the entire ensemble is evaluated (e.g. Brier Score, CRPS, spread-error relation). Verification of individual members explicitly looks at their forecast bias and at their forecast quality. If the members are similar in bias and quality, there is evidence that ensemble spread is not simply generated by differing biases and also that members may be treated as equally likely. The parameter perturbation is rated as suitable for operational use only if these criteria are fulfilled. In addition to the statistical evaluation, visual inspection of individual fields is beneficial, because it can detect unrealistic behaviour in observation-sparse regions (e.g. over sea).

As the development involves a manual selection and testing procedure, the current approach requires some effort in maintenance. Adaptations become necessary whenever the model version or the user-specific aspects change. Recently, the renewable energy sector has become a potential user of the forecast, so there is an incentive to improve COSMO-DE-EPS with regard to additional aspects (e.g. low-level clouds and solar radiation, low-level jet and wind in 100m height). Similarly to other convective-scale ensembles, these variables suffer from a lack of ensemble spread in COSMO-DE-EPS. Current development has attained improvements by enlarging the set of perturbed parameters, so COSMO-DE-EPS will soon perturb 9 parameters (9 default values and 13 alternative values).

Further attempts at optimization require much effort with a relatively small gain. One issue is the optimal combination of different parameter values in each member. Current development aims at replacing the fixed setting by a randomized one, so the combination of different parameter values would be a random result and not subject to optimization. The random scheme would be active at each forecast start. It would assign the various parameter values to each member and they would still be constant during forecast integration. Verification indicates that forecast quality is not degraded by the randomly combined parameter perturbations.

The multi-parameter approach covers an incomplete portion of the entire model uncertainty. A fully stochastic approach is believed to have more potential if appropriately developed. As a medium-term goal, DWD's research team on "physical processes" is developing a stochastic perturbation of model tendencies (E. Machulskaya, DWD, see poster at this workshop). The approach consists in a prognostic equation of model error. The equation contains parameters which specify noise amplitude and its autocorrelation in space and time. By applying a proxy for model error, it can be shown that these parameters are statistically related to resolved model variables. This may be an opportunity to use resolved model variables as predictors for the evolution of model error, resulting in a flow-dependent model of the model error.

As a long-term goal, DWD also supports research via the Hans Ertel Centre for Weather Research. Within this framework, a stochastic parameterization of shallow cumulus convection is developed at the Max Planck Institute for Meteorology (Sakradžija, 2015). Also within this framework, a model error representation within ensemble based data assimilation is developed at the Ludwig-Maximilians University in Munich (M. Sommer, LMU, see poster at this workshop).

References:

Gebhardt, C., Theis, S.E., Paulat, M. and Z. Ben Bouallègue (2011): Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. Atmospheric Research 100, 168-177.

Keil, C., Heinlein, F. and G.C. Craig (2013): The convective adjustment time-scale as indicator of predictability of convective precipitation. Q. J. R. Meteorol. Soc. 140, 480-490.

Peralta, C., Ben Bouallègue, Z., Theis, S.E., Gebhardt, C. and M. Buchhold (2012): Accounting for initial condition uncertainties in COSMO-DE-EPS. J. Geophys. Res., 117 (D7), doi:10.1029/2011JD016581.

Sakradžija, Mirjana (2015): A stochastic parametrization of shallow cumulus convection for high-resolution numerical weather prediction and climate models. Reports on Earth System Science, Max-Planck-Institute for Meteorology, Hamburg, 144 pp.

Session 4
Latest research

Stochastic parameterization: Towards a new view of Weather and Climate Models	Judith Berner
Model uncertainty in global ocean models: Stochastic parametrizations of ocean mixing	Stephan Juricke
Representing model uncertainty for climate forecasts	Antje Weisheimer
On the diagnosis of model error statistics using weak-constraint data assimilation	Neill Bowler
A comparison of the model error and state formulations of weak-constraint 4D-Var	Amos Lawless
Using trajectories of ensemble analyses and tools from weak constraint 4D-Var to represent model uncertainty	Craig Bishop
Model error representation in convection-permitting forecasts and ensemble data assimilation	Glen Romine
Improving the Stochastically Perturbed Parametrisation Tendencies scheme using High-Resolution Model Simulations	Hannah Christensen
Impact of a stochastic deep convection scheme using cellular automata in the meso-scale ensemble prediction system; Harmon-EPS	Lisa Bengtsson
Parameter uncertainty of chaotic systems	Heikki Haario
Inducing tropical cyclones to undergo brownian motion	Daniel Hodyss
Observational based stochastic convection parameterization	Jesse Dorrestijn

Parameter uncertainty of chaotic systems	Heikki Haario
Inducing tropical cyclones to undergo brownian motion	Daniel Hodyss
Observational based stochastic convection parameterization	Jesse Dorrestijn

Stochastic Parameterization: Towards a new view of Weather and Climate Models

Judith Berner

National Center for Atmospheric Research, Boulder, Colorado, USA

The last decade has seen the success of stochastic parameterizations in short-term, medium-range and seasonal forecasts: operational weather centers now routinely use stochastic parameterization schemes to better represent model inadequacy and improve the quantification of forecast uncertainty.

Developed initially for numerical weather prediction, the inclusion of stochastic parameterizations not only provides better estimates of uncertainty, but it is also promising for reducing longstanding climate biases and is relevant for determining the climate response to forcing such as an increase of CO2.

Recent work from different research groups is reviewed. It shows that the stochastic representation of unresolved processes in the atmosphere, oceans, land surface and cryosphere of comprehensive weather and climate models (a) gives rise to more reliable probabilistic forecasts of weather and climate and (b) reduces systematic model bias.

We make a case that the use of mathematically stringent methods for the derivation of stochastic dynamic equations will lead to substantial improvements in our ability to accurately simulate weather and climate at all scales. Recent work in mathematics, statistical mechanics and turbulence is reviewed, its relevance for the climate problem demonstrated, and future research directions outlined.

Model uncertainty in global ocean models: Stochastic parametrizations of ocean mixing

Stephan Juricke[1], Tim Palmer[1], Antje Weisheimer[1,2], Laure Zanna[1]

[1]University of Oxford, Atmospheric, Oceanic and Planetary Physics, Department of Physics
[2] ECMWF

Model resolution of current state-of-the-art ocean models, especially on timescales of seasons to decades, is of the order of 100km. Mesoscale eddies in the ocean are, however, about an order of magnitude smaller. Therefore most ocean models used for seasonal or decadal predictions utilize resolutions where mesoscale eddies are not or only partially resolved. As a consequence, the effects of unresolved eddies on the resolved, large scale circulation need to be parametrized.

Eddies are just one of many sub-grid scale ocean processes that are not explicitly resolved in ocean models. Most of the commonly implemented parametrizations deal with unresolved horizontal and vertical sub-grid scale mixing processes that can vary strongly with time and location. Oftentimes these parametrizations and their parameters are imperfectly constraint, due to missing process understanding or unavailable observations. In this context, stochastic parametrizations can help to introduce a measure of uncertainty estimation in the model that deals specifically with the uncertainty originating from parametrized processes. Furthermore, stochastic perturbations can be used to represent not only the mean impact of the sub-grid scales on the resolved flow but also reintroduce some of the sub-grid scale variability that is not captured by classical deterministic parametrization schemes.

One approach to reintroduce sub-grid scale variability as well as implement uncertainty estimates in current state-of-the-art climate models is to identify crucial, imperfectly constrained parameters or parametrization tendencies and perturb those in a symmetric, multiplicative way. We identified three different parametrizations in the NEMO global ocean model with a 1 degree horizontal resolution for which certain parameters meet these criteria:

- The Gent-McWilliams parametrization parametrizes unresolved eddy advection of temperature and salinity, especially in the Southern Ocean. It generally leads to a flattening of overly steep isopycnal slopes, but the exact amplitude of this process is quite uncertain.
- The strength of vertical mixing in the NEMO ocean model is based on a parametrization using a prognostic turbulent kinetic energy formulation, which defines the intensity of vertical mixing especially in the upper ocean.
- In the case of unstable stratification, an enhanced vertical mixing parametrization is used to stabilize the water column.

The amplitudes and timescales of these three mixing parametrizations were perturbed by stochastically perturbing important parameters used in their formulation. The applied random perturbations were tuned in their amplitude and exhibited temporal and spatial correlations.

The results show that in uncoupled forced ocean-only simulations the perturbations to sub-grid scale mixing parametrizations lead to an increase in low frequency variability in eddy-active regions for a variety of variables, even though the perturbations themselves exhibit high frequency variability. Interannual variability for sea surface temperature, sea surface height, integrated heat content and zonally averaged streamfunction was increased predominantly in the Southern Ocean and along western boundary currents such as the Kuroshio region. This is in accordance with missing low frequency variability in these regions when compared to observations and reanalysis products. Therefore, including high frequency perturbations in parametrizations of horizontal and vertical mixing improves the representation of low frequency variability in the ocean model, an effect that can also be achieved by increased resolution but at increased computational costs. However, the effect of the stochastic perturbations is not sufficient to fully compensate for the effects of the missing eddy variability in the 1 degree ocean model.

In coupled ECMWF seasonal forecasts with the same ocean model and horizontal ocean resolution, the stochastic perturbations increase the ensemble spread again especially in the eddy-active regions of the Southern Ocean and the western boundary currents, for variables such as sea surface temperature and upper ocean heat content (see figure). This is the case for months 3 to 10 of the 10-month forecasts. The increase in spread leads to an increase in forecast reliability in the Southern Ocean, where ensemble spread strongly underestimates the forecast error. The forecast error itself is affected by the stochastic perturbations as well, but while the error is reduced in some regions, in other regions the stochastic perturbations lead to an increase in forecast error. However, it should be noted that the seasonal forecasts are not retuned after implementation of the stochastic schemes. Also, the effect of the increased spread is generally larger than the effect of the schemes on forecast error. Future studies will analyse the seasonal forecasts in more detail and will also introduce new stochastic perturbation schemes.

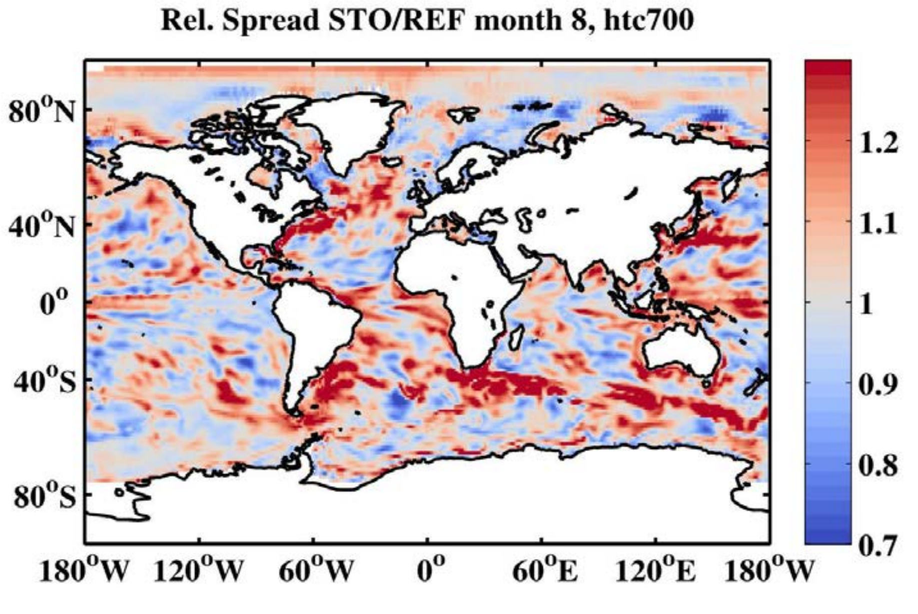


Figure 1 Relative changes in spread between the stochastically perturbed forecasts (STO) and the reference forecasts (REF) for 20-member ensembles, averaged over the years 1981 to 2005, initialised in May and integrated for 10 months. Shown are the results for 700 meter heat content and month 8, i.e. December. The spread in the eddy-active regions is increased by more than 30 %.

Representing model uncertainty for climate forecasts

Antje Weisheimer

ECMWF & University of Oxford, UK

Initialised forecasts on climate time scales from months to seasons ahead are routinely issued at ECMWF using its coupled Earth System model. The physical basis for such estimates arises from the effect on the atmosphere of predictable seasonal-timescale signals arising from the ocean and the land surface. Monthly and seasonal predictions provide estimates of forecast of weekly and seasonal- means of the coming month and season.

ECMWF's currently operational seasonal forecast model (System 4 or S4) consists of the atmospheric component IFS that contains an explicit representation of model uncertainties through the Stochastically Perturbed Physical Tendencies (SPPT) and the Stochastically Perturbed kinetic energy BackScatter (SPBS) schemes, and the ocean model NEMO. A set of retrospective seasonal forecasts over the period 1981-2010 with and without stochastic parametrisations was used to estimate the impact of SPPT and SPBS on the model climatology and biases and on the forecast performance (for details see *Weisheimer et al., 2014*).

It was found that the schemes (primarily SPPT) lead to a reduction of the overly active tropical convection and a reduction of the associated model biases for OLR, total cloud cover, precipitation and winds especially in the tropical Western Pacific which is a crucial geographical region for ENSO. It was further found that SPPT improves the frequency of MJO events in all 8 phases and increases their amplitude. For a discussion of the impact of the different SPPT scales on tropical precipitation see also *Subramanian et al., 2016*. In terms of forecast quality, several examples of improvements due to stochastic perturbations were presented: more skilful and reliable tropical temperature and precipitation forecasts up to two weeks, significant increases in the MJO spread, and a better calibrated forecasts of SSTs in the western tropical Pacific. These improvements are a combination of the effects of having a beneficial increase in the ensemble spread and a reduction in the ensemble-mean forecast errors with stochastic perturbations.

Some hypotheses have been put forward as to why SPPT leads to a systematic shift of the distribution of precipitation in the tropics. These include mathematical effects of the product of two distributions of random variables (as in the multiplicative SPPT scheme), the existence of non-linear physical thresholds affected by the stochastic perturbations, e.g. the trigger for deep convection or super-saturation, the tapering of the boundary layer in SPPT and related inconsistencies, e.g. in the surface fluxes, the asymmetric nature of specific humidity and precipitation, and the tuning of the model in its deterministic formulation rather than the stochastic one (noise-induced drift). Work is currently under way to better understand the physical mechanisms behind those.

A problem with the surface moisture fluxes in SPPT was found which indicated a drying in the atmosphere compared to the unperturbed control members. This lead to large P-E imbalances which are not acceptable for climate simulations with SPPT (*Davini et al., 2016*). A fix that empirically corrects for the loss of humidity was introduced and showed that the flux problem could be eliminated. At the time of writing it is very likely that this SPPT fix to conserve humidity will become operational in the next IFS model update.

The land surface is a key component for seasonal prediction due to its inherent longer time scales. However, there exist large uncertainties in poorly constrained land surface parameters that are often unquantified. We have introduced different schemes to account for such uncertainties by explicitly representing them in the land surface model of the coupled ECMWF model (*MacLeod et al., 2015*). The schemes perturb two key hydrological parameters in either a static or stochastic way. We have also tested a stochastic tendency perturbation scheme for soil moisture and soil temperature using different settings of the spectral pattern generator. The results are promising and show improved probabilistic forecasts for cases of strong land-atmosphere coupling like the European heat summer of 2003. We also find a general improvement in the reliability of extreme soil moisture forecasts when the parameter perturbations are activated.

Acknowledgement: I would like to thank Jost von Hardenberg (ISAC-CNR, Torino, Italy), David MacLeod and Aneesh Subramanian (University of Oxford) and Simon Lang (ECMWF) for their input to this workshop contribution.

References

Davini, P., J. von Hardenberg, S. Corti, A. Subramanian, H. Christensen, S. Juricke, P.A.G Watson, A. Weisheimer and T.N. Palmer (2016). Climate SPHINX: evaluating the impact of resolution and stochastic physics parametrisations in climate simulations. submitted to *GMD*.

Macleod, D.A, L.H. Cloke, F. Pappenberger and A. Weisheimer (2015). Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land. Q. J. R. Meteorol. Soc. , doi: 10.1002/qj.2631. Available from http://onlinelibrary.wiley.com/doi/10.1002/qj.2631/abstract?campaign=wolacceptedarticle

Subramanian, A., A. Weisheimer, T.N. Palmer, F. Vitart and P. Bechtold (2016). Impact of stochastic physics on tropical precipitation in the coupled ECMWF model. submitted to *Q. J. R. Meteorol. Soc.*

Weisheimer, A., S. Corti, F. Vitart and T.N. Palmer (2014). Addressing model error through atmospheric stochastic physical parameterisations: Impact on the coupled ECMWF seasonal forecasting system. *Phil. Trans. R. Soc. A* , **372**, 201820130290, doi: 10.1098/rsta.2013.0290. Available from http://rsta.royalsocietypublishing.org/content/372/2018/20130290.abstract

On the diagnosis of model error statistics using weak-constraint data assimilation

N. E. Bowler, Met Office, FitzRoy Road, Exeter, EX1 3PB, UK

May 13, 2016

Abstract

Outputs from a data assimilation system may be used to diagnose observation and background error statistics, as has been demonstrated by previous researchers. In this study, that technique is extended to diagnose model-error statistics using a weak-constraint data assimilation. It deals with a set of observations over a time window and uses the temporal distribution to separate model errors from errors in the background forecast. In idealised tests this method is shown to be able to successfully distinguish between model, background and observation errors. The success of this method depends on the prior assumptions included in the weak-constraint data assimilation and how well these describe the true nature of the system being modelled.

1 Introduction

It has long been recognised that computer models of complex physical processes are imperfect. What has been less clear is how to estimate the magnitude and structure of these imperfections. In particular, how does one differentiate errors in the numerical model from those in the observations and from any chaotic growth of small errors intrinsic to the system being modelled?

Recently Todling (2015) introduced a method to diagnose the model-error covariance from a pair of data assimilations — one of which is a filter and the other a smoother (able to use future observations). This system is described as sequential, since it is devised for a set of observations which are available at discrete times, rather than being spread over a given time window.

2 Weak-constraint data assimilation

To find the analysis in a system which is affected by model error one can use weak-constraint data assimilation. At a set of times we allow the analysis trajectory to depart from the solution given by the nonlinear model according to

$$\mathbf{x}_i = M_i(\mathbf{x}_{i-1}) + \boldsymbol{\eta}_i \tag{1}$$

where $\mathbf{x}_i$ is the model state at time i and M_i is the nonlinear model propagator from time $i-1$ to time i. At each time we are permitting a modification of the model state of $\boldsymbol{\eta}_i$.

If we use this perturbed model to define a four-dimensional state $\underline{\mathbf{x}}$, then we can write the weak-constraint cost function as (Trémolet, 2006)

$$\begin{aligned} J_{\text{weak}} = \frac{1}{2}(\mathbf{x}-\mathbf{x}_{\text{b}})^{\text{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_{\text{b}}) + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\eta}_i^{\text{T}}\mathbf{Q}_i\boldsymbol{\eta}_i + \\ \frac{1}{2}(\mathbf{y}-H(\underline{\mathbf{x}}))^{\text{T}}\mathbf{R}^{-1}(\mathbf{y}-H(\underline{\mathbf{x}})). \end{aligned} \tag{2}$$

where $\mathbf{Q}_i$ is the model-error covariance for time i and n is the total number of times at which a modification is allowed. If we assume that the modification at each time is the same, then we may write the total effect of the modifications on the trajectory as

$$\underline{\mathbf{N}}\boldsymbol{\eta} = \begin{pmatrix} 0 \\ \mathbf{M}_{1\leftarrow 1} \\ \mathbf{M}_{2\leftarrow 2} + \mathbf{M}_{2\leftarrow 1} \\ \cdots \\ \sum_{i=1}^{t}\mathbf{M}_{t\leftarrow i} \end{pmatrix}\boldsymbol{\eta} \tag{3}$$

where $\mathbf{M}_{j\leftarrow i}$ is the linear propagator of the numerical model from time i to time j. The final term in equation (2) then becomes

$$\begin{aligned} J_o = \frac{1}{2}(\mathbf{y} - H(\underline{M}(\mathbf{x}_{\text{b}})) - \mathbf{H}\underline{\mathbf{M}}\mathbf{d}_{\text{b}}^{\text{a}} - \mathbf{H}\underline{\mathbf{N}}\boldsymbol{\eta})^{\text{T}}\mathbf{R}^{-1} \\ (\mathbf{y} - H(\underline{M}(\mathbf{x}_{\text{b}})) - \mathbf{H}\underline{\mathbf{M}}\mathbf{d}_{\text{b}}^{\text{a}} - \mathbf{H}\underline{\mathbf{N}}\boldsymbol{\eta}) \end{aligned} \tag{4}$$

where $\mathbf{d}_{\text{b}}^{\text{a}}$ is the increment applied to the initial condition. From this we can derive an expression for the model forcing term as

$$\boldsymbol{\eta} = \mathbf{K}^{\text{q}}\mathbf{d}_{\text{b}}^{\text{o}} \tag{5}$$

where

$$\mathbf{K}^{\text{q}} = \mathbf{Q}\underline{\mathbf{N}}^{\text{T}}\mathbf{H}^{\text{T}}\left(\mathbf{H}\underline{\mathbf{M}}\mathbf{B}\underline{\mathbf{M}}^{\text{T}}\mathbf{H}^{\text{T}} + \mathbf{R} + \mathbf{H}\underline{\mathbf{N}}\mathbf{Q}\underline{\mathbf{N}}^{\text{T}}\mathbf{H}^{\text{T}}\right)^{-1} \tag{6}$$

and $\mathbf{d}_{\text{b}}^{\text{o}}$ is the innovation (the difference between the observations and the background trajectory).

2.1 Diagnosis using weak-constraint DA

Desroziers *et al.* (2005) introduced a method to diagnose the observation-, background- and analysis-error covariance matrices from data assimilation statistics. To extend this technique to model errors, we first need to calculate the covariance of the innovations. Following the assumption we made earlier we take the model errors to be constant during the DA window, but uncorrelated with background and observation errors. In this case the innovation covariance is

$$E((\mathbf{d}_{\text{b}}^{\text{o}})\,(\mathbf{d}_{\text{b}}^{\text{o}})^{\text{T}}) = \mathbf{R}^{\text{o}} + \mathbf{H}\underline{\mathbf{M}}\mathbf{B}^{\text{o}}\underline{\mathbf{M}}^{\text{T}}\mathbf{H}^{\text{T}} + \mathbf{H}\underline{\mathbf{N}}\mathbf{Q}^{\text{o}}\underline{\mathbf{N}}^{\text{T}}\mathbf{H}^{\text{T}} \tag{7}$$

where $\mathbf{Q}^{\text{o}}$ is the observed model-error covariance. For observations at the end of the window the last term is proportional to the number of time-steps squared, n^2, since each $\underline{\mathbf{N}}$ contains a summation of n terms.

Thus, the cross-covariance between the model forcing term and the innovation will be

$$\begin{aligned} E(\mathbf{H}\underline{\mathbf{N}}\boldsymbol{\eta}(\mathbf{d}_{\text{b}}^{\text{o}})^{\text{T}}) &= \mathbf{H}\underline{\mathbf{N}}\mathbf{K}^{\text{q}}E(\mathbf{d}_{\text{b}}^{\text{o}}(\mathbf{d}_{\text{b}}^{\text{o}})^{\text{T}}) \\ &= \mathbf{H}\underline{\mathbf{N}}\mathbf{Q}\underline{\mathbf{N}}^{\text{T}}\mathbf{H}^{\text{T}}\Delta\mathbf{K}^{\text{w}} \end{aligned} \tag{8}$$

where $\Delta\mathbf{K}^{\mathrm{w}}$ is given by

$$\Delta\mathbf{K}^{\mathrm{w}} = \begin{array}{l}\left(\mathbf{H}\underline{\mathbf{M}}\mathbf{B}\underline{\mathbf{M}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} + \mathbf{R} + \mathbf{H}\underline{\mathbf{N}}\mathbf{Q}\underline{\mathbf{N}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right)^{-1} \\ \left(\mathbf{H}\underline{\mathbf{M}}\mathbf{B}^{\mathrm{o}}\underline{\mathbf{M}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}^{\mathrm{o}} + \mathbf{H}\underline{\mathbf{N}}\mathbf{Q}^{\mathrm{o}}\underline{\mathbf{N}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right).\end{array} \tag{9}$$

To simplify the estimating procedure we use only observations from the first time in the data assimilation window, since the above expression will not then include the tangent linear model.

3 Experimental setup

To investigate the behaviour of the diagnostics tests were completed using the model of Lorenz (1995) which is based on the idea of waves propagating around a latitude circle. This circle is divided into 40 grid-points, and at each time step the grid-points are updated according to

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F \tag{10}$$

where the variables x_i, $i = 1, 2, ..., N$, are defined on a cyclic chain such that $x_{-1} = x_{N-1}$, $x_0 = x_N$ and $x_1 = x_{N+1}$. These experiments use a forcing term $F = 8$ which is within the chaotic regime. The Runge-Kutta 4^{th} order method was used to perform the time stepping, for intervals of $\delta t = 0.05$.

To create a model which is affected by model error, we follow an approach similar to that of Todling (2015). For each time-step in the truth run a random term is added to equation (10) of the following form

$$\delta\mathbf{r} = \mathbf{G}^{1/2}\delta\mathbf{p} \tag{11}$$

where $\mathbf{G}^{1/2}$ is the symmetric square-root of the covariance matrix $\mathbf{G}$. For the first half of the domain $\mathbf{G}$ takes values given by a Gaussian function of the distance between the points, using a length-scale of 5 grid-points. For the second half of the domain all the elements are zero, meaning that only the first half of the model is perturbed.

The data assimilation was run using weak-constraint 4DVar. This was given observations every time-step, and the data-assimilation window used observations from three times. Observations were produced by perturbing the truth run with errors sampled from $N(0, 0.1^2)$. By choosing small observation errors we ensure that the analysis errors are small, and the tangent-linear approximation used by 4DVar is valid.

4 Results

Figure 1 shows the estimates of the single-step model-error covariance matrix for the Lorenz '95 system. The initial input to the data assimilation (top-left) is an homogenous and nearly-diagonal covariance matrix. This is taken from the background-error covariance matrix estimated from an experiment using the Lorenz '95 model without model error and scaled to give reasonable results. As an approximation to the true model-error covariance matrix (top-right) it is quite poor.

The diagnostic estimate of the model-error covariance is shown in the bottom-left. The second half of the domain does not experience model error, and the estimated model error covariance is much reduced in this region. There is still an imprint of the initial model-error covariance in the estimated matrix,

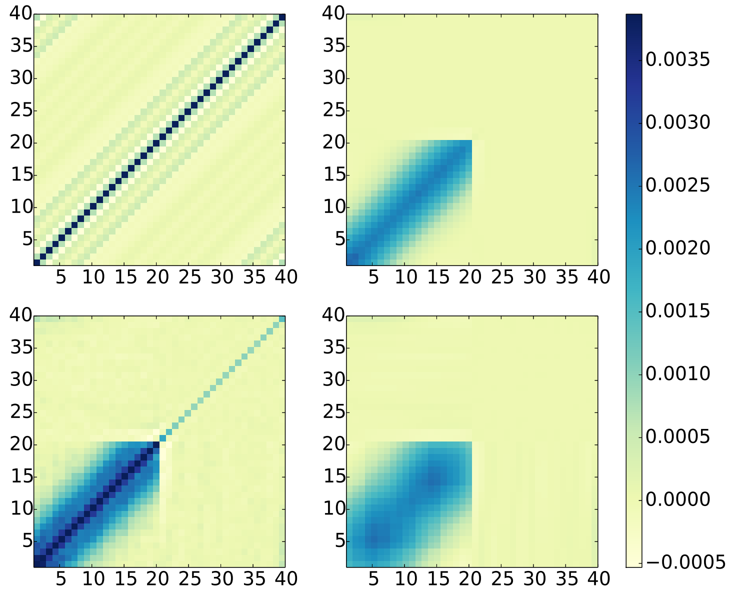


Figure 1: Single-step model error covariance matrix for the Lorenz '95 model. In these graphs the top-left graph shows the scaled input provided to the data assimilation, the top-right shows the true covariance matrix which is the target for the estimation. The bottom-left graph shows the estimate from the first run of the DA, and the bottom-right graph shows the estimate from the tenth run.

but the magnitude is much reduced. The diagonal elements in the first part of the domain are also reduced. However the off-diagonal elements are increased, reflecting the correlations in the true error covariance matrices. This is iterated by placing the diagnosed $\mathbf{B}$, $\mathbf{Q}$ and $\mathbf{R}$ matrices as input in the next run of the data assimilation. After 10 iterations the diagnosed $\mathbf{Q}$ matrix (bottom right) is very close to the true $\mathbf{Q}$ matrix (top right).

Acknowledgement

This work was developed following a series of discussions with Mike Cullen. The author would like to thank him for his very helpful input.

References

Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. *QJR Meteorol Soc* **131**: 3385–3396. doi: 10.1256/qj.05.108.

Lorenz EN. 1995. Predictability: a problem partly solved. In: *Proceedings of the seminar on predictability*, vol. I. ECMWF: Reading, Berkshire, UK, pp. 1–18.

Todling R. 2015. A lag-1 smoother approach to system-error estimation: sequential method. *QJR Meteorol Soc* **141**: 1502–1513. doi: 10.1002/qj.2460.

Trémolet Y. 2006. Accounting for an imperfect model in 4D-Var. *QJR Meteorol Soc* **132**: 2483–2504. doi: 10.1256/qj.05.224.

A comparison of the model error and state formulations of weak-constraint 4D-Var

Amos S. Lawless, Adam El-Said, Nancy K. Nichols

University of Reading, School of Mathematical and Physical Sciences

In this study we compare two formulations of the weak-constraint 4-dimensional variational data assimilation problem (wc4DVAR) - the ‘model error formulation’, in which the initial state and the model errors are estimated, and the ‘state formulation’, which estimates the model state throughout the assimilation window. The accuracy and efficiency with which the problems can be solved are determined by the condition numbers of the Hessians of the respective cost functions. Here we compare the sensitivities of the condition numbers and the convergence behaviour in the assimilation to changes in the input assimilation parameters.

We compare the two formulations applied to assimilation on a model of the linear advection equation. Using identical twin experiments we demonstrate the sensitivities of the condition number, convergence and solution of both algorithms to changes in the input data, such as observation accuracy, number of observations, correlation length-scales and assimilation window length. We show that both formulations are sensitive to these parameters. However, for some parameters the sensitivities can be quite different between the two formulations.

Using trajectories of ensemble analyses and tools from weak constraint 4DVAR to represent model uncertainty

Craig H. Bishop

Naval Research Laboratory, Monterey, California, USA

Weak constraint 4DVAR can be formulated to yield estimates of the stochastic component of model error. The accuracy of these estimates depends on the accuracy and number of assimilated observations, and the accuracy of the model error covariance matrix used in the weak constraint 4DVAR algorithm. To better understand and quantify these dependencies, weak constraint 4DVAR stochastic model error retrieval methods are applied to an idealized medium dimensional (240 variables) stochastic coupled model in which the stochastic forcing is precisely known. The experiments have indicated that even in this idealized system, for which all sources of uncertainty can be accurately quantified, an extremely high degree of observational coverage and accuracy is required to accurately recover individual stochastic noise realizations.

As a perhaps more promising alternative, experiments have been performed using the weak constraint 4DVAR apparatus to adjust trajectories of EnKF ensemble analyses to (a) enforce consistency between the model and observation space analyses and (b) estimate "balanced" model error realizations. The approach significantly improves the analyses in model space. The presentation will include results from attempts to quantify the extent to which stochastic perturbations based on an archive of such variationally derived model error realizations would be superior to perturbations based on the difference between forecasts and their verifying analyses.

Model error representation in convection-permitting forecasts and ensemble data assimilation

Glen Romine

National Center for Atmospheric Research (NCAR), Boulder, CO, USA

Regional ensemble prediction systems are typically deficient in error growth leading to under-dispersive predictions. There are a number of challenges within a regional modelling framework to improve forecast reliability, spanning initial and lateral boundary condition diversity along with a wide range of approaches to represent error in the forecast model. The effectiveness of these error growth sources depends on the forecast duration, domain size, and horizontal grid spacing among other factors. For continuously cycled ensemble data assimilation systems, insufficient error growth is often compensated through various inflation approaches. Meanwhile, with convection-permitting ensemble forecasts approaches include initial and lateral boundary condition diversity, multi-physics, and stochastic model error schemes. This talk will describe the framework behind NCAR's real-time convection-permitting ensemble prediction system along with outlining the current challenges and our ongoing activities in convection-permitting ensemble forecast system design.

Improving the Stochastically Perturbed Parametrisation Tendencies Scheme using High-Resolution Model Simulations

Hannah Christensen[1], Sarah-Jane Lock[2], Andrew Dawson[1] and Tim Palmer[1]

[1]Atmospheric, Oceanic and Planetary Physics, University of Oxford
[2]European Centre for Medium-Range Weather Forecasts

Stochastic parametrisations are used in weather forecasts as a physically motivated way of representing model uncertainty due to unresolved processes. In particular, the "Stochastically Perturbed Parametrisation Tendencies" (SPPT) scheme has become widely implemented in operational centres. SPPT is a holistic approach that represents uncertainty in all the sub-grid physics parametrisation schemes. It is easy to implement, and has beneficial impacts on medium range, seasonal and climate timescales. In the SPPT approach, the tendencies from the different parametrisation schemes are added together, and a spatially and temporally correlated random number is used to multiply this total tendency:

$$T = D + (1+e)\sum_{i=1} P_i$$

where T is the final total tendency, D the tendency from the dynamics scheme P_i, the tendency from the ith parametrisation scheme, and e, the spatially and temporally correlated zero-mean random number (Palmer et al, 2009).

However, despite the widespread use of the SPPT approach, little work has focused on providing a firm physical basis for the SPPT scheme. The scheme involves several assumptions. Firstly, the errors from different parametrisation schemes are assumed perfectly correlated. Secondly, the imposed spatial and temporal correlations are not based on observations, though the optimal magnitude of the noise has been justified.

The sensitivity of the scheme to the first of these assumptions can be tested. A generalised version of SPPT is developed whereby the individual parameterisation schemes are perturbed with an independent stochastic perturbation field. This 'independent SPPT' (iSPPT) approach assumes the errors from the different schemes are uncorrelated, and allows the user to set the noise magnitude and spatial and temporal characteristics separately for each parametrisation scheme. A 'partially independent' approach can also be used, whereby some schemes are perturbed with the same stochastic pattern, while others are perturbed independently.

A series of 21 member ensemble forecasts were performed in CY41R1 at T_L255 and in CY42R1 at $T_{CO}255$. Standard SPPT was compared to iSPPT and to a partially independent SPPT in which two patterns were used, one for the moist and one for the dry processes. iSPPT led to a significant improvement in ensemble forecast reliability in the tropics, increasing the spread, reducing the ensemble mean error, and improving the continuous ranked probability skill score across a range of variables. However, in the southern extratropics the iSPPT scheme leads to a slight increase of error. The two- pattern approach seems to be a good compromise, performing well in all regions across many variables. Both iSPPT and partially independent SPPT have the largest impact in regions with significant convective activity, correcting the underdispersive nature of the ensemble in those regions.

However it is likely that the true parametrisation errors are neither perfectly correlated, as in SPPT, nor uncorrelated, as in iSPPT. We propose the use of high-resolution model simulations to explicitly measure the difference between the parametrised and 'true' sub-grid tendencies: in this way, we characterise the error in the tendency that stochastic schemes such as SPPT seek to represent. This allows for a more systematic approach towards improving SPPT.

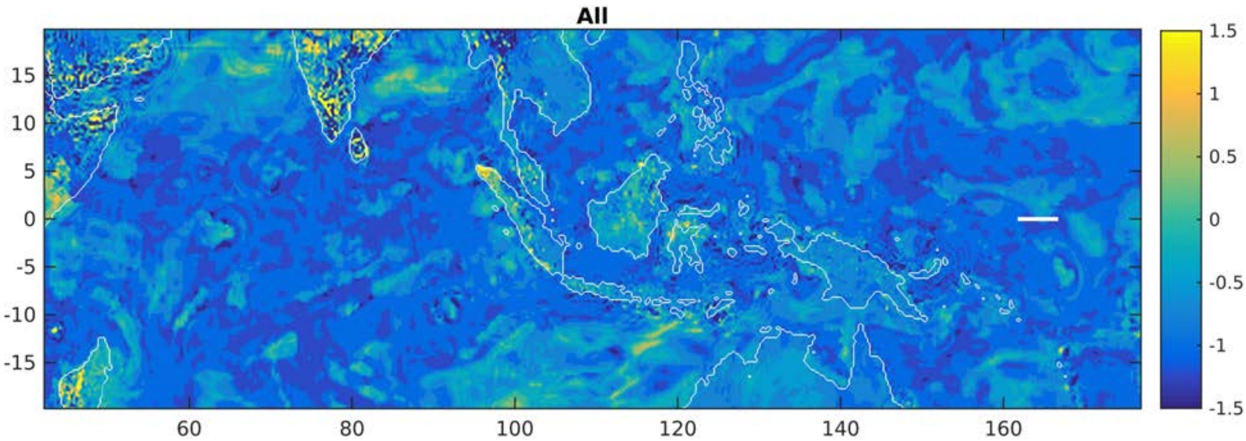
We use data from a high-resolution convection permitting integration with the UK Met Office limited area model as 'truth'. The model is run at 4km horizontal resolution with 70 levels in the vertical up to the model top at 40 km, and covers a large tropical domain (15,500 km x 4,500 km), focusing on regions of tropical convection in the Indian Ocean and West Pacific. This integration was performed as part of the CASCADE project (Holloway and Woolnough, 2013).

The high-resolution data is coarse-grained to the resolution of the ECMWF ensemble prediction system. This is used to provide initial conditions and forcing data for the IFS single column model (SCM). Short-range predictions with the SCM (15 minutes to one hour) are compared to the coarse-grained CASCADE data to derive the error statistics.

The SPPT equation can be rearranged to give:

$$T - D - \sum_{i=1} P_i = e \sum_{i=1} P_i$$

T is the 'true' total tendency from the CASCADE dataset, while D and P_i are the dynamics and physics tendencies from the SCM. All tendencies are a function of height at each location, while e, the optimal multiplicative perturbation, is a scalar. We can therefore solve this equation as a function of position and time to calculate a snap-shot of the optimal perturbation field, such as that shown below:



By repeating this calculation for all time steps within the CASCADE dataset, we can build up statistics of the optimal perturbation to be used in SPPT. We can also repeat the calculation allowing each parametrisation scheme to be perturbed independently, as in iSPPT. This allows us to investigate the error characteristics of each parametrisation scheme separately. It is hoped these measurements will improve both holistic and process based approaches to stochastic parametrisation.

References

Palmer, T. N., R. Buizza, F. Doblas-Reyes, et al, (2009) Stochastic parametrization and model uncertainty. Technical Report 598, ECMWF Shinfield park, Reading.

Holloway, C. E. and S. J. Woolnough, (2013). The effects of explicit versus parameterized convection on the MJO in a large-domain high-resolution tropical case study. Part I: Characterization of large-scale organization and propagation. J. Atmos. Sci., 70 (5), 1342–1369.

Impact of a stochastic deep convection scheme using cellular automata in the meso-scale ensemble prediction system; Harmon-EPS

Lisa Bengtsson

SMHI

There is a long standing discussion in the Numerical Weather Prediction (NWP) community whether model error (arising from model physics) should be represented in form of stochastic physics applied a posteriori after the tendencies from the physical parameterizations have been computed, or whether it should be included at the source of uncertainty, at the sub-grid, in form parameter perturbations or a stochastic parameterization (see for instance various methods discussed in Berner et al. 2016). Although the latter sounds appealing, since one could really address the model error in a more physically motivated manner, the former has proven more successful in many ensemble prediction systems, using the stochastically perturbed parameterization tendency (SPPT) scheme (e.g Buizza et al., 1999, Palmer et al., 2009), and the stochastic kinetic energy backscatter scheme (SKEBS) (Shutts, 2005). One reason for the success of these schemes is the fact that the scheme are correlated in space and time over scales far beyond the sub-grid scale, whereas sub-grid representation of model error, such as parameter perturbations or white noise stochastic parameterizations, are limited to the "column physics", and thus there are no spatial and temporal correlations of the perturbations.

One way to include both spatial correlations, and temporal memory to a stochastic parameterization included on the sub-grid, was to couple a cellular automaton to the deep convection parameterization of the NWP model, ALARO (Bengtsson et al. 2013). The cellular automaton is in this scheme acting on the sub-grid of the NWP model and can form clusters which can organize themselves across the NWP model's grid-boxes to enhance deep convective organization. This way the cellular automata can form a "fraction of the NWP grid-box" that are later coupled back to the model via the updraft mesh-fraction in the closure assumptions of the deep convection parameterization. The cellular automata is only coupled back to the deep convective evolution of the updraft mesh-fraction if the value of CAPE exceeds a certain threshold.

The aim of the study presented at this workshop was to understand whether such a stochastic parametrization on the sub-grid scale could be used in order to represent the uncertainty associated with deep convection, and provide more reliable probabilistic forecasts in an ensemble prediction system for the mesoscale. Describing the statistical effect that deep convection has on the large- scale flow in a stochastic manner means that the resolved scale variables, such as the vertical and horizontal wind or temperature, would respond differently to the convection scheme each time the model was run. Thus, the stochasticity of the implemented parametrization can be studied by examining the ensemble spread in resolved model variables generated from the sub-grid scheme. The expectation is to capture better the range of possible convective responses given by the ensemble members in the resolved fields of the model.

The study has been published in Bengtsson and Körnich, 2016. The ensemble prediction system used is called Harmon-EPS, it is an ensemble prediction system aimed at the mesoscale and is based on the Hirlam Aladin Regional/Mesoscale Operational NWP In Europe (HARMONIE) forecast system, which is developed within the two HIgh Resolution Limited Area Model (HIRLAM)–Aire Limit´ee Adaptation dynamique D´eveloppement InterNational (ALADIN) consortia, a collaboration on NWP development between 26 countries in Europe.

It was found that the stochastic cumulus parametrization does in general give an increased amount of convective precipitation in regions where CAPE is large. This resulted in a larger ensemble spread of convective precipitation. However, as a consequence of the parametrization, resolved precipitation was reduced due to an increase in sub-grid precipitation, giving a decreased ensemble spread over the domain average for total precipitation. Such reduction in resolved precipitation also led to an improvement in precipitation bias for the test-period, which in general improved the skill. Thus, from the classical view of introducing stochastic physics in order to increase the ensemble spread, doing so within the physical sub-grid parametrization is not straightforward. Various feedbacks within the physical parameterization even lead to a reduced spread in some instances where the skill was improved.

Overall, for 6 h accumulated precipitation, the BSS, CRPS and mean bias were for the most part improved (with the exception of small thresholds for precipitation in the BSS), which suggests that the ensemble forecast was improved, however, not by reason of increasing the spread but instead by improving the skill, due to the inclusion of the dependence of CAPE in the closure and the memory and lateral communication of the cellular automata.

In the future the scheme's impact will be tested on equatorially coupled waves over the Tropics in order to understand further the potential for "upstream" representation of model error when the coupling between the deep convective organization and the atmospheric flow is large.

References:

Bengtsson, L., Steinheimer, M., Bechtold, P. and Geleyn, J.-F. (2013), A stochastic parametrization for deep convection using cellular automata. Q.J.R. Meteorol. Soc., 139:1533–1543. doi:10.1002/qj.2108

Bengtsson, L. and Körnich, H. (2016), Impact of a stochastic parametrization of cumulus convection, using cellular automata, in a mesoscale ensemble prediction system. Q.J.R. Meteorol. Soc., 142:1150–1159. doi:10.1002/qj.2720

Berner et al. 2016: "Stochastic Parameterization: Towards a new view of Weather and Climate Models", Bulletin of the American Meteorological Society. Under review.

Buizza, Roberto, M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system."QJR Meteorol. Soc, 125, 2887-2908.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic Parametrization and Model Uncertainty. ECMWF Technical Memorandum, 598, available at http://www.ecmwf.int/publications/.

Shutts, G, 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems." Quarterly Journal of the Royal Meteorological Society (2005): 3079-3102

Parameter uncertainty of chaotic systems

Heikki Haario[1]

Lappeenranta University of Technology, Lappeenranta, and Finnish Meteorological Institute, Helsinki, Finland

(Dated: 30 May 2016)

Many physical processes, such as boundary layer turbulence or cloud microphysics, are represented in numerical weather prediction (NWP) models by physical parameterization schemes. These schemes contain closure parameters to express some unresolved variables by predefined parameters rather than by explicit modelling. The increasing complexity of NWP models makes it very demanding to optimally specify parameters values by manual techniques using limited samples of test forecasts. This may partly explain the increasing difficulties encountered in integration of new physical parametrizations schemes into model dynamics. Development of algorithmic tools to make statistical inference about the closure parameters would be helpful to facilitate and speed-up NWP model development. The same question of parameter optimization and uncertainty quantification is equally crucial for climate models. Additionally, the situation is complicated here by the challenge of unpredictability: due to chaoticity there is no unique solution for the long time model integrations, even with fixed model parameters, see, e.g., Ref.[1] for more discussion. Here we present recent methods for parameter estimation of chaotic systems, both for short-time and long-time situations.

Several approaches have been proposed for joint estimation of static parameters and dynamic state variables. It is relatively straightforward to augment the state vector in filtering applications with the static model parameters and treat them as artificial model states. A drawback is that parameter values tend to change from one filtering step to the next, in accordance with the changing atmosphere and observing network, although they are static or quasi-static. Moreover, filtering requires additional tuning parameters which may lead to bias for the model parameters Ref.[2]. Another way of employing the filtering approach is to construct a filter-based likelihood to be optimized with respect to the parameters Ref.[3]. For large systems such iterative optimization is prohibitively CPU demanding, however. For the same reason other approaches such as particle filtering are excluded in state estimation of large systems.

The idea of the EPPES concept Ref.[4], Ref.[5] is to create a 'CPU-free' NWP model parameter estimation by slightly modifying an existing EPS system: an operational ensemble prediction system is added with a functionality to perturb model parameters and to learn which ones tend to perform well. So the massive amount of model simulations of ensemble prediction would be utilized for on-line model optimization, practically without any additional CPU demand. The original EPPES concept is based on the steps, repeated for each assimilation window, of (i) sampling candidate parameter values from a Gaussian proposal distribution, (ii) launching each ensemble member of the prediction model with different candidate parameter values, (iii) evaluating the performance of the parameters against a cost function, and (iv) adapting the proposal distribution according to the parameter performance. The adaptation is done in a Monte Carlo way, based on the importance weights of the cost function values.

The approach was successfully applied to improve the performance of the already highly tuned IFS system in Ref.[6], Ref.[7]. The selection of the cost function, however, reveals a problem: while the performance of the model can be improved according to the criteria selected as part of the cost function, some other aspects of the prediction may deteriorate. This calls for a multicriteria optimization approach, where no relevant part of the model performance is allowed to converge towards unacceptable values. In Ref.[8] we apply an evolutionary optimization approach, the Differential Evolution (DE), for this purpose. Each assimilation window may be interpreted as a generation and the ensemble as the respective population. With slight modifications (due to the stochastic nature of the cost functions) the DE algorithm may then be employed to optimize the model parameters. The special requirements of various optimization criteria may be taken into account by, e.g., the desirability function method. Otherwise the implementation is similar to that with EPPES, in the sense that an existing EPS system is used, without any essential new CPU demand. As an optimization approach the DE version typically gives faster convergence, while the sampling-type original EPPES algorithm provides an uncertainty quantification for the parameter identification.

The closure parameters of a large scale climate model, ECHAM5, were studied in Ref.[9] using several summary statistics, such as temporal and spatial averages of the key balance factors of the climate, as the cost function. While parameter estimation was technically possible to perform, see Ref.[10] for the methods, the results remained inconclusive. The reason was the difficulty of selecting the cost function terms that would be sensitive enough with respect to the closure parameters. The standard way of estimating parameters of dynamical systems is based on the residuals between the data and the model responses, both given at the time points of the measurements. Supposing the statistics of the measurement error is known, a well defined likelihood function can be written. The maximum likelihood point is typically considered as the best point estimator, and it coincides with the usual least squares fit in the case of Gaussian noise. The full posterior distribution of parameters can be sampled by Markov chain Monte Carlo (MCMC) methods. The approach has become routine for the parameter estimation of deterministic models in Bayesian inference, see Refs.[11,13] for further references. The estimation of the parameters of chaotic models can not be performed in this way. After an initial time period where the system is predictable, the model responses, even with just slightly varying initial condition or some settings of a numerical solver employed, diverge so that the concept of a given model response at a given time point loses the meaning. The same effect can be seen when some infinitesimal changes to the model parameters are made. In this sense, there is no unique model trajectory corresponding to a fixed model parameter vector. But while all such trajectories are different, they approximate the same underlying attractor and should be considered in this sense equivalent. Here we discuss a statistical approach presented in Ref.[14] to quantify such "sameness" of trajectories, and to distinguish trajectories that are significantly different. The basic idea is to create a summary statistics that takes into account the geometry of the attractor, rather than using direct averages or other (linear) projections such as used in Ref.[9]. Various formulations of fractal dimensions have been developed to characterise the internal geometry of such attractors. Here we modify one of these, the so-called correlation dimension[15], to develop a way to quantify the variability of samples of an attractor by mapping the respective phase space trajectories onto vectors, whose statistical distribution can be empirically estimated. The distributions turn out to be Gaussian, which provides us a well defined statistical tool to compare the trajectories. We use the approach for the task of parameter estimation of chaotic systems. Other applications are pointed out as well.

[1] J. Rougier, Intractable and unsolved: some thoughts on statistical data assimilation with uncertain static parameters. Phil. Trans. R. Soc. A 371, 20120297 (2013).

[2] J. Hakkarainen, A. Solonen, A. Ilin, J. Susiluoto, M. Laine, H. Haario, and H. Järvinen, Tellus A 65, 20147 (2013).

[3] J. Hakkarainen, A. Ilin, A. Solonen, M. Laine, H. Haario, J. Tamminen, E. Oja, and H. Järvinen, Nonlin. Processes Geophys. 19, 127 (2012).

[4] H. Järvinen, M. Laine, A. Solonen, and H. Haario, Q. J. R., Meteorol. Soc. 138, 281 (2012).

[5] M. Laine, A. Solonen, H. Haario, and H. Järvinen, Q. J. R. Meteorol. Soc. 138, 289 (2012).

[6] P. Ollinaho, P. Bechtold, M. Leutbecher, M. Laine, A. Solonen, H. Haario, and H. Järvinen, Nonlin. Processes Geophys. 20, 1001 (2013).

[7] Ollinaho, P., Järvinen, H., Bauer, P., Laine, M., Bechtold, P., Susiluoto, J., and Haario, H.: *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, 2014.

[8] Vladimir Shemyakin, Heikki Haario: *Optimizing Ensemble Prediction System parameters with Differential Evolution*. Submitted.

[9] Järvinen, H., Räisänen, P., Laine, M., Tamminen, J., Ilin, A., Oja, E., Solonen, A., and Haario, H.: *Estimation of ECHAM5 climate model closure parameters with adaptive MCMC*, Atmos. Chem. Phys., Vol. 10, nro. 2, 9993-10002, 2010.

[10] A. Solonen, P. Ollinaho, M. Laine, H. Haario, J. Tamminen, and H. Järvinen, Bayesian Analysis 7, 1 (2012).

[11] C. Robert and G. Casella, Monte Carlo Statistical Methods, Springer, 2nd ed. (2004).

[12] H. Haario, E. Saksman, and J. Tamminen, Bernoulli 7, 223 (2001).

[13] H., Haario, M. Laine, A. Mira, and E. Saksman, Stat. Comput. 16, 339 (2006).

[14] Heikki Haario, Leonid Kalachev, Janne Hakkarainen *Generalized Correlation integral vectors: A distance concept for chaotic dynamical systems.* Chaos, 25, 2015

[15] P. Grassberger and I. Procaccia, Phys. Rev. A. 28, 2591 (1983).

Inducing Tropical Cyclones to Undergo Brownian Motion

Daniel Hodyss

Naval Research Laboratory, Monterey, California, USA

Recent work in ensemble forecasting has focused on validating the impact of general forms of stochastic forcing on TC forecasts (e.g. Snyder et al. 2011, Lang et al. 2012) and has shown some beneficial impact. We aim here to specifically address the basic issues of under-dispersiveness and biases in ensemble-based TC track distributions through a stochastic parameterization that induces TCs to undergo Brownian motion. Because a characteristic of Brownian motion is an increasing ensemble position variance with time this allows for the inflation of forecasted distributions by user-defined amounts. The proper application of a stochastic parameterization however requires a choice of stochastic calculus. There exist two standard stochastic calculi that are commonly studied in the theory of stochastic differential equations (Kloeden and Platten 1991). The first is that of Itô (1951) and the second is that of Stratonovich (1966). The most important point about the choice of stochastic calculus is that each one implies a distinctly different algorithm is required to obtain a particular result. The algorithmic differences implied by the choice of stochastic calculus and their impact upon the structure and life cycle of TCs are the subject of this talk. In the course of this talk we will show that the naïve implementation of a stochastic parameterization without properly accounting for the appropriate stochastic calculus will lead to undesirable results. In the cases presented here these undesirable results will manifest as overly intense TCs, which, depending on the strength of the forcing, could lead to problems with numerical stability and physical realism.

References

Itô, K., 1951: On stochastic differential equations. Mem. Amer. Math. Soc. **4**, 1-51.

Kloeden, P.E., Platen, E., 1999: Numerical solution of stochastic differential equations. *Springer-Verlag*, Berlin, Germany, 636 pgs.

Lang, S.T.K., M. Leutbecher, and S.C. Jones, 2012: Impact of perturbation methods in the ECMWF ensemble prediction system on tropical cyclone forecasts. *Q.J.R. Meteorol. Soc*., **138**, 2030-2046.

Snyder, A., Z. Pu, and C.A. Reynolds, 2011: Impact of stochastic convection on ensemble forecasts of tropical cyclone development. *Mon. Wea. Rev.,* **139**, 620-626.

Stratonovich, R.L., 1966: A new representation for stochastic integrals and equations. *J SIAM Control*, **4**, 362-371.

Observational-based Stochastic Convection Parameterization

Jesse Dorrestijn (CWI),
Pier Siebesma (KNMI, TU-Delft), Daan Crommelin (CWI, UvA),
Harmen Jonker (TU-Delft), Frank Selten (KNMI)

July 8, 2016

Inadequate representations of convection and clouds in weather and climate models by parameterizations are sources of model error and lead to uncertainties in model predictions. In this presentation, an approach for the construction of stochastic convection parameterizations has been presented. Results are shown for convection parameterizations that are constructed by using observational data.

Firstly, motivations for the usage of stochastic convection parameterization in weather and climate models are given: GCM model resolutions tend to increase and are getting close to the Grey Zone. Furthermore, the increased variability in convective response in a smaller model column can be captured by using *stochastic* rather than the more classically applied *deterministic* parameterizations.

Then, the building blocks of our models are presented and explained: data-driven Conditional Markov Chains (Crommelin and Vanden-Eijnden (2008)) and stochastic multicloud models (Khouider et al. (2010)). A choice can be made between the source of data that is used to infer the Markov chains: high-resolution observations or high-resolution model data. The size of the area covered by the data is shown to be larger for the observations, which is an advantage for the usage of observations.

The construction of the Markov chain multicloud model from observational data is explained in detail. It is shown how data of a rain radar in Darwin in Australia can be divided into classes such that only a few states are possible for each pixel of the radar domain. The time series that are obtained can be used to infer transition probabilities of a Markov chain that switches between states every 10 minutes. After construction of the multicloud model, it produces convective area fractions that can serve as a closure for the mass flux at cloud base in the convective scheme of a GCM.

Furthermore, with cross-correlation analysis, it is shown that for the observations, the large-scale vertical velocity ω correlates strongly with convection. The relationship between ω and the convective area fraction σ is explained in a figure.

Results from the multicloud model are shown. Convective area fractions produced by the multicloud model are compared to observations in Darwin: statistics are well comparable. Then, it is shown that the multicloud model can be adapted to the size of the model column for which it has to produce convective area fractions. For larger model columns, it is able to produce fractions with smaller fluctuations.

Results are shown for the implementation of several convection parameterizations in a GCM of intermediate complexity: SPEEDY (Dorrestijn et al. (2016)). Time series of the mass flux at cloud base in one grid column are shown and their statistics are shown for the entire tropical belt. A Markov-chain-based scheme similar to the scheme of Gottwald et al. (2016), is shown to have a remarkably well similarity with the observations. We present autocorrelation functions of the mass flux at cloud base and probability density functions of the daily accumulated precipitation produced by SPEEDY with several stochastic schemes and compare to control. Also Hovmöller-diagrams and Wheeler-Kiladis diagrams are presented for the precipitation in the tropics. The stochastic schemes are shown to have a large impact on the precipitation. A novel approach for the assessment of the power of the convectively coupled equatorial waves has been presented.

In the conclusion section, it is concluded that the stochastic models are able to capture variability related to convection; that observations are more useful than high-resolution model data (at the moment); that realistic time-series of the mass flux at cloud base can be produced in SPEEDY when the convective schemes are conditioned on ω; that the average strength of the mass flux at cloud base affects the simulation of the MJO and equatorial Kelvin waves; and finally that the skill of these equatorial waves can be expressed in a single scaler, which enables modelers to tune models.

- D. Crommelin and E. Vanden-Eijnden. Subgrid-scale parameterization with conditional markov chains. J. Atmos. Sci., 65:2661–2675, 2008.
- J. Dorrestijn, D.T. Crommelin, A.P. Siebesma, H.J.J. Jonker, and F. Selten. Stochastic convection parameterization with Markov chains in an intermediate complexity GCM. J. Atmos. Sci., 73:1367–1382, 2016.
- G.A. Gottwald, K. Peters, and L. Davies. A data-driven method for the stochastic parametrisation of subgrid-scale tropical convective area fraction. Quarterly Journal of the Royal Meteorological Society, 142(694):349–359, 2016.
- B. Khouider, J. Biello, and A.J. Majda. A stochastic multicloud model for tropical convection. Comm. Math. Sci., 8:187-216, 2010.