Uncertainty Prediction Across a Range of Scales: From short-range weather forecasting to climate uncertainty

Judith Berner[⋆]

[⋆] *National Center for Atmospheric Research*
Boulder, CO, USA
berner@ucar.edu

ABSTRACT

A stochastic kinetic-energy backscatter scheme and other model-refinements are employed to represent model uncertainty and model error across a range of models targeted at different temporal and spatial scales. In a mesoscale ensemble prediction system using the Weather Research and Forecasting model (WRF) including a uncertainty representation scheme leads to significant improvements in probabilistic skill. When used for ensemble data assimilation the inclusion of model-error schemes decreases the the root-mean squared innovations and analysis increments indicating a better match between model first-guess and observations. When included in the ECMWF model run at horizontal resolutions used in climate models and on seasonal timescales, the stochastic kinetic-energy backscatter scheme results in a reduction of systematic model bias and a better representation of certain aspects of tropical tropical climate variability. Some of the improvements can also be obtained by increasing horizontal resolution or improving the deterministic parameterizations, pointing to the importance of unresolved scales and the need to include them in some manner.

1 Introduction

The central concern of numerical weather prediction is to predict meso- and synoptic scales of atmospheric motion as accurately as possible. An objective way to estimate not only the most likely weather evolution, but also its uncertainty, is to run an ensemble prediction system, which accounts for initial-condition error and provides a probabilistic forecast of the atmospheric evolution. On these scales model error manifests itself in that ensemble forecasts – even with elaborate initial-condition uncertainty representations – tend to be underdispersive and underestimate the true uncertainty of the atmospheric evolution (Buizza et al., 2005). This leads to unreliable and overconfident probabilistic forecasts, and in particular to a poor representation of large anomalies such as extreme weather events.

On seasonal to climatic timescales the trajectories of ensemble systems continue to not diverge fast enough, leading to even more underdispersive ensemble systems (e.g., Berner et al., 2008, Doblas-Reyes et al., 2009). However, another component of model error becomes more evident, namely the emergence of biases, i.e., differences between the mean climate in the model and that in nature. There is increasing evidence that many climate biases emerge during the the first steps of an integration (e.g. Hannay et al., 2009) and can be analyzed by running climate models in "weather mode" (Klinker and Sardeshmukh, 1997, Rodwell and Palmer 2007).

Both weather and climate models can be characterized in terms of a) their dynamical cores and b) their physical parameterizations. The dynamical core describes the truncation of the underlying – essentially infinite-dimensional– differential equations onto a finite grid, and the corresponding algorithms that evolve the resolved scales of motion on this grid. Complementary to this, the parameterizations are traditionally deterministic formulae that provide estimates of the gridscale effect of processes (such as deep convection) which cannot be resolved by the dynamical core.

These formulae necessarily involve a number of free parameters which in principle should be determined by observations, but in practice may not be. This may be partially because of insufficient observations, but more typically is because the formulae themselves are not accurate bulk representations of the sub-grid processes. This necessitates a certain amount of model "tuning": in the case of climate, for example, by fixing parameters though the fit of the climate model as a whole to the globally averaged net outgoing longwave radiation.

In recent years, the development of methods to estimate flow-dependent uncertainty in weather and climate model predictions in addition to initial condition uncertainty has become an important addendum. Different techniques have been developed, the most commonplace being multi-model ensembles (e.g Houtekamer et al., 1996, Krishnamurti et al., 2000; Hagedorn et al., 2005; Solomon et al., 2007), perturbed parameter (e.g., Stainforth et al., 2004) and multi-physics approaches (Murphy, 2004, Berner et al. 2011).

However, a third technique has emerged. Specifically, the basis of stochastic-dynamic parameterization lies in the treatment of tendencies due to unresolved processes by stochastic rather than bulk-formula deterministic representations (Epstein and Pitcher, 1972, Pitcher 1977, Palmer 2001). In this approach, the effect of uncertainties due to the finite truncation of the model can be estimated by running ensembles of integrations, whose members are based on independent realizations of a stochastic processes that describe these truncation uncertainties. One advantage of stochastically perturbed models is that all ensemble members have the same climatology and model bias in contrast to multi-parameter, multi-parameterization and multi-model ensembles in which each ensemble member is *de facto* a different model with its own dynamical attractor.

One such stochastic parameterization is the stochastic kinetic energy-backscatter scheme (SKEBS) whose origin lies in Large-Eddy Simulation modeling (Mason and Thompson, 1992) and has recently been extended to weather and climate scales (Shutts and Palmer, 2003, Shutts, 2005). The key idea is that energy associated with sub grid processes, instead of being dissipated within the gridbox, is injected back onto the grid using a stochastic pattern generator. This method has been successfully used for operational and research forecasts across a range of scales from the meso-scale (Berner et al 2011), over synoptic scales (Berner et al., 2009, Charron et al. 2010, Tennant et al., 2011) to seasonal scales (Berner et al., 2008, Doblas-Reyes et al., 2009). At the point of writing this report, studies to look at the effect of stochastic parameterizations on climatic scales are under way, but not yet published (e.g. Doblas-Reyes, Berner, pers. Comm.).

2 Short-range ensemble forecasts

Here we present results for the short-range weather forecasts in a mesoscale ensemble prediction system based on the Weather and Research forecast model (WRF, Skamarock et al., 2008). The horizontal resolution is 45km and the domain is limited to the area of the continuous United States. This work compares the results from four ensemble systems: the control ensemble with a unique set of physics (CNTL), a multi-physics ensemble using different physics packages for each ensemble member (PHYS), an ensemble with control physics and stochastic kinetic energy-backscatter scheme (STOCH) and finally an ensemble system combining stochastic and multi-physics perturbations (PHYS_STOCH). More details of this work can be found in Berner et al., 2011a and Hacker et al., 2011.

2.1 Spread-error consistency

A measure of reliability of an ensemble system is the degree of consistency between root-mean-square (RMS) ensemble-mean error and spread. The spread around the ensemble mean and the RMS error of

the ensemble mean are computed for horizontal wind and temperature from all ensemble systems as a function of pressure level at 12-h and 60-h forecast lead times, and averaged over the entire domain. Prior to the computation of spread and error, the monthly mean bias for each ensemble member was removed. For a perfectly reliable ensemble system the flow-dependent initial uncertainty should be fully represented by the total ensemble spread and thus spread and RMS error should grow at the same rate so that the uncertainty of the forecast is well represented by the ensemble spread. Here, verifications of zonal wind u against observed soundings are presented.

On average, all ensemble systems in this study are underdispersive i.e., the error exceeds the ensemble spread for all forecast lead times (Fig. 1). It needs to be stressed that the ensemble systems appears more underdispersive than they are, because the observation error was not included in the spread computation. However, Berner et al. 2011 confirm that the ensemble systems continue to be underdispersive, if observation error is accounted for. In the spatial and temporal average of the u-wind component, the error and spread curves have a similar vertical structure with smallest values near the surface, a local maximum near 95 kPa and vertically increasing values above 75 kPa (Fig. 1). Comparing the spatially and temporally averaged error and spread curves confirms that the under-dispersiveness in the PHYS and STOCH ensembles has greatly improved from CNTL, while the RMS error remains almost the same or is even reduced (Fig. 1).

Overall, STOCH is best at generating the uncertainties in the horizontal winds for the entire atmosphere while PHYS is best at representing the uncertainties in temperature at the surface and in the planetary boundary layer. The most dispersive ensemble system is PHYS_STOCH, characterized by both perturbations from multiple physics schemes and stochastic perturbations. We note that the combination of both model-error schemes increases the spread in most levels, but not in an additive manner. When comparing PHYS_STOCH to PHYS, we see that the former has considerably more spread, but the RMS error of the ensemble mean is hardly different.

2.2 Brier-score profiles

Verifying the spread-error consistency gives us an estimate of the predictive accuracy of the ensemble-mean forecast. However, the true value of an ensemble system lies in predicting the variability of the potential outcomes. Latter is best assessed by a probabilistic verification. Therefore we perform such a probabilistic verification using the most commonly used score, the Brier score (e.g., Wilks, 1995). The Brier score was categorized in four different verification events signifying positive or negative "common anomalies" and "extreme events", where "common anomalies" are defined as an anomaly of less than one standard deviation from the climatological mean, and "extreme events" as an anomaly of more than one standard deviation. At each isobaric spherical coordinate $r = (\lambda, \phi, p)$, we compute the climatological standard deviation, $\sigma_x(r)$ of a variable $x \in \{u, T\}$ with regard to their respective monthly mean, and take the weighted average over the number of months in the verification period. Then Brier score profiles as function of pressure level are computed for the following events: $x(r) < -\sigma_x(r)$, $-\sigma_x(r) < x(r) < 0$, $0 < x(r) < \sigma_x(r)$ and $\sigma_x(r) < x$. Since the results for negative and positive "common anomalies" are very similar, we only show the scores for the first, third and fourth event. The Brier score is computed after removing the monthly mean bias from each ensemble member.

The largest (i.e., worst) Brier scores are found at the surface, and generally the probabilistic forecast skill improves with height (Fig. 2 a-c). The differences between CNTL and the ensembles with model-error schemes (panels d-f) are generally small compared to the height dependence of the Brier score, but nevertheless tend to be statistically significant at the 95% confidence level (denoted by filled and empty markers) in most levels. The sign is defined so that positive differences signify an improvement over CNTL and negative differences a deterioration. The verification confirms that the ensemble systems with model-error representation PHYS, STOCH and their combination PHYS_STOCH perform without exception better than the control ensemble. For u, STOCH performs better than PHYS, especially aloft,

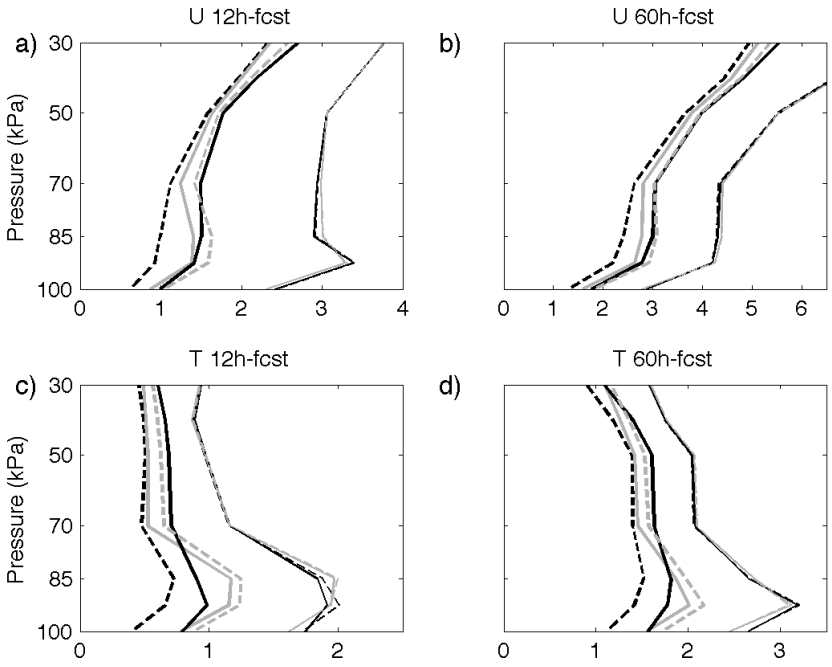

Figure 1: *Spread around ensemble mean (thick curves) and root-mean-square (RMS) error of ensemble mean (thin curves) relative to observations for u in m s^{-1} (a,b) and T in K (c,d). Spread and error curves are shown for four ensemble systems: control physics CNTL (black dashed), multi-physics PHYS (gray solid), stochastic backscatter STOCH (black solid), and multi-physics combined with stochastic backscatter PHYS_STOCH (gray dashed). The ensemble systems are debiased with regard to their respective mean monthly bias. Forecast lead time is 12 h (a,c) or 60 h (b,d). (From Berner et al., 2011a)*

but PHYS and especially PHYS_STOCH are significantly better in the boundary layer and at the surface.

3 Use of uncertainty representations for ensemble data assimilation

Recently data from observations networks have been used increasingly not only to come up with the best estimate for the initial state, but to detect systematic model biases, which often manifest themselves already in the first timesteps. The basic idea is that if there is a systematic bias in the analysis increments averaged over many different synoptic situations, and if the observations have been de-biased, then this difference is due to systematic model error (e.g., Klinker and Sardeshmukh, 1997, Rodwell and Palmer, 2005, Hannay et al., 2009). The data assimilation research testbed (DART, Anderson et al., 2009) uses the ensemble Kalman Filter to assimilate observations into short-term forecasts produced by the Weather and Research Forecast (WRF) model. Since the short-term forecasts started from cycled analyses (the "prior") are underdispersive due to over-constraining certain geographical regions, an inflation algorithm is introduced to account for this "sampling error' (Anderson, 2007). An ensemble of analyses is obtained by updating the priors with the observations.

Preliminary results of ensemble mean analyses and priors using the ensemble systems described in the last section are shown in Fig.3 for the period 06/01/2008-06/11/2008 (Ha and Snyder, 2011, Snyder at al., 2011). The zig-zag pattern is the result of showing the prior (peaks in spread/valleys in innovations curves) and posterior (valleys in spread/peaks innovations curves) in the same figure. We find that the RMS innovations for the ensemble with stochastic parameterizations have the smallest RMS value, indicating the closest match between observations and short-term forecasts from the model, as well as

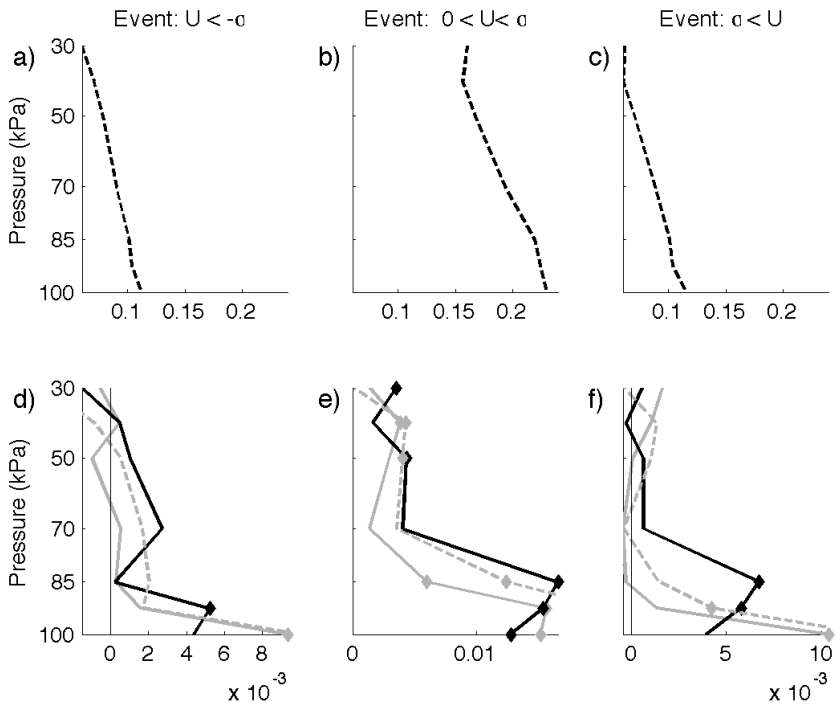

Figure 2: *Brier score of ensemble system with control physics CNTL (black dashed) relative to observations, for zonal wind u, as function of pressure, for three verification events: a) $u(r) < -\sigma_u(r)$, b) $0 < u(r) < \sigma_u(r)$ and c) $\sigma_u(r) < u$; where $\sigma_u(r)$ is the climatological standard deviation of u as function of longitude $\lambda = r_1$, latitude $\phi = r_2$ and pressure $p = r_3$. The score for the event $-\sigma_u(r) < u(r) < 0$ is very similar to that in (b) and hence not shown. A smaller Brier score denotes better skill. d-f) Brier Score differences of PHYS (gray solid), STOCH (black solid) and PHYS_STOCH (gray dashed) from CNTL for the same three events. The sign is defined so that positive differences signify an improvement over CNTL and negative differences a deterioration. Filled (empty) markers denote statistically significant improvement (deterioration) with regard to the skill of CNTL at the 95% confidence level. The forecast lead time is 60 h. (From Berner et al., 2011a)*

the closest agreement between the ensemble mean analysis and observations. The multi-physics scheme also produces a reduction in RMS innovations and additionally is characterized by an increase in spread, but the improvement is not as marked as with the stochastic perturbations. Overall these initial results are very promising and suggest that the inclusion of a model-error representation in the context of ensemble data assimilation can lead to an improved analysis system.

4 Model error on seasonal to climatic time scales

Next we are investigating the impact of model-error schemes on seasonal to climatic timescales. The impact on systematic model error is studied by comparing seasonal runs of ECMWF's Integrated Forecasting System (IFS), version CY31R1 to renanalysis and analyses data. Here we present results based on 15 years of Northern Hemispheric winter forecasts from 1990-2005. The horizontal resolution was chosen to be T95 which is a resolution typical for climate integrations. Three different kinds of subgrid-scale refinements were studied: an increase in horizontal resolution to T599 (HIGHRES), the inclusion of a stochastic kinetic-energy backscatter scheme (STOCH) and advancements to the deterministic parameterizations, namely the ones used in model version CY36R1. For further details on the experiment setup and a results beyond the ones presented here, we refer to Berner et al., 2011b.

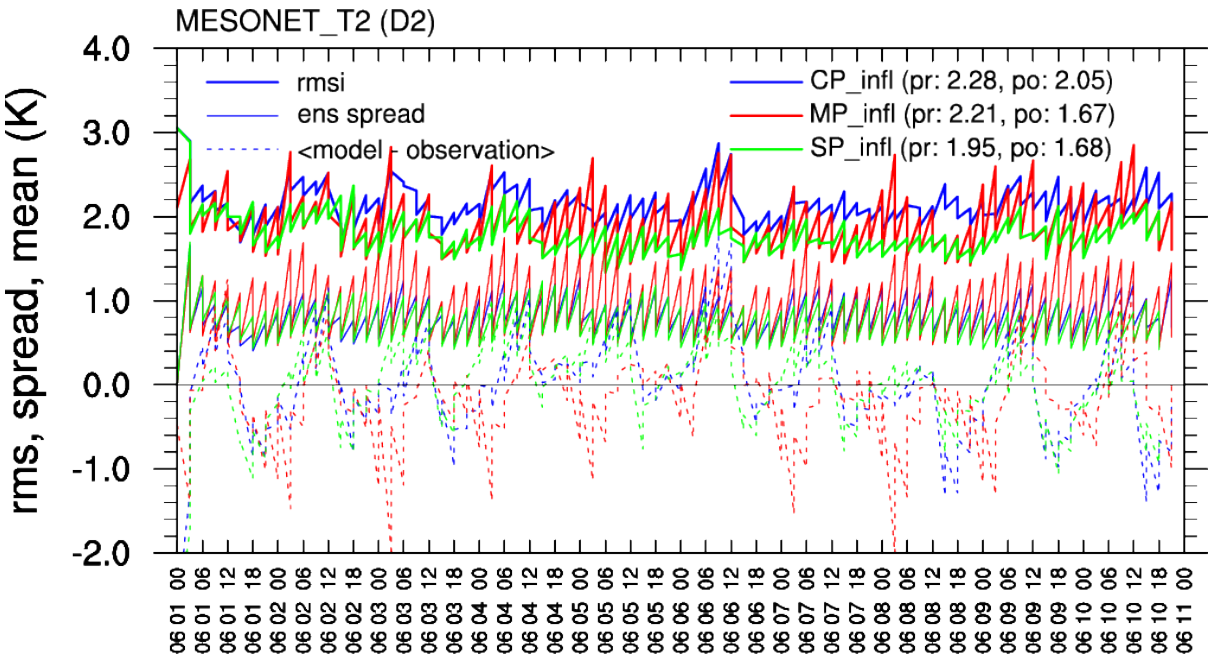

Figure 3: *Spread around ensemble mean (thin curves) and root-mean-square innovation (rmsi) of ensemble mean (thick curves) for temperature T in 2m (in K) verified against an independent data (Mesonet observations) over the continental United States. Spread and error curves are shown for three ensemble systems: control physics CNTL (blue), multi-physics PHYS (red) and stochastic backscatter STOCH (green). The ensemble mean bias for each ensemble system is shown by the dashed lines.*

4.1 Systematic Biases in the 500hPa Geopotential Height Fields

First we investigate the impact of the different model refinements on the mean bias in 500 hPa geopotential height fields (Z500, hereafter) in the Northern Hemisphere extra-tropics during boreal winter. (Fig. 4). The geopotential heights are averaged over the four winter months December-March and all years, and the bias is calculated as the difference between the mean winter climate in the analysis and each experiment. The bias in LOWRES signifies an overly zonal flow, especially in the storm-track regions. This systematic error has been persistent across a number of model versions for more than a decade (e.g., Jung et al. 2010). Including a stochastic model error representation reduces the systematic Z500 bias in both, the North Pacific/North American, and North Atlantic regions substantially, but the circulation continues to be somewhat too zonal. Increasing horizontal resolution leads to a very similar improvement in systematic model bias in Z500 (Fig. 4 c). Note that the increased horizontal resolution of HIGHRES not only affects the atmospheric dynamics, but also the representation of orographic details. Resolving orography in itself is expected to lead to a better representation of the Northern Hemispheric circulation (Jung et al., 2010) The biggest reduction in systematic bias is achieved by improving the physical parameterizations (Fig. 4 d): the dipole pattern characterizing the error over the Western Pacific has disappeared completely and over the European/North Atlantic region only a residual error remains. To determine statistical significance of the bias, a Student's t-test at each gridpoint was performed. Hatched areas denote the regions where the model climate differs from the observed climate at the 95% confidence level.

Our analysis shows that the mean bias in Z500 can be reduced in more than one way: increasing horizontal resolution, introducing stochastic perturbations or changing the physics packages all lead to significant improvements. While the improved physical parameterizations of the experiment PHYS results overall in the closest match to the analysis, differences on a regional level exist. For example, the large bias of LOWRES over the Western United States only completely disappears in the experiment STOCH (Fig. 4 b). It remains to be shown if these improvements are the result of a better representation of local

processes or if they are a teleconnection response to a better representation of tropical variability.

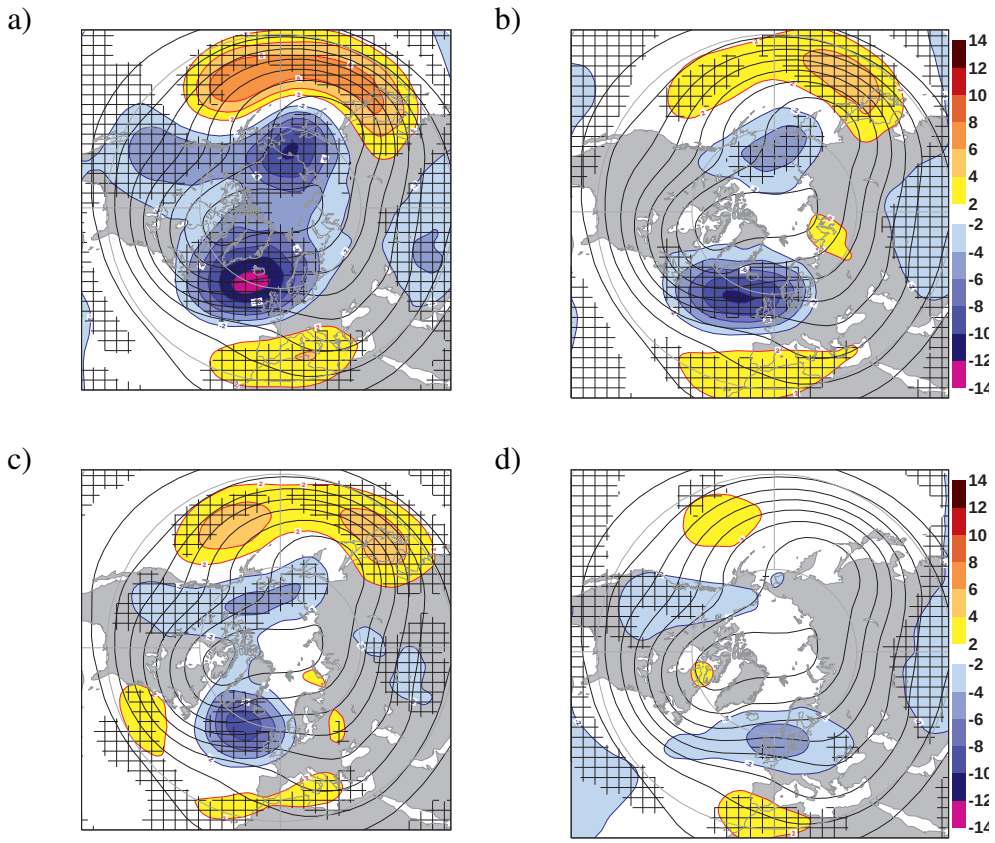

Figure 4: Mean systematic error of 500 hPa geopotential height fields (drawn as shading) for extended boreal winters (December–March) of the period 1990-2005. Errors are defined with regard to the observed mean field (contours), consisting of a combination of ERA-40 (1990-2001) and operational ECMWF analyses (2002-2005). Shown are the systematic error of experiments (a) low-resolution (LOWRES), (b) stochastic kinetic-energy backscatter (STOCH), (c) high-resolution (HIGHRES) and (d) improved deterministic parameterizations (PHYS). Significant differences at the 95% confidence level based on a Student's t-test are hatched.

4.2 Convectively-coupled Tropical Waves

To determine the impact of the different model error schemes in the tropics, we investigate convectively-coupled tropical waves (e.g., Lin et al. 2006). Following Wheeler and Kiladis (1999), wavenumber-frequency spectra for observed and simulated anomalies of out-going longwave radiation (OLR) in the latitude band 5°S - 5°N have been computed. Here, we focus on the component that is symmetric about the equator.

The total symmetric spectrum in the observations is characterized by strong eastward propagating and weaker westward propagating Rossby waves (Fig. 5a). Generally the opposite is true for most model experiments, where there is more power in the westward propagating than in the eastward propagating modes (Fig. 5c,e,g). The exception is the experiment PHYS, which exhibits strongly westward propagating waves. However, even in this experiment the spectra remain too steep, indicating too little power in the higher frequencies.

In the observations, tropical variability associated with the Madden-Julian oscillation (MJO, Madden

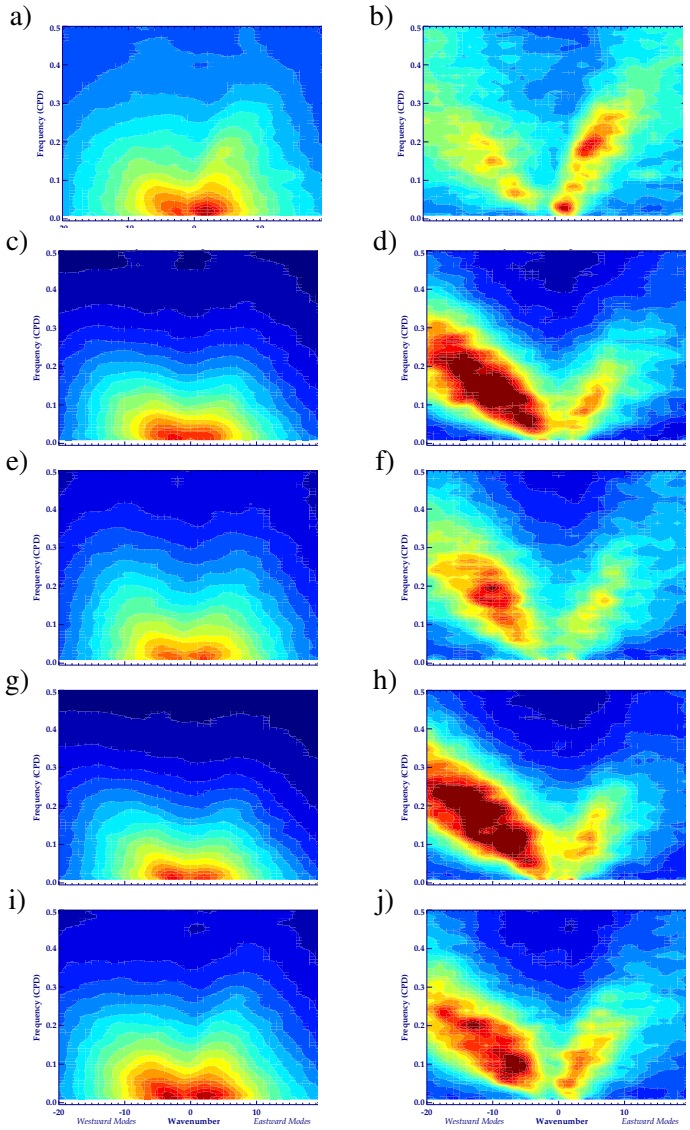

Figure 5: Mean wavenumber-frequency spectra for the symmetric component of daily tropical outgoing longwave radiation anomalies (in K^2) during December-March for the period (1990.3005) before (a,c,e,g,i) and after (b,d,f,h,j) a background has been removed (see text for details). The mean annual cycle has been removed prior to the computation of the wavenumber-frequency spectra. The wavenumber-frequency spectra are depicted for (a,b) satellite data from NOAA and for experiments (c,d) LOWRES, (e,f) STOCH, (g,h) HIGHRES and (i,j) PHYS. (From Berner et al., 2011b)

and Julian 1971) leads to a distinct peak in the spectrum of eastward propagating modes of wavenumber 1-3, which has markedly more power than its counterpart in the westward modes (Fig. 5a). None of the model experiments captures the MJO peak correctly.

Following Wheeler and Kiladis (1999), we define a signal-to-noise spectrum by removing the background spectrum for each experiment separately. The erroneous dominance of westward propagating modes and missing MJO-variability is even more evident in these signal-to-noise spectra, (Fig. 5, right column). The signal-to-noise spectra of experiments STOCH and PHYS resemble the observed signal closest, although they both misrepresent the MJO peaks. PHYS is slightly better at capturing the amplitude of eastward propagating modes, while STOCH is more successful at damping westward propagating Kelvin waves. HIGHRES shows no improvement over the wavenumber-frequency spectrum of LOWRES (Fig. 5g,h).

In summary, increasing horizontal resolution did not improve the wavenumber-frequency spectra of OLR, but introducing a stochastic kinetic-energy backscatter scheme or improving the deterministic parameterizations resulted in less westward and more eastward propagating modes and an overall better match to the observed spectra. The closest match was obtained by refining the physical parameterization, especially the convective parameterization scheme (Bechthold et al., 2008).

5 Conclusion

This report summarizes work from different publications or manuscripts in preparation, which study the impact of model-error schemes in general and stochastic kinetic-energy schemes in particular on different timescales. The time-scales span from 3-hourly cycled ensemble data assimilation experiments in WRF (Ha and Snyder , 2011; Snyder et al., 2011) over 60h-short range ensemble forecasts with WRF (Berner et al., 2011, Hacker et al., 2011) to seasonal climate predictions with ECMWF's IFS. On all timescales the inclusion of a stochastic parameterization or other model refinements – either aimed at model-error or subgrid-scale fluctuation representation – was beneficial in terms of variability and a reduction of model bias and mean error. We interpret this as evidence that the subgrid-scale fluctuations play an important role and must be better represented. Ideally, stochastic fluctuations should be part of the physics-package development and not added on *a posteriori*, and some work along the lines has emerged in recent years (e.g., Plant and Craig, 2008). In the meanwhile, uncertainty predictions utilizing stochastic parameterizations are beginning to be a true alternative to multi-model approaches.

Acknowledgments

The author would like to acknowledge the the following collaborators, who all contributed essentailly to the presented body of work: Aimé Fournier, Soyoung Ha, Joshua Hacker, Thomas Jung, Tim Palmer and Chris Snyder.

References

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

Anderson, J. L. (2007). An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A 59*, 210–224.

Bechtold, P., M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. Rodwell, F. Vitart, and G. Balsamo (2008). Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quart. J. Roy. Meteor. Soc. 134*, 1337–1351.

Berner, J., F. J. Doblas-Reyes, T. N. Palmer, G. Shutts, and A. Weisheimer (2008). Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal predicition skill of a global climate model. *Phil. Trans. R. Soc A 366*, 2561–2579.

Berner, J., S.-Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder (2011). Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multi-physics representations. *Mon. Wea. Rev. 139*, 1972–1995.

Berner, J., T. Jung, and T. N. Palmer (2011). Systematic model error: The impact of increased horizontal resolution versus improved stochastic and deterministic parameterizations. *J. Climate submitted*.

Berner, J., G. Shutts, M. Leutbecher, and T. Palmer (2009). A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci. 66*, 603–626.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu (2005). A comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev. 133*, 1076–1097.

Cavallo, S., J. Berner, and C. Snyder (2011). Initial tendencies and model bias in WRF. *Mon. Wea. Rev. in preparation*.

Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin (2010). Toward random sampling of model error in the canadian ensemble prediction system. *Mon. Wea. Rev. 138*, 1877–1901.

Doblas-Reyes, F. (A. Weisheimer, M. Déqué, N. Keenlyside, M. McVean, J.M. Murphy, P. Rogel, D. Smith, T.N. Palmer, 2009). Addressing model uncertainty in seasonal and annual dynamical seasonal forecasts. *Quart. J. Roy. Meteor. Soc. 135*, 1538–1559.

Epstein, E. S. and E. J. Pitcher (1972). Stochastic analysis of meteorological fields. *J. Atmos. Sci. 29*, 244–257.

Ha, S. and S. Synder (2011). The impact of surface data assimilation in ensemble data assimilation using the ensemble kalman filter. *Mon. Wea. Rev. in preparation*.

Hacker, J. P., S.-Y. Ha, C. Snyder, J. Berner, F. A. Eckel, E. Kuchera, M. Pocernich, S. Rugg, J. Schramm, and X. Wang (2011). The U.S. Air Force Weather Agencys mesoscale ensemble: Scientific description and performance results. *Tellus A,* Early View, DOI: 10.1111/j.1600-0870.2010.00497.x.

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting – i. basic concept. *Tellus 57A*, 219–233.

Hannay, C., D. Williamson, J. J. Hack, J. T. Kiehl, J. G. Olson, S. A. Klein, C. S. Bretherton, and M. Köhler (2009). Evaluation of forecasted southeast pacific stratocumulus in the NCAR, GFDL, and ECMWF models. *J. Climate 22*, 2871–2889.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell (1996). A system simulation approach to ensemble prediction. *Mon. Wea. Rev. 124*, 1225–1242.

Jung, T. (G. Balsamo, P. Bechtold, A. Beljaars, M. Köhler, M. Miller, J.-J. Morcrette, A. Orr, M.J. Rodwell and A.M. Tompkins, 2010). The ECMWF model climate: Recent progress through improved physical parametrizations. *Quart. J. Roy. Meteor. Soc. 136*, 1145–1160.

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

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, , and S. Surendran (2000). Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate 13*, 4196–4216.

Lin, J. (N. Kiladis, B. Mapes, K. Weickmann, K. Sperber, W. Lin, M. Wheeler, S. Schubert, A. Del Genio, L. Donner, S. Emori, J.-F. Gueremy, F. Hourdin, P. Rasch, E. Roeckner and J. Scinocca, 2006). Tropical intraseasonal variability in 14 IPCC AR4 climate models. *J. Climate 19*, 2665–2690.

Madden, R. A. and P. R. Julian (1971). Detection of a 40-50 day oscillation in the zonal wind in the tropical pacific. *J. Atmos. Sci. 5*, 702–708.

Mason, P. and D. Thomson (1992). Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech. 242*, 51–78.

Murphy, J., D. Sexton, D. Barnett, G. Jones, M. Webb, M. Collins, and D. Stainforth (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature 430*, 768–772.

Palmer, T. N. (2001, January). A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterization in weather and climate prediction. *Quart. J. Roy. Meteor. Soc. 127*, 279–304.

Pitcher, E. J. (1977). Application of stochastic dynamic prediction to real data. *J. Atmos. Sci. 34*, 3–21.

Plant, R. S. and . G. C. Craig (2008). A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci. 64*, 87–105.

Rodwell, M. and T. Palmer (2007). Using numerical weather prediction to assess climate models. *133*, 129–146.

Shutts, G. J. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc. 612*, 3079–3102.

Shutts, G. J. and T. N. Palmer (2003). The use of high resolution numerical simulations of tropical circulation to calibrate stochastic physics schemes. In *ECMWF/CLIVAR workshop on simulation and prediction of intra-seasonal variability with the emphasis on the MJO*, available from http://www.ecmwf.int/publications, pp. 83–102. ECMWF.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers (2008). A Description of the Advanced Research WRF version 3. Technical report, NCAR Tech. Note, NCAR/TN-475+STR, 113pp.

Solomon, E., S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller (2007). *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, New York.

Stephenson, D., A. Hannachi, and A. O'Neill (2004). On the existence of multiple climate regimes. *Quart. J. Roy. Meteor. Soc. 130*, 583–605.

Synder, S., S. Ha, and J. Berner (2011). Performance of model error schemes in ensemble data assimilation using the ensemble kalman filter. *Mon. Wea. Rev. in preparation*.

Tennant, W. J., G. J. Shutts, and S. A. T. A. Arribas (2011). Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev. 139*, 1190–1206.

Wheeler, M. and G. N. Kiladis (1999). Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber-frequency domain. *J. Atmos. Sci. 56*, 374–399.

Wilks, D. (1995). *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, San Diego, 464pp.