

Stochastic forcing, ensemble prediction systems, and TIGGE

Ensemble prediction systems are increasingly able to provide useful probabilistic forecasts. The representation of near-surface and model uncertainties in ensemble systems, and the correction and filtering of forecast errors remain as significant challenges.

Thomas M. Hamill¹ and Richard Swinbank²

¹ NOAA Earth System Research Lab, Physical Sciences Division, Boulder, Colorado, USA

² Met Office, Exeter, UK

ABSTRACT

Many operational NWP centres now produce global medium-range (≤ 14 day) and higher-resolution, limited-area, shorter-range (≤ 3 day) ensemble forecasts. These provide probabilistic guidance and early warning of the likelihood of high-impact weather. There are two main challenges in the design of ensemble prediction systems: (1) properly simulating the initial condition uncertainty, including the definition of the initial ocean, land, and sea-ice states, and (2) properly simulating the uncertainty due to inadequate representations of physical processes, especially parameterizations.

Post-processing the output from the ensemble prediction systems using past forecasts and observations/analyses can dramatically reduce systematic errors in forecast products and improve skill and reliability. The generation of products from multi-model ensembles (facilitated by the TIGGE database, sharing global operational ensemble forecasts) has also been shown to frequently improve the skill and reliability of ensemble predictions.

1. INTRODUCTION

From the earliest days of weather forecasting, there has been an appreciation that there are inevitable uncertainties in weather prediction. Admiral FitzRoy, the founder of the UK Met Office, wrote in a letter to the British *Times* newspaper some 150 years ago that “*forecasts are expressions of probabilities – and not dogmatic predictions.*” However, only in the last two decades has it become computationally feasible to apply methods that objectively calculate the state-dependent uncertainties in weather forecasts. Prior to this, forecast guidance typically consisted of one model integration, and forecasts were expressed deterministically. Now, while some deterministic models are often run at higher resolutions, an increasing role is played by

ensembles of forecasts that are integrated from sets of slightly different initial conditions, and employ methods to simulate the uncertainty of the forecast model itself. The intent is to make sharp (specific) yet reliable state-dependent probabilistic forecasts (Gneiting et al. 2007) directly from the ensemble guidance.

The penetration of probabilistic concepts throughout the forecast process is not yet complete. Customers of weather forecasts still generally expect deterministic expressions for upcoming forecasts, even though in many cases more appropriate decisions could be made when leveraging probabilistic information (Zhu et al. 2002). Further, though ensemble prediction waxes and deterministic predictions wanes at operational centres, prediction systems are still commonly evaluated with deterministic verification methods. As discussed by Palmer (2014), this can lead us to inappropriate conclusions about whether we have improved our prediction systems. Common deterministic verification metrics include root mean square (RMS) errors, anomaly correlations, and threat scores (Wilks 2011, Joliffe and Stephenson 2012). Unfortunately, such measures often penalize the forecasting of small-scale features if they are not predictable; a smoother forecast lacking such scales of motion is assessed as providing higher skill. The existence of smaller-scale phenomena is of course realistic and consistent with a continuous energy spectrum across scales (Nastrom and Gage 1985). In contrast, the use of probabilistic verification metrics such as the continuous ranked probability skill score (CRPSS; Wilks 2011, Chapter 7) suffers no such consequences, though observation errors should be accounted for (Candille and Talagrand 2008). A forecast is rewarded for accurately predicting the probability as specifically as possible, subject to their being reliable. If an ensemble of forecasts is missing small scales present in the analysis, it will be penalized for this lack of variability.

Despite such impediments, over the last decade ensemble predictions have matured, in part from a more thorough understanding of the underlying theoretical concepts, in part from better and higher-resolution prediction systems afforded by larger computers, and in part from research. This includes collaborative studies facilitated by WMO's THORPEX program and its shared TIGGE (THORPEX Interactive Grand Global Ensemble, Bougeault et al. 2009) database. This chapter reviews the current state of the science of ensemble prediction and suggests fruitful areas for further research. In section 2, we touch upon the advances in construction of initial conditions; the operational centres now have as many similarities in their methods as differences, thanks to advances in ensemble-based data assimilation. We also discuss uncertainty in the state of the lower boundary (ocean, land, ice). Section 3 focuses on a current key development in ensemble prediction, the use of stochastic forcing methods to treat model uncertainty. Those techniques are at an earlier stage of the development cycle than initial-condition uncertainty. Section 4 describes various practical methods to

address these systematic errors, both by combining predictions from different centres' ensemble prediction systems and by using more objective statistical methods that correct today's forecast based on discrepancies noted between past forecasts and observations/analyses. Finally, in section 5 we discuss some applications of ensembles to produce risk-based weather forecasts, particularly as they enable improved forecasts of high-impact weather events. Some related material on ensemble design and post-processing is also available in the chapter "Global Environmental Prediction."

2. ENSEMBLE INITIAL CONDITIONS.

2.1 Introduction.

At the dawn of operational ensemble prediction, the scientific debate about ensembles focused largely on the method of construction of the ensemble of initial atmospheric model states, meant to sample the uncertainties in the initial conditions. On the basis of trying to provide a medium-range ensemble that explained as much as possible of the forecast error, ECMWF scientists proposed the use of "singular vectors" (Buizza and Palmer 1995, Molteni et al. 1996, Li et al. 2008), perturbations that grow the fastest in time given the chosen initial and final norms for measuring perturbation size. For both times, ECMWF chose the total-energy norm. In more recent years, ECMWF has blended in perturbations generated by running parallel, reduced-resolution 4D-Var cycles that assimilate perturbed observations. They refer to this as "*ensembles of data assimilations*," or EDA (Bonavita et al. 2012; Lang et al. 2014).

Under the assumption that the most critical initial-condition errors would be inherited from the background forecast in the data assimilation process, NCEP initially used the "bred vector" method (Toth and Kalnay 1993, 1997). Initially random perturbations were repeatedly forecast forward in time to the next assimilation cycle, then rescaled and adjusted in amplitude to be generally consistent with presumed analysis uncertainties. Analysis uncertainties were described by a "mask," i.e., a field of spatially varying analysis variances. Over the last five years, NCEP has used a modified version of the breeding technique known as "ensemble transform with rescaling," or "ETR" (Wei et al. 2008). This procedure added ortho-normalization so that pairs of perturbations did not result in forecasts with as highly correlated errors.

The Canadian Meteorological Centre (CMC) initially used the "perturbed observations" method to quantify the effect of observation errors on the uncertainties in the initial state (Houtekamer and Derome 1995). Parallel 3D-Var data assimilation cycles were conducted, with each member cycle updated with perturbed observations consisting of the control observations plus realizations of random noise consistent with observation-error statistics. In recent years, CMC has migrated to the use of an

ensemble Kalman filter, or “EnKF” (e.g., Evensen 1994; Houtekamer and Mitchell 1998, Burgers et al. 1998, Hamill 2006) whereby the ensemble provides background-error statistics for the data assimilation. In 2015, NCEP will also migrate to using initial perturbations from 6-hour forecasts generated from EnKF analysis perturbations. The UK Met Office uses a related technique known as the ensemble transform Kalman filter, or ETKF (Wang and Bishop 2003, Bowler et al. 2008).

There actually is a theoretical basis for the optimal choice of initial condition perturbations, outlined in Ehrendorfer and Tribbia (1997). Under assumptions of Gaussianity, linearity of error growth, and the choice of final time’s norm, analysis-error covariance singular vector initial perturbations will provide the maximum amount of explained forecast error variance at the chosen final time. Restated, these are the select perturbations that are initially consistent with analysis-error statistics while growing most quickly. With this theoretical result in mind, we have the ability to understand the various approximations used by the various centres. While ECMWF’s total-energy singular vector perturbations grow quickly, their singular vectors are sub-optimal due to the choice of total energy rather than analysis-error covariance as the initial norm. Their singular-vector perturbations may have too low amplitudes near the tropopause and the surface relative to what they would have were they consistent with the initial analysis error covariance norm (Barkmeijer et al. 1998, 1999, Hamill et al. 2002b). This may result in unrealistic perturbation amplitudes in very short-range forecasts. ECMWF’s more recent EDA technique will encounter sampling error from the limited number of perturbed data assimilations conducted and uses an initial covariance that is not flow-dependent (note: this will change to a fully flow-dependent cycling in the next model cycle; personal communication, F. Rabier, 2015). These perturbations are not optimized to grow as quickly as possible, either.

Bred and ETR perturbations only approximately are consistent with initial analysis error statistics; they are not explicitly estimating the analysis-error covariances from data assimilation cycle. The procedure only rescales (and in the case of ETR, orthogonalizes) the forecast perturbations. Thus, they cannot account for the randomization effects from observation assimilation (Hamill et al. 2002a). Further, they are optimized for past forecast error growth rather than future growth.

CMC’s and NCEP’s EnKF approach are conceptually appealing, in that the resulting initial perturbations are more closely designed to represent analysis-error statistics. In practice, the realism of these perturbations may be limited by several factors. These include the fidelity of model-error representations used in the generation of the ensemble (Mitchell et al. 2002, Zhang et al. 2004, Hamill and Whitaker 2005, 2011, Anderson 2009, Whitaker and Hamill 2012), the underlying assumption of

Gaussian error statistics, and the limited ensemble size which requires the introduction of ad-hoc procedures like “covariance localization” (Houtekamer et al. 2001, Hamill et al. 2001) that can introduce imbalance. The initial perturbations are typically random rather than pre-selected for optimal forecast-error growth. The Met Office ETKF represents a low-dimensional approximation to the EnKF that doesn’t involve a full assimilation cycling and covariance localization (Bowler et al. 2008, Bowler and Mylne 2009). Consequently, the rescaling and rotation process of the ETKF strips out too much variance from the prior forecast ensemble, and hence perturbations must be dramatically scaled up in size before use.

In general, limited computational capacity requires making simplifications of one sort or another. The choice of which simplification to apply has varied between the operational centres. Still, over the past two decades the various centres have evolved toward using methods that are increasingly consistent with the theoretical ideal outlined by Tribbia and Ehrendorfer (1997).

2.2 Underpinning research.

While methods for initializing ensembles of atmospheric states are becoming more similar, the ensemble prediction systems continue to have too little spread near the earth’s surface. This is likely because the uncertainty in the land, water, or ice state is currently not treated at all or is treated sub-optimally. Addressing this is an important direction for future research in the design of ensemble prediction systems.

Consider the land state. As outlined in Sutton et al. (2006), near-surface temperature forecasts and precipitation forecasts can be sensitively dependent on the initial state of soil moisture, and furthermore the analyses of soil moisture are often quite error-prone. The soil moisture states are commonly estimated through the offline cycling of a land-surface model forced by analysed temperatures, humidities, and precipitation. These analyses can be highly imperfect, and the land-surface model itself can have significant imperfections, such as mis-specifications of model constants such as soil hydraulic conductivity or surface roughness length. Various centres have recently introduced some methods for increasing near-surface variability, such as by perturbing soil moisture in some form (e.g., Lavaysse et al. 2013, Tennant and Beare 2014). Still, there is much room for improvement. Research is needed into improved ways to simulate the range of model structural and initial-condition uncertainties near the surface (e.g., Hacker et al. 2007, Hacker and Rostkier-Edelstein 2007), thereby providing more realistic ensembles of near-surface temperatures and humidities as well as ranges of initial soil states.

There is also increasing demand for the medium-range ensemble predictions to be extended to sub-seasonal time scales. For these longer-lead predictions, the spread of the atmospheric ensemble will usually grow to near its climatological variability. What predictive skill remains may be in only a few low-frequency modes of oscillation, some related to ocean-state oscillations such as El Niño. Hence, for these extended-range predictions, it may be necessary to quantify the initial uncertainty of the ocean state and how that uncertainty evolves through the duration of the forecast. Ideally, there would be numerical consistency between atmospheric perturbations and ocean perturbations; for example, in a member of the ensemble that has larger wind speeds than average, there might be greater vertical mixing in the ocean state for that ensemble member. Since synoptic-scale variability peaks at $O(1 \text{ week})$ while ocean variability peaks at $O(1 \text{ year})$, this discrepancy of time scales makes the direct coupling of ocean and atmospheric initialization via methods such as the EnKF potentially problematic, and in need of further exploration (e.g., Yang et al. 2009, Ueno et al. 2010). Relatedly, ensemble predictions out to sub-seasonal time scales may be sensitively dependent on the initialization of sea-ice amount and thickness (Juricke et al. 2014) as well as land snow cover (Jeong et al. 2013).

There are complicated ensemble initialization challenges associated with very high-resolution, shorter-range forecasts as well. The shorter-range, convection-permitting models have the ability to provide forecasts with detail at the scale of individual thunderstorms (e.g., Hohenegger et al. 2008, Clark et al. 2009, 2010, Schwartz et al. 2010, 2014, Johnson and Wang 2012, Duc et al. 2013). However, data assimilation systems like the EnKF have underlying assumptions such as Gaussian error statistics (Lawson and Hansen 2004), assumptions that may be more frequently unrealistic at the convective scale. One would not a priori expect Gaussian ensemble error statistics of cloud liquid water, for example, in the region of a thunderstorm; they might have two dominant modes, no cloud water (with no thunderstorm) and ample cloud water (with thunderstorm). The optimal methods for data assimilation and ensemble initialization in the presence of such non-Gaussian error statistics are not clear. Further, methods are desired that simultaneously perform a high quality analysis and initialization of both the larger and smaller scales of motion. There are some suggestive directions; there is much research into particle-filter methods as a potential solution (e.g., Gordon et al. 1993, Doucet et al. 2001) but there are also concerns that the “curse of dimensionality” may make such methods impractical with the very-high dimensional systems common in weather prediction (Snyder et al. 2008). Other new directions to address non-Gaussianity include the rank-histogram ensemble filter (Anderson 2010, Metref et al. 2014) and methods that deal with position errors (Ravela et al. 2007, Nehrkorn et al. 2015). ETKF methods are being tried by the German Weather Service (Harnisch and Keil, 2014).

A common approach for providing spatially detailed forecasts at short leads is to use a high-resolution, limited-area ensemble prediction system (Hamill and Colucci 1997, 1998, Frogner et al. 2006, Bowler et al. 2008, 2009, Bowler and Mylne 2009, Aspelien et al. 2011, Romine et al. 2014) commonly with lateral boundary conditions provided by a global ensemble prediction system. There are a host of challenges associated with the use of limited-area ensemble prediction systems with one-way interactive nests, with many outlined in Warner et al. (1997). The one-way nesting prohibits scale interactivity, whereby developing features inside the limited-area domain can affect the larger scales of motion outside the domain. If the limited-area prediction is carried out in a very small domain, the ability to predict detailed features may be overwhelmed by the “sweeping” in of lower-resolution information from the global model. In an ensemble context, it is also important to provide lateral boundary conditions with appropriate variability (Nutter et al. 2004, Torn et al. 2006).

2.3 Linkages and requirements.

Several WMO-sponsored projects will help address many of the research questions identified here. For example, the Sub-seasonal to Seasonal (S2S) prediction project (WMO 2013) will facilitate coordinated explorations of issues related to prediction at these time scales, including methods for initializing coupled models and how to generate ensembles. WMO is also developing a program for improving forecasts of high-impact weather (Jones and Golding 2014) that will facilitate further exploration in the design of ensemble prediction systems. Inter-comparisons of prediction systems such as Buizza et al. (2005) have been facilitated in recent years by the TIGGE project (Bougeault et al. 2009, Swinbank et al. 2015); more inter-comparisons to understand the evolving relative strengths and weaknesses of various ensemble methodologies are still needed.

Despite the widespread interest in short-range, high-resolution ensemble prediction, collaboration between prediction centres on this avenue of research has been more difficult; each centre typically chooses a more limited region of interest driven by their country’s need and the tight production timelines for short-range forecast guidance. Still, given the success with global multi-model ensemble prediction (see section 4 below), there may be significant benefit in future coordination between prediction centres (Paccagnella et al. 2011). For example, were each centre in Europe to enlarge the domains of their high-resolution ensemble forecasts to a general common domain while each decreased their ensemble size, it may be possible to have a neutral impact on computational expense at each centre while leveraging the advantages of the multiple dynamical cores, initialization methods, and parameterization suites demonstrated with global models under TIGGE. This may result in ensembles that do a

better job of spanning the forecast uncertainty. Such approaches would require greater coordination between the prediction centres.

The validation of ensemble prediction systems for high-impact variables such as precipitation amount and type are hindered by challenges in the sharing of data and by different data processing methods. Member countries are highly encouraged to share their raw and synthesized data with each other.

3. STOCHASTIC FORCING.

3.1. Introduction.

Finite grid spacings result in many meso- and microscale phenomena being unresolved in atmospheric models. In such cases the effects of the unresolved scales upon the resolved scales are typically “parameterised,” that is, estimated with closure assumptions that depend on the resolved-scale weather parameters. Parameterisation schemes may be developed in several ways (Craig 2014). One way is to base parameterisations on process models that are based on theory developed independent of the atmospheric model. A prediction of mean effects are produced, and sometimes higher moments are estimated as well. Another approach is to use systematic truncations of equations in high-resolution. A third approach is to use ad-hoc formulations to produce a desired effect in the model (for example, hyperdiffusion to control grid-scale noise).

Commonly, these parameterisations are deterministic; two grid boxes with identical grid-scale states will have the same parameterised forcings diagnosed for each box, even though the sub-grid scale details may vary between them. Where the model grid size is large, it is assumed that the phenomena being parameterised is likely to be much smaller than the size of the grid box, and its mean statistical properties of the sub-grid effects can be estimated with reasonable approximation (Fig. 1a). As models are run using finer grid spacings, two challenges become increasingly apparent. First, note that Fig. 1a suggests a simplification that commonly is not valid, that the unresolved phenomena has one characteristic scale clearly smaller than the grid scale. Much more commonly, there is a spectrum of motions (Fig. 1b); for example, not all thunderstorms are in fact the same size. Second, there may no longer be a clear scale separation between the phenomena in question and the size of the grid box; the spread of estimates of sub-grid effect between grid boxes with nearly identical large-scale states can be of comparable magnitude to the mean estimate (Plant and Craig 2008). A consequence of neglecting this in the parameterisation design is that the ensemble predictions may be unduly similar to each other. This deficiency of spread leads to over-confident probabilistic forecasts.

Apart from the issue of scale separation, there are several other sources of uncertainty in parameterizations. The grid-scale effects of some subgrid-scale processes may depend on unknowable sub-grid details, not just grid-scale details. An example of this is the effect of greatly variable cloud droplet size spectra on the reflection, absorption, and transmission of short-wave radiation. Some details of grid-scale characteristics may be mis-estimated and should be treated as random variables. For example, a mis-estimation of soil characteristics could lead to errors in energy balance affecting surface-temperature forecasts and convective initiation (Sutton et al. 2006).

3.2 Underpinning research.

A variety of methods have been used to simulate model uncertainty. One of the simplest methods is to estimate probabilities using multi-model ensembles. Since weather prediction centres tend to have developed their models and parameterisations suites somewhat independently, the combination of several ensembles may account for some of the knowledge uncertainty. Multi-model ensembles will be discussed in more detail in section 4 below.

A related approach is to use multiple parameterisations amongst the ensemble members (e.g., Charron et al. 2010, Berner et al. 2011). For example, member 1 may use a Kain-Fritsch deep convective parameterisation, member 2 a Tiedtke parameterisation, and so on. Experimentally, many have noted significant increases in spread and some improved skill with the use of such approaches. Still, there are significant drawbacks to the multi-parameterisation approach. It becomes necessary to maintain a library of multiple parameterisations rather than one, increasing software maintenance expense. Also, should one of the parameterisations be improved significantly, it would be desirable to use that one consistently, so that its improvement affects all ensemble members, rather than retaining out-of-date parameterisations for the sake of diversity. Finally, there can be a loss of “*exchangeability*,” the desirable property that all ensemble members have similar error statistics.

In a class of schemes referred to as “perturbed parameters,” a number of key parameterisation constants are identified whose values are uncertain but which have a significant effect on the model tendencies (e.g., Bowler et al. 2008, Charron et al. 2010). A range of plausible but different parameterization constants are used across the ensemble members to represent the uncertainty in those parameter values. In some versions of this approach the perturbed parameters are held constant for a given ensemble member, while in other versions they are varied with time. Such methods are defensible to the extent that parameters are perturbed consistent with their uncertainty

and are fully tested in combination with other parameterisation constants; without rigorous testing and randomizing the values of the constants over times, it's possible that specific combinations of perturbed parameters will result in members with non-exchangeable statistics and growing systematic errors.

A very promising approach to representing model uncertainties is the use of stochastic forcing methods, i.e., changing the design of parameterisations that are known to have uncertainties so that they provide stochastic rather than deterministic estimates of the sub-grid effects. One of the earliest methods is “*stochastically perturbed physical tendencies*” (SPPT; Buizza et al. 1999, Palmer et al. 2009, Bouttier et al. 2012). SPPT is a somewhat ad-hoc method that multiplies the total parameterised tendency by a random number that fluctuates in time and space. These methods have been shown by several centres to increase ensemble spread to be more consistent with mean error.

Realistic representation of model uncertainties requires that, as far as is possible, takes into account an understanding of the relevant physical processes. Stochastic kinetic energy backscatter (SKEB; Shutts et al. 2008, Berner et al. 2008), attempts to take account of the unphysical energy loss that typically occurs in models as a consequence of numerical diffusion, mountain drag, and deep convection. This method has had the practical effect of making the model energy spectra in ensemble members look more like $-5/3$ spectral slope expected by theory for the mesoscale (Nastrom and Gage 1985), thus notably increasing spread in the smaller scales of motion. It is yet unclear, however, whether SKEB is preferable to running ensemble prediction systems with reduced hyperdiffusion; the two would likely have similar effects.

These considerations have stimulated significant new research into physically based stochastic parameterizations, whereby stochasticity is incorporated into the parameterization in physically realistic ways. Examples of research into stochastic parameterization for deep convection include Lin and Neelin (2000, 2003), Majda (2007), Tompkins and Berner (2008), Plant and Craig (2008), Teixeira and Reynolds (2008), Frenkel et al. (2012), Peters et al. 2013, Grell and Freitas (2014), and Keane et al. (2014). Stochastic parameterization concepts touching on cloud microphysics includes Posselt and Vukicevic (2010), and van Lier-Walqui et al. (2012).

A further consideration is that models need to take into account the uncertainties in the forcing of the atmosphere from the lower boundary. Work addressing the uncertainties in land-surface processes includes Lavaysse et al. (2013) and Tennant et al. (2013); ocean-atmospheric uncertainty has been studied by McClay et al. (2012). A longer list of reading material on this subject is available from the 2011 WMO- and ECMWF-sponsored workshop on model uncertainty (ECMWF, 2011).

3.3 Linkages and requirements.

The 2011 workshop mentioned above provided a succinct summary of the necessary research. The summary stated,

“the stochastic parametrization paradigm needs further development at the process level, and hence needs to be incorporated as part of general parameterisation development. Key tools will include sophisticated analyses of observational datasets, output from cloud resolving models, and analyses from objective data assimilation. Data assimilation techniques themselves will benefit from better representations of model uncertainty.”

The development of physically based stochastic parameterisations at the individual process level will proceed more rapidly if the scientists involved in the parameterisation development and the ensemble system development collaborate. The parameterisation methods need to be consistent with the physical laws while faithfully replicating the statistics of the processes. For some parameterisations, the entire formulation might be cast probabilistically, such as the representations of the sub-grid distributions of vertical velocity, temperature, and cloud liquid water as joint PDFs, (e.g., Larson and Golaz 2005, Larson et al. 2012, Bogenschutz and Krueger 2013). With such a formulation, generating a realistic range of parameterization outputs for the ensemble is somewhat more straightforward.

Determining the space- and time-dependent relationships of the PDFs will require further study. Parameterisation inputs may have correlated errors; for example, a microphysics parameterization may infer a different drop-size distribution in the presence of few vs. many aerosols, and the aerosol concentration likely is correlated from one grid cell to the next. Stochastic parameterisation outputs should also have appropriately correlated structures. For example, the parameterisation of convection should be “non-local” in many circumstances; the organization of convection is important for the realistic simulation of mesoscale convective systems and their influence on larger-scale phenomena.

Modern ensemble-based data assimilation methods provide one useful way of evaluating whether changes in simulating model uncertainty are realistic. With an improved model uncertainty method, the space-time background-error (first-guess) covariances should become more realistic, resulting in a more appropriate adjustment of the background forecast(s) to the observations. Over time, then, the statistics of the

mean absolute error of observations minus the forecasts¹ should decrease. There are other novel methods (e.g., Scheuerer and Hamill 2014) for evaluating relationships between forecast state elements and whether they resemble the relationships between observed states.

As noted earlier in this section, parameterisation originally relied on the large separation of scales between resolved and unresolved scales. We have discussed the use of stochastic methods to treat the situation where the scales of physical processes start to approach the model grid scales. As model resolutions improve further the scales of some physical processes start to overlap the grid-scales. For example, kilometre-scale models are now resolving deep convection, although they may also need to parameterize shallow convection. The regime where processes are partially resolved is often referred to as the grey zone. In this regime, it is a particular challenge to represent turbulent transport of heat, moisture and momentum. At the convective scale it is particularly important to properly represent physical processes in order to represent the uncertainties in the initiation of convection, and the formation of fog and low cloud,

The problem is addressed by the WMO “[Grey Zone](#)” project which is coordinated by the Global Atmospheric Systems Studies (GASS) within the WMO World Climate Research Program (WCRP) and the Working Group on Numerical Experimentation (WGNE). Many scientists are contributing to the development of cloud-resolving model simulations that can be coarse-grained and used to provide some ground truth for the development of stochastic parameterizations (e.g. Palmer et al. 2009).

Another international project providing valuable data to support the development of stochastic parameterizations is the Protocol for the Analysis of Land-Surface Models ([PALS](#)), organized by the WMO/WCRP Global Energy and Water Cycle Exchanges Project (GEWEX). This project provides data sets suitable for testing and evaluating land-surface models, and can be leveraged to test land-surface models that incorporate physically based stochastic parameterizations.

The WWRP has set up a new working group on Predictability, Dynamics and Ensemble Forecasting (PDEF). One of the main scientific challenges that the group will address is the representation of model uncertainty using stochastic techniques. The working group will support the WWRP projects, including the three THORPEX legacy projects on Subseasonal to Seasonal prediction (S2S), Polar Prediction Project (PPP) and High-impact Weather (HIWeather). Extending the data available from TIGGE, the S2S project will collect ensemble forecasts out to a range of several months, suitable for

¹ processed through the data assimilation system’s “*forward operator*.”

the intercomparison of operational methods for simulating model uncertainty and the uncertainty associated with coupled-state interactions. The WMO's Polar Prediction Project (PPP) will collect databases of Arctic conditions that can be used for the refinement of stochastic parameterizations in the Arctic.

4. MODEL COMBINATION AND STATISTICAL POST-PROCESSING.

4.1. Background.

Although much effort has been applied to reduce systematic errors in NWP models, given past history they are likely to be large enough to be of consequence for decades to come. Deterministic NWP output may often exhibit systematic mean error, and ensembles may also be under-dispersive. Although the spread of ensembles has been significantly improved for many variables in recent years (e.g., Gagnon et al. 2014), the customers of weather guidance are requesting reliable, unbiased probabilistic guidance right now for all variables. Hence, other methods such as multi-model or multi-centre ensembles and statistical post-processing are commonly applied to improve forecast reliability and skill.

Multi-centre ensemble combinations exhibit improved skill and reliability of the forecasts (Swinbank et al. 2014 and references therein), especially at the larger scales of motion that are within the predictive capacity of these systems. The different systems commonly exhibit varying systematic errors, and hence their combination provides some increased spread and reduction of error through cancellation. Multi-centre combinations are now a regular part of the post-processing used by weather forecasting centres. Indeed, the North American Ensemble Forecast System (NAEFS; Candille, 2009) has been established by the weather services of the USA, Canada and Mexico to provide multi-model ensemble forecast products as an operational counterpart to the TIGGE research project. The benefit of combining predictions from different ensembles also extends to cyclone track predictions (e.g., Yamaguchi et al. 2012). The applications section below continues the discussion of multi-model techniques.

Beyond simple model combination, many statistical post-processing methods may be applied that address the systematic errors of ensemble predictions. The general approach is to adjust current model guidance using relationships between past forecasts and observations/analyses. Many approaches have been proposed in the last several years, including Bayesian Model Averaging (BMA; Raftery et al. 2005, Wilson et al. 2007, Sloughter et al. 2007, Hamill 2007, Fraley et al. 2010) and related techniques (Wang and Bishop 2005, Glahn et al. 2009, Unger et al. 2009), non-homogeneous Gaussian regression (NGR; Gneiting et al. 2005, Hagedorn et al. 2012), logistic, extended logistic, and heteroscedastic extended logistic regression (e.g., Hamill et al.

2008, Wilks 2004, Roulin and Vannitsem 2012, Messner et al. 2014), analog methods (Hamill and Whitaker 2006, delle Monache et al. 2013), and many other methods (e.g., Hamill and Colucci 1998, Eckel and Walters 1998, Cui et al. 2012, Flowerdew 2013, van Schaeybroeck and Vannitsem 2014, Scheuerer 2014, Scheuerer and König 2014). For the calibration of uncommon events such as heavy precipitation or for longer-lead forecasts where the signal is small and errors are large, a large amount of training data may be needed. One approach to provide extra sample sizes is to run many forecasts of historical cases with a current NWP system (often referred to as reforecasts; Hamill et al. 2006, 2013, Hagedorn 2008, Fundel et al. 2010, Fundel and Zappa 2011). Another approach to increasing sample sizes may be to compose training data across many locations (e.g., Charba and Samplatsky 2011ab, Hamill et al. 2008, 2015).

Statistical post-processing methods are commonly applied independently for each forecast point and lead time. Some applications such as hydrological prediction can benefit from additional information on the joint probabilities between many locations, information which can be lost when processing the data independently. One particularly useful approach for providing correlative information may be through “*ensemble copula coupling*” (ECC, Schefzik et al. 2013 and references therein). Figure 2 shows how the ECC technique can be used to restore the spatial structure in ensemble members that have previously been calibrated using BMA. Flowerdew (2013) also used a similar approach to ensure spatial coherence of the calibrated ensemble members as part of his reliability-based ensemble calibration method. A complementary approach where multivariate relationships are set using climatological data is known as the Schaake Shuffle (Clark et al. 2004). Wilks (2014) provided a comparative evaluation of two.

4.2 Underpinning research.

Despite the proliferation of methods for statistical post-processing, it appears that continued research is needed into improved methods. Methods that may be optimal for one forecast problem (e.g., heavy precipitation) may not be optimal for another (e.g., precipitation type forecasting). Hence, continued research into the development of improved algorithms is desired, especially for variables related to high-impact weather (tropical cyclone intensity, precipitation type, tornado probability, calibration of joint probabilities, and so forth). Methods that facilitate exploratory data analysis for the very high-dimensional data common in post-processing would be helpful; we need to identify predictors and classes of methods that will work adequately from dry to moist locations, from tropical to extra-tropical.

In many situations now, the differences in skill between several credible post-processing algorithms may not be as large as the differences in skill for one algorithm for small vs. large training data sets. Extensive reforecast data sets are expensive to compute, so additional research would be helpful to inform how best to construct the reforecasts. For example, which provides the most useful data across a range of post-processing applications: is a ten-member, twice weekly reforecast spanning ten years preferable to a five-member, twice weekly reforecast over 20 years? What about a five-member, four times weekly reforecast spanning ten years? Should reforecast samples be more frequent for shorter-lead forecasts than for longer-lead forecasts, or vice versa? Hamill et al. (2014) provides some guidance behind possible choices, but more research is needed.

Another challenge with reforecasts is that their statistical characteristics should resemble those of the current operational forecast model. Ideally, this would mean that they would be initialized with reanalyses and ensemble initialization methods that were the same as used operationally; the same forecast model at the same resolution, the same data assimilation methodology. Extensive reanalyses may be impractical for every operational centre to compute for each forecast model. This then raises the practical question as to whether the reanalyses from a different modelling system can be used with or without some modification for reforecast initialization. This is a very new area of research and is currently being pursued at the Canadian Meteorological Centre (personal communication, N. Gagnon, 2014) and at Météo-France (personal communication, M. Boisserie).

4.3 Linkages and requirements.

The current literature is replete with the testing of large varieties of methods for statistical post-processing. Unfortunately, there are no standardized test data sets that have been published, so it is often difficult to know whether a proposed new methodology is better than an older one, since they likely were not tested with the same data. The development of some standardized forecast and observation/analysis data sets would be helpful.

Since post-processing is dramatically improved with large samples, should the operational centres embrace the reforecast methodology, they will also need dramatically enlarged disk space to make the data accessible to in-house and external developers. Such costs should be incorporated when soliciting bids for future high-performance computing systems.

Finally, post-processing skill will only be as good as the data used to train the method. High-quality, shared analyses of high-impact variables are needed, including precipitation and precipitation type, surface temperature, winds, and humidity, and so forth.

5. Applications.

5.1 Introduction.

A major motivation for running ensemble prediction systems is to provide better predictions of the risk of high-impact weather. The THORPEX Global Interactive Forecast System (GIFS)-TIGGE working group to foster the development of multi-centre ensemble-based products to support this goal using the TIGGE dataset.

An initial focus of this work was on the developments of experimental products to support tropical cyclone forecasting. Most of the weather prediction centres that participated in TIGGE provided forecasts of tropical cyclone tracks, and in some cases additional data including intensity. Those were exchanged using a new XML-based format (known as CXML, see <http://www.bom.gov.au/cyclone/cxmlinfo/>), to support the THORPEX Pacific Asian Regional Campaign (T-PARC) and subsequent forecast demonstration projects. With the help of TIGGE data, several different types of products have been developed to support forecasts of tropical cyclones. A simple approach is to plot the individual tracks from each ensemble member. Another approach that has been widely adopted is to calculate a “strike probability” map, showing the probability that the cyclone will pass within a set distance (normally 120 km) of any point (van der Grijn et al. 2004). Figure 3 shows an example of this type of plot for typhoon Hagupit that struck the Philippines in early December 2014. Other useful products include ellipses (Hamill et al. 2011) and graphs showing time series of various measures of the forecast cyclone intensities (e.g., central pressure, maximum wind speed, or vorticity, such as Fig. 3 of Hamill et al. 2012).

As noted above, objective verification scores indicate that combining ensembles together is generally beneficial, especially giving improved measures of the uncertainties in cyclone track forecasts, (e.g., Yamaguchi et al. 2012). With the exchange of cyclone forecast information by TIGGE partners, it has been straightforward to use those data to produce strike probabilities and other products based a multi-model grand ensemble. Indeed, the examples shown in Figure 3 are based on the combination of three ensembles.

TIGGE has also been used more recently to highlight the risks of heavy rainfall, strong wind and extreme temperatures. Using TIGGE data, Matsueda and Nakazawa

(2014) developed a prototype suite of ensemble-based early warning products for severe weather events, using both single-model (ECMWF, JMA, NCEP, and Met Office) and multi-model grand ensembles. These products estimate the forecast probability of the occurrence of heavy rainfall, strong winds, and severe high/low temperatures, based on each model's climatology. The procedure attempts to calibrate the products by using the climatological probability density function from each ensemble to determine appropriate thresholds for severe weather events. Objective verification of the products confirms that the combination of four ensembles gives more reliable forecasts than the equivalent product based on a single ensemble. In another initiative developed using TIGGE data, forecasts of humidity are being used to help forecast meningitis outbreaks in the "meningitis belt" south of the Sahara (Hopson, 2014).

With the sophistication of modern NWP systems, unless data is shared, there is now potentially large gap between the forecast information available in the most highly developed countries and that available in less developed countries. Since developing countries are often particularly vulnerable to severe weather events, it is important to address that gap. The WMO has established a Severe Weather Forecast Demonstration Project (SWFDP) that enables some of the less-developed regions of the world to benefit from state of the art numerical predictions. This is achieved by global NWP centres supplying graphical products tailored to support regional SWFDP initiatives. Some of the products developed from the TIGGE data are now being rolled out for use in the SWFDP.

The SWFDP was first established in Southern Africa, and following its early success, the initiative was extended to more countries. The SWFDP was then established in a second region – the South Pacific, and has since been extended to South-east Asia and East Africa. It would be very beneficial to extend it to other regions, as funding allows.

5.2 Underpinning research.

In recent years there has been considerable improvement in the skill of tropical cyclone track forecasts, and ensembles have contributed a lot to quantifying the uncertainties in the track forecasts. However, the skill of forecasts of cyclone intensity has not improved at the same rate. One factor leading to this poor skill could be that global NWP models, in particular those used for ensemble forecasting, are still relatively coarse in resolution compared with the size of storm's inner core. Some recent studies have shown very encouraging results using very high-resolution, limited-area models to simulate the evolution of tropical cyclones (Gall et al. 2013 and references therein). However, the initialization of mesoscale structures is hindered by challenges with data

and assimilation methodologies, including the relative paucity of inner-core data in many circumstances as well as other challenges such as substantial position errors and non-Gaussian error statistics (Chen and Snyder 2007, Geer and Bauer 2011, Nehrkorn et al. 2015; and section 2 of this chapter). Continued research on assimilation methods and high-resolution modelling of tropical cyclones is warranted.

While it has proved feasible to provide high-quality forecasts of tropical cyclones tracks a few days ahead, the next challenge is to look further ahead, which entails forecasting cyclone formation as well as evolution. Recent work using the TIGGE ensembles shows that it is now becoming possible to give probabilistically based forecasts of the likelihood of tropical cyclone formation, and an indication of their future tracks, as shown in Figure 4. There are obvious societal benefits from further improvements to both cyclone intensity forecasts and the extension of the forecast range. Given that tropical cyclone formation is sometimes over- or under-forecast in models, the use of reforecasts to assess the deviations of genesis from their climatological probabilities may be helpful (e.g., Fig. 4 from Hamill et al. 2012).

Ensembles provide a wealth of information, but this information needs to be synthesized into products that are most useful for decision makers. There is still a significant role for social scientists to play in helping meteorologists determine the best ways to convey probabilistic information (e.g., Joslyn and Savelli 2010, Savelli and Joslyn 2013, Novak et al. 2014, Ash et al. 2014).

5.2 Linkages and requirements

The WMO/WWRP High-Impact Weather (HIWeather) project is laying out a research and development agenda that includes the development of applications for using ensemble and multi-model ensemble data. The goal is to

“Promote cooperative international research to achieve a dramatic increase in resilience to high impact weather, worldwide, through improving forecasts for timescales of minutes to two weeks and enhancing their communication and utility in social, economic and environmental applications.”

The reader is encouraged to consider the implementation plan (Jones and Golding, 2014). This program will coordinate the activities of physical scientists and social scientists to address major high-impact weather phenomena, including urban flooding, wildfire, localised extreme wind, disruptive winter weather, and urban heat waves/air pollution. Databases of regional and global weather ensembles and intraseasonal ensembles will continue to be very helpful in the research and development of experimental forecast products.

6. Conclusions

In the past two decades, considerable progress had been made in quantifying uncertainties in weather forecasts using ensemble prediction systems. Ensemble Prediction Systems need represent both the uncertainties in the initial conditions, and how those uncertainties evolve during the course of the forecast. The quantification of initial errors is closely linked to the data assimilation problem. The problem of representing the effect of model errors in a physically reasonable and yet statistically correct manner remains a major challenge, requiring the deployment of sophisticated stochastic modelling techniques. The challenge is magnified as model resolutions increase, so that physical processes are partly (but not fully) resolved – the so-called “grey zone” problem. Stochastic forcing methods will remain a key area for further research and development in the coming decades.

The next part of the challenge is the translation of ensemble output to create probabilistic weather forecasts – and particularly to alert people to the risks of severe weather events. Especially using TIGGE data, there have been considerable developments in recent years on using both statistical methods and the combination ensembles to reduce systematic errors and provide reliable probabilistic forecast products. The communication and application of such probabilistic forecasts is a further challenge, addressed by the User, Applications and Social Science component of this conference.

Acknowledgments

We thank all presenters at the WWRP Open Science Conference, especially those who shared images used in this chapter (T. N. Palmer, M. Yamaguchi, and S. Lang).

References

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83.

Anderson, J. L., 2010: A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.*, **138**, 4186–4198.

Ash, K. D., R. L. Schumann III, and G. C. Bowser, 2014: Tornado warning trade-offs: evaluating choices for visually communicating risk. *Weather, Climate, and Society* **6**, 104-118.

Aspelien, T., T. Iversen, J. B. Bremnes, and I. L. Frogner, 2011: Short-range probabilistic forecasts from the Norwegian limited-area EPS: long-term validation and a polar low study. *Tellus A*, **63**, doi: <http://dx.doi.org/10.3402/tellusa.v63i3.15820>.

Barkmeijer, J., M. van Gijzen, and F. Bouttier, 1998: Singular vectors and estimates of the analysis error covariance metric. *Quart. J. Roy. Meteor. Soc.*, **124**, 1695–1713.

Barkmeijer, J., R. Buizza, and T. N. Palmer, 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2333–2351.

Berner J., G. Shutts, M. Leutbecher, and T.N. Palmer, 2008: "[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.

Berner, J., S.-Y. Ha, J. P. Hacker, A. Fournier, C. Snyder, 2011: [Model uncertainty in a mesoscale ensemble prediction system: stochastic versus multiphysics representations](#). *Mon. Wea. Rev.*, **139**, 1972-1995, doi: <http://dx.doi.org/10.1175/2010MWR3595.1>

Bogenschutz, P. A., and S. K. Krueger, 2013: A simplified PDF parameterization of subgrid-scale clouds and turbulence for cloud-resolving models, *J. Adv. Model. Earth Syst.*, **5**, 195–211. doi:[10.1002/jame.20018](https://doi.org/10.1002/jame.20018).

Bonavita, M., L. Isaksen and E. Hólm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. *Q.J.R. Meteorol. Soc.*, **138**: 1540–1559. doi: [10.1002/qj.1899](https://doi.org/10.1002/qj.1899)

Bougeault, P., and others, 2009: [The THORPEX Interactive Grand Global Ensemble \(TIGGE\)](#). *Bull Amer. Meteor. Soc.*, **91**, 1059-1072.

Bouttier, F., B. Vié, O. Nuissier, and L. Raynaud, 2012: Impact of stochastic physics in a convection-permitting ensemble. *Mon. Wea. Rev.*, **140**, 3706–3721.

doi: <http://dx.doi.org/10.1175/MWR-D-12-00031.1>

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Royal Meteor. Soc.*, **134**, 703–722.

Bowler, N.E. , A. Arribas, S. E. Beare, K. R. Mylne, and G. J. Shutts, 2009: The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Quart. J. Royal Meteor. Soc.*, **135**, 767–776. DOI: 10.1002/qj.394.

Bowler, N. E., and Mylne, K. R., 2009: Ensemble transform Kalman filter perturbations for a regional ensemble prediction system. *Quart. J. Royal Meteor. Soc.*, **135**, 757–766. doi: 10.1002/qj.404

Buizza, R., and T. N. Palmer, 1995: The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 1434-1456.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Royal Meteor. Soc.*, **125**, 2887-2908.

Buizza, R., P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097. doi: <http://dx.doi.org/10.1175/MWR2905.1>

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 420-436.

Candille, G. and O. Talagrand, O., 2008: Impact of observational error on the validation of ensemble prediction systems. *Quart. J. Royal Meteor. Soc.*, **134**, 959–971. doi: 10.1002/qj.268

Candille, G., 2009: The multiensemble approach: the NAEFS example. *Mon. Wea. Rev.*, **137**, 1655-1665. DOI: 10.1175/2008MWR2682.1

Charba, J. P., and F. G. Samplatsky, 2011a: Regionalization in fine-grid GFS MOS 6-h quantitative precipitation forecasts. *Mon. Wea. Rev.*, **139**, 24–38. doi: <http://dx.doi.org/10.1175/2010MWR2926.1>

Charba, J. P., and F. G. Samplatsky, 2011b: High-resolution GFS-based MOS quantitative precipitation forecasts on a 4-km grid. *Mon. Wea. Rev.*, **139**, 39–68. doi: <http://dx.doi.org/10.1175/2010MWR3224.1>

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. doi: <http://dx.doi.org/10.1175/2009MWR3187.1>

Chen, Y., and C. Snyder, 2007: Assimilating vortex position with an ensemble Kalman filter. *Mon. Wea. Rev.*, **135**, 1828–1845. doi: <http://dx.doi.org/10.1175/MWR3351.1>

Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan and R. Wilby, 2004: The Schaake Shuffle: a method for reconstructing space-time variability in forecasting precipitation and temperature fields. *J. Hydrometeor.*, **5**, 243–262.

Clark, A. J., W. A. Gallus Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140.

Clark, A. J., W. A. Gallus, M. Xue, and F. Kong, 2010: Growth of spread in convection-allowing and convection-parameterizing ensembles. *Wea. Forecasting*, **25**, 594–612, doi:10.1175/2009WAF2222318.1.

Craig, G. C., 2014: *Physically based stochastic parameterisation*. [Presentation to WWRP Open Science Conference](#). SCI-PS181.01.

Cui, B., Z. Toth, Y. Zhu, and D. Hou, 2012: Bias correction for global ensemble forecast. *Wea. Forecasting*, **27**, 396–410. doi: <http://dx.doi.org/10.1175/WAF-D-11-00011.1>

DelSole, T., 2011: State and Parameter Estimation in Stochastic Dynamical Models. *ECMWF Proceedings, Workshop on Representing Model Uncertainty and Error in Numerical Weather and Climate Prediction Models*. [\[PDF\]](#)

Delle Monache, L., F. A. Eckel, D. L. Rife, B. Nagarajan, and K. Searight, 2013: Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev.*, **141**, 3498–3516. doi: <http://dx.doi.org/10.1175/MWR-D-12-00281.1>

Doucet, A., N. de Freitas, and N. Gordon, 2001: An introduction to sequential Monte Carlo methods. *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds., Springer-Verlag, 2–14.

Duc, L., K. Saito, and H. Seko, 2013: Spatial–temporal fractions verification for high-resolution ensemble forecasts. *Tellus*, **65A**, 18171, doi:10.3402/tellusa.v65i0.18171.

Eckel, F. A., and M. K. Walters, 1998: Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. *Wea. Forecasting*, **13**, 1132–1147.

ECMWF, 2011: *Proceedings of the Workshop on Representing Model Uncertainty and Error in Numerical Weather and Climate Prediction Models*. Available from <http://old.ecmwf.int/publications/library/do/references/list/201106>

Ehrendorfer, M., and J. J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, **54**, 286–313.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10143-10162.

Flowerdew J., 2013: Calibrating ensemble reliability whilst preserving spatial structure. *Tellus A* **66**, 22662.

Fraley, C., Raftery, A. E., and Gneiting, T., 2010: [Calibrating multi-model forecast ensembles with exchangeable and missing members using Bayesian model averaging](#). *Mon. Wea. Rev.*, **138**, 190-202.

Frenkel, Y., A. J. Majda, and B. Khouider, 2012: Using the stochastic multcloud model to improve tropical convective parameterization: a paradigm example. *J. Atmos. Sci.*, **69**, 1080–1105. doi: <http://dx.doi.org/10.1175/JAS-D-11-0148.1>

Frogner, I.-L., Haakenstad, H., and Iversen, T., 2006: Limited-area ensemble predictions at the Norwegian Meteorological Institute. *Quart. J. Royal Meteor. Soc.*, **132**, 2785–2808.

Fundel, F., Walser A., Liniger, M.A., Frei, C., and Appenzeller, C., 2010. Calibrated precipitation forecasts for a limited-area ensemble forecast system using reforecasts. *Mon. Wea. Rev.*, **138**, 176–189. DOI:10.1175/2009mwr2977.1.

- Fundel, F., and M. Zappa, 2011: Hydrological ensemble forecasting in mesoscale catchments: sensitivity to initial conditions and value of reforecasts. *Water Resour. Res.*, **47**, DOI: 10.1029/2010wr009996.
- Gagnon, N., R. Frenette, M. Charron, S. Beauregarde and A. Erfani, 2014: Are we still lacking spread in medium-range ensemble forecasts? WWOSC presentation.
- Gall, R., J. Franklin, F.D. Marks E.N. Rappaport, and F. Toepfer, 2013: The Hurricane Forecast Improvement Project. *Bull. Amer. Meteor. Soc.*, **94**, 329-343.
- Geer, A. J. and Bauer, P., 2011: Observation errors in all-sky data assimilation. *Quart. J. Royal Meteor. Soc.*, **137**, 2024–2037. doi: 10.1002/qj.830 .
- Glahn, B., M. Peroutka, J. Wiedenfeld, J. Wagner, G. Zylstra, B. Schuknecht, and B. Jackson, 2009: MOS uncertainty estimates in an ensemble framework. *Mon. Wea. Rev.*, **137**, 246–268. doi: <http://dx.doi.org/10.1175/2008MWR2569.1>
- Gneiting, T., Raftery, A. E., Westveld, A. H., and Goldman, T., 2005: [Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation](#). *Mon. Wea. Rev.*, **133**, 1098-1118.
- Gneiting, T., Balabdaoui, F., and Raftery, A. E., 2007: [Probabilistic forecasts, calibration and sharpness](#). *J. Royal Stat. Soc. Series B: Statistical Methodology*, **69**, 243-268.
- Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proc.*, **140**, 107–113.
- Grell, G. A., and S. R. Freitas, 2014: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos. Chem. Phys.*, **14**, 5233-5250.
- Hacker, J. P., J. L. Anderson, and M. Pagowski, 2007: Improved vertical covariance estimates for ensemble-filter assimilation of near-surface observations. *Mon. Wea. Rev.*, **135**, 1021–1036. doi: <http://dx.doi.org/10.1175/MWR3333.1>
- Hacker, J. P., and D. Rostkier-Edelstein, 2007: PBL state estimation with surface observations, a column Model, and an ensemble filter. *Mon. Wea. Rev.*, **135**, 2958–2972. doi: <http://dx.doi.org/10.1175/MWR3443.1>

- Hagedorn, R., 2008: Using the ECMWF reforecast data set to calibrate EPS reforecasts. *ECMWF Newsletter No. 117*, ECMWF, Reading, United Kingdom, 8–13.
- Hagedorn, R., Buizza, R., Hamill, T. M., Leutbecher, M., and T. N. Palmer, 2012: [Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF ensemble forecasts](#). *Quart. J. Royal Meteor. Soc.*, **138**, 1814-1827.
- Hamill, T. M., and S. J. Colucci, 1997: [Verification of Eta/RSM Short-Range Ensemble Forecasts](#). *Mon. Wea. Rev.*, **125**, 1312-1327.
- Hamill, T. M., and S. J. Colucci, 1998: [Evaluation of Eta/RSM ensemble probabilistic precipitation forecasts](#). *Mon. Wea. Rev.*, **126**, 711-724.
- Hamill, T. M., Whitaker, J. S., and C. Snyder, 2001: [Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter](#). *Mon. Wea. Rev.*, **129**, 2776-2790.
- Hamill, T. M., C. Snyder, and R. E. Morss, 2002a: [Analysis-error statistics of a quasigeostrophic model using 3-dimensional variational assimilation](#). *Mon. Wea. Rev.*, **130**, 2777-2790.
- Hamill, T. M., C. Snyder, and J. S. Whitaker, 2002b: [Ensemble forecasts and the properties of flow-dependent analysis-error covariance singular vectors](#). *Mon. Wea. Rev.*, **131**, 1741-1758.
- Hamill, T. M., and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon. Wea. Rev.*, **133**, 3132–3147.
- Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: [Reforecasts, an important dataset for improving weather predictions](#). *Bull. Amer. Meteor. Soc.*, **87**,33-46.
- Hamill, T. M., and J. S. Whitaker, 2006: [Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application](#). *Mon. Wea. Rev.*, **134**, 3209-3229.
- Hamill, T. M., 2006: [Ensemble-based atmospheric data assimilation](#) Chapter 6 of *Predictability of Weather and Climate*, Cambridge Press, 124-156.
- Hamill, T. M., 2007: [Comments on "Calibrated surface temperature forecasts from the Canadian ensemble prediction system using Bayesian Model Averaging](#). *Mon. Wea. Rev.*, **135**, 4226-4230.

Hamill, T. M., R. Hagedorn, and J. S. Whitaker, 2008: [Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: precipitation.](#) *Mon. Wea. Rev.*, **136**, 2620-2632.

Hamill, T. M., and J. S. Whitaker, 2011: [What constrains spread growth in forecasts initialized from ensemble Kalman filters?](#) *Mon. Wea. Rev.*, **139**, 117-131.

Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. J. Benjamin, 2011: [Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter.](#) *Mon. Wea. Rev.*, **139**, 668-688.

Hamill, T. M., M. J. Brennan, B. Brown, M. DeMaria, E. N. Rappaport, and Z. Toth, 2012: NOAA's future ensemble-based hurricane forecast products. *Bull. Amer. Meteor. Soc.*, **93**, 209–220. doi: <http://dx.doi.org/10.1175/2011BAMS3106.1>

Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Jr., Y. Zhu, and W. Lapenta, 2013: [NOAA's second-generation global medium-range ensemble reforecast data set.](#) *Bull Amer. Meteor. Soc.*, **94**, 1553-1565.

Hamill, T. M., and others, 2014: *White paper: A Recommended Reforecast Configuration for the NCEP Global Ensemble Forecast System.* Available at <http://www.esrl.noaa.gov/psd/people/tom.hamill/White-paper-reforecast-configuration.pdf>

Hamill, T. M., M. Scheuerer, and G. T. Bates, 2015: [Analog probabilistic precipitation forecasts using GEFS Reforecasts and Climatology-Calibrated Precipitation Analyses.](#) *Mon. Wea. Rev.*, in review. Also: online [appendix A](#) and [appendix B](#).

Harnisch, F. and C. Keil, 2014: Initial ensemble perturbations provided by convective-scale ensemble data assimilation. WWOSC presentation.

Hohenegger, C., A. Walser, W. Langhans, and C. Schär, 2008: Cloud-resolving ensemble simulations of the August 2005 Alpine flood. *Quart. J. Roy. Meteor. Soc.*, **134**, 889–904, doi:10.1002/qj.252.

Hopson, T., 2014: TIGGE ensemble forecasts with useful skill-spread relationships for African Meningitis and Asia Streamflow forecasting. WWOSC presentation.

Houtekamer, P. L., and J. Derome, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.*, **123**, 2181-2196.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123-137.

Jeong, J.-H., H. W. Linderholm, S.-H. Woo, C. Folland, B.-M. Kim, S.-J. Kim, and D. Chen, 2013: Impacts of snow initialization on subseasonal forecasts of surface air temperature for the cold season. *J. Climate*, **26**, 1956–1972.
doi: <http://dx.doi.org/10.1175/JCLI-D-12-00159.1>

Johnson, A., and X. Wang, 2012: Verification and calibration of neighborhood and object-based probabilistic precipitation forecasts from a multimodel convection-allowing ensemble. *Mon. Wea. Rev.*, **140**, 3054–3077, doi:10.1175/MWR-D-11-00356.1.

Joliffe, I. T., and D. B. Stephenson, 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd Edition, Wiley, 292pp.

Jones, S., and B. Golding, 2014: *HIWeather: A research activity on High-Impact Weather Within the World Weather Research Programme*. Implementation plan available at
http://www.wmo.int/pages/prog/arep/wwrp/new/high_impact_weather_project.html .

Joslyn, S., and S. Savelli, 2010: Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteor. Appl.*, **17**, 180–195.

Juricke, S., H. F. Goessling, and T. Jung, 2014: Potential sea ice predictability and the role of stochastic sea ice strength perturbations, *Geophys. Res. Lett.*, **41**,
doi:10.1002/2014GL062081.

Keane, R. J., G. C. Craig, C. Keil, and G. Zängl, 2014: The Plant–Craig stochastic convection scheme in ICON and its scale adaptivity. *J. Atmos. Sci.*, **71**, 3404–3415.
doi: <http://dx.doi.org/10.1175/JAS-D-13-0331.1>

Lang, S., M. Leutbecher and M. Bonavita, 2014: Defining the initial conditions for medium-range ensemble forecasts with an ensemble of data assimilations. WWOSC presentation. SCI-PS207.01 [Presentation to WMO/WWRP Open Science Conference](#).

Larson, V. E., and J.-C. Golaz, 2005: Using probability density functions to derive consistent closure relationships among higher-order moments. *Mon. Wea. Rev.*, **133**, 1023-1042.

Larson, V. E., D. P. Schanen, M. Wang, M. Ovchinnikov, and S. Ghan, 2012: PDF parameterization of boundary layer clouds in models with horizontal grid spacings from 2 to 16 km. *Mon. Wea. Rev.*, **140**, 285–306. doi: <http://dx.doi.org/10.1175/MWR-D-10-05059.1>

Lawson, W. and Hansen, J., 2004: Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Wea. Rev.*, **132**, 1966–1981.

Lavaysse, C., M. Carrera, S. Bélair, N. Gagnon, R. Frenette, M. Charron, and M. K. Yau, 2013: Impact of surface parameter uncertainties within the Canadian regional ensemble prediction system. *Mon. Wea. Rev.*, **141**, 1506–1526. doi: <http://dx.doi.org/10.1175/MWR-D-11-00354.1>

Li, X., M. Charron, L. Spacek, and G. Candille, 2008: A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbations. *Mon. Wea. Rev.*, **136**, 443–462. doi: <http://dx.doi.org/10.1175/2007MWR2109.1>

Lin, J.W.-B., and J. D. Neelin, 2000: Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophys. Res. Lett.*, **27**, 3691-3694.

Lin, J.W.-B., and J. D. Neelin, 2003: Toward stochastic deep convective parameterization in general circulation models. *Geophys. Res. Lett.*, **30**, 1162, doi:10.1029/2002GL016203.

Majda, A. J., 2007: Multiscale models with moisture and systematic strategies for superparameterization. *J. Atmos. Sci.*, **64**, 2726–2734. doi: <http://dx.doi.org/10.1175/JAS3976.1>

Matsueda, M. and T. Nakazawa, 2014: Early warning products for severe weather events derived from operational medium-range ensemble forecasts. *Meteor. Apps.*, **119**, doi: 10.1002/met.1444

McClay, J. G., M. K. Flatau, C. A. Reynolds, J. Cummings, T. Hogan, and P. J. Flatau, 2012: Inclusion of sea-surface temperature variation in the U.S. Navy ensemble-

transform global ensemble prediction system. *J. Geophys. Res.*, **117**, D19120, doi:10.1029/2011JD016937.

Metref, S., Cosme, E., Snyder, C., and Brasseur, P., 2014: A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation, *Nonlin. Processes Geophys.*, **21**, 869-885, doi:10.5194/npg-21-869-2014.

Messner, J.W., G. J. Mayr, A. Zeileis, and D. S. Wilks, 2014: Heteroscedastic extended logistic regression for postprocessing of ensemble guidance. *Mon. Wea. Rev.*, **142**, 448-456.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 2791–2808.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroligis, 1996: The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Royal. Meteor. Soc.*, **122**, 73-119.

Nastrom, G. D., and Gage, K. S., 1985: A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.*, **42**, 950-960.

Nehrkorn, T., B. K. Woods, R. N. Hoffman, and T. Auligne, 2015: Correcting for position errors in variational data assimilation. *Mon. Wea. Rev.*, conditionally accepted. Available from TNehrkor@aer.com.

Novak, D. R., K. F. Brill, and W. A. Hogsett, 2014: Using percentiles to communicate snowfall uncertainty. *Wea. Forecasting* **29**, 1259-1265.

Nutter, P., M. Xue, and D. J. Stensrud, 2004: Application of lateral boundary condition perturbations to help restore dispersion in limited-area ensemble forecasts. *Mon. Wea. Rev.*, **132**, 2378–2390.

Paccagnella, T., J. Hacker, C. Marsigli, A. Montani, F. Pappenberger, D. Parsons, R. Swinbank, and Z. Toth, 2011: *THORPEX Interactive Grand Global Ensemble Limited Area Model Plan (TIGGE LAM)*. WMO THOREX No. 17, 39 pp. Available from http://www.wmo.int/pages/prog/arep/wwrp/new/documents/THORPEX_17_TIGGE_LAM

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model

uncertainty. *ECMWF Technical Memorandum 598*, 42 pp. Available from http://old.ecmwf.int/publications/library/ecpublications/_pdf/tm/501-600/tm598.pdf.pdf

Palmer, T. N., 2012: Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction. *Quart. J. Royal Meteor. Soc.*, **138**, 841–861. doi: 10.1002/qj.1923

Palmer, T.N., 2014: Strategic goals for NWP centres: minimising RMS error or maximising forecast reliability? SCI-PS172.01, [Presentation to WMO/WWRP Open Science Conference 2014](#).

Peters, K., C. Jakob, L. Davies, B. Khouider, and A. J. Majda, 2013: Stochastic behavior of tropical convection in observations and a multcloud model. *J. Atmos. Sci.*, **70**, 3556–3575. doi: <http://dx.doi.org/10.1175/JAS-D-13-031.1>

Plant, R. S., and G.C. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87-105.

Posselt, D. J., and T. Vukicevic, 2010: Robust characterization of model physics uncertainty for simulations of deep moist convection. *Mon. Wea. Rev.*, **138**, 1513–1535.

Raftery, A., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155-1174.

Ravela, S., K. Emanuel, and D. McLaughlin, 2007: [Data assimilation by field alignment](#). *Physica D*, **230**, 127-145.

Romine, G. S., C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.*, **142**, 4519–4541. doi: <http://dx.doi.org/10.1175/MWR-D-14-00100.1>

Roulin, E., and S. Vannitsem, 2012: Postprocessing of ensemble precipitation predictions with extended logistic regression based on hindcasts. *Mon. Wea. Rev.*, **140**, 874–888. doi: <http://dx.doi.org/10.1175/MWR-D-11-00062.1>

Savelli, S., and S. Joslyn, 2013: The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied Cognitive Psychology*, **27**, 527-541, doi: 10.1002/acp.v27.4.

Scheuerer, M., and T. M. Hamill, 2014: [Variogram-based proper scoring rules for probabilistic forecasts of two multivariate quantities](#). *Mon. Wea. Rev.*, submitted.

Schefzik, R., T.L. Thorarinsdottir, and T. Gneiting, 2013: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, **28** (4), 616-640

Scheuerer, M., 2014: Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quart. J. Royal Meteor. Soc.*, **140**, 1086-1096.

Scheuerer, M. and G. König, 2014: Gridded locally calibrated, probabilistic temperature forecasts based on ensemble model output statistics. *Quart. J. Royal Meteor. Soc.*, doi: 10.1002/qj.2323

Schwartz, C. S., and coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280, doi:10.1175/2009WAF2222267.1.

Schwartz, C. S., G. S. Romine, K. R. Smith, and M. L. Weisman, 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Wea. Forecasting*, **29**, 1295–1318. doi: <http://dx.doi.org/10.1175/WAF-D-13-00145.1>

Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Royal Meteor. Soc.*, **131**, 3079-3102.

Sloughter, J. M., Raftery, A. E., Gneiting, T., and Fraley, C., 2007: [Probabilistic quantitative precipitation forecasting using Bayesian model averaging](#). *Mon. Wea. Rev.*, **135**, 3209-3220.

Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J., 2008: Obstacles to high-dimensional particle filtering, *Mon. Wea. Rev.*, **136**, 4629–4640.

Sutton, C. J., T. M. Hamill, and T. T. Warner, 2006: [Will Perturbing Soil Moisture Improve Warm-Season Ensemble Forecasts? A Proof of Concept](#). *Mon. Wea. Rev.*, **134**, 3174-3189.

Swinbank, R., and others, 2014: [The THORPEX Interactive Grand Global Ensemble \(TIGGE\) and its achievements](#). *Bull. Amer. Meteor. Soc.*, conditionally accepted.

Teixeira, J., and C. A. Reynolds, 2008: Stochastic nature of physical parameterizations in ensemble prediction: a stochastic convection approach. *Mon. Wea. Rev.*, **136**, 483-496.

Tennant, W. and Beare, S., 2014: New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS. *Quart. J. Royal Meteor. Soc.*, **140**, 1150–1160. doi: 10.1002/qj.2202

Tompkins, A. M., and J. Berner, 2008, "[A stochastic convective approach to account for model uncertainty due to unresolved humidity variability](#)", *J. Geophys. Res.*, **113**, D18101, DOI:10.1029/2007JD009284.

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490–2502. doi: <http://dx.doi.org/10.1175/MWR3187.1>

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317-2330.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

Ueno, G., Higuchi, T., Kagimoto, T. and Hirose, N., 2010: Maximum likelihood estimation of error covariances in ensemble-based filters and its application to a coupled atmosphere–ocean model. *Quart. J. Roy. Meteor. Soc.*, **136**, 1316–1343. doi: 10.1002/qj.654 .

Unger, D. A., H. van den Dool, E. O’Lenic, and D. Collins, 2009: Ensemble regression. *Mon. Wea. Rev.*, **137**, 2365–2379. doi: <http://dx.doi.org/10.1175/2008MWR2605.1>

Van der Grijn, G, J.E. Paulsen, F. Lalauette and M. Leutbecher, 2004, ‘Early medium-range forecasts of tropical cyclones’. *ECMWF newsletter*, **102**, 7-14

Van Lier-Walqui, M., T. Vukicevic, and D. J. Posselt, 2012: Quantification of cloud microphysical parameterization uncertainty using radar reflectivity. *Mon. Wea. Rev.*, **140**, 3442–3466. doi: <http://dx.doi.org/10.1175/MWR-D-11-00216.1>

Van Schaeybroeck, B., and S. Vannitsem, 2014: Ensemble post-processing using member-by-member approaches: theoretical aspects. *Quart. J. Royal Meteor. Soc.*, doi: 10.1002/qj.2397 .

Wang, Y. M. Bellus, J.-F. Geleyn, X. Ma, W. Tian, and F. Weidle, 2014: A new method for generating initial condition perturbations in a regional ensemble prediction system: blending. *Mon. Wea. Rev.*, **142**, 2043-2059.

Wang, X., and C.H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.* **60**: 1140–1158.

Wang, X. and Bishop, C. H., 2005: Improvement of ensemble reliability with a new dressing kernel. *Quart. J. Royal Meteor. Soc.*, **131**, 965–986. doi: 10.1256/qj.04.120

Wang, X., D.M. Barker, C. Snyder, T.M. Hamill, 2008a: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Weather Rev.* **136**: 5116–5131.

Wang, X., D.M. Barker, C. Snyder, T.M. Hamill, 2008b: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiments. *Mon. Weather Rev.* **136**: 5132–5147.

Warner, T.T., R.A. Peterson, and R.E. Treadon, 1997: A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **78**, 2599-2617, DOI: [10.1175/1520-0477\(1997\)078<2599:ATOLBC>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2599:ATOLBC>2.0.CO;2).

Wei, M., Toth, Z., Wobus, R. and Zhu, Y., 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus A*, **60**, 62–79. doi: 10.1111/j.1600-0870.2007.00273.x

Whitaker, J. S., and T. M. Hamill, 2012: [Evaluating methods to account for system errors in ensemble data assimilation](#). *Mon. Wea. Rev.*, **140**, 3078-3089.

Wilks, D. S., 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. *Met. Apps*, **16**, 361–368. doi: 10.1002/met.134

Wilks D., 2011: *Statistical Methods in the Atmospheric Sciences*, 3rd edition, Academic Press, 704 pp.

Wilks, D. S., 2014: Multivariate ensemble Model Output Statistics using empirical copulas. *Quart. J. Royal Meteor. Soc.* doi: 10.1002/qj.2414

Wilson, L. J., S. Beauregard, A. E. Raftery, and R. Verret, 2007: Calibrated surface temperature forecasts from the Canadian ensemble prediction system using Bayesian Model Averaging. *Mon. Wea. Rev.*, **135**, 1364–1385.

doi: <http://dx.doi.org/10.1175/MWR3347.1>

WMO, 2013: *Subseasonal to seasonal prediction research implementation plan*. 63 pp. Available at

http://www.wmo.int/pages/prog/arep/wwrp/new/documents/S2S_Implem_plan_en.pdf

Yamaguchi, M., T. Nakazawa, and S. Hoshino, 2012: On the relative benefits of a multi-centre grand ensemble for tropical cyclone track prediction in the western North Pacific. *Quart. J. Roy. Meteor. Soc.*, **138**, 2019–2029.. doi: 10.1002/qj.1937.

Yamaguchi, M., 2014: Multi-model ensemble forecasts of tropical cyclones using TIGGE. [Presentation to WWRP Open Science Conference](#). SCI-PS244.01.

Yang, S.-C, C. Keppenne, M. Rienecker, and E. Kalnay, 2009: Application of coupled bred vectors to seasonal-to-interannual forecasting and ocean data assimilation. *J. Climate*, **20**, 2850–2870.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253.

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.*, **83**, 73–83.

Tables and Figures

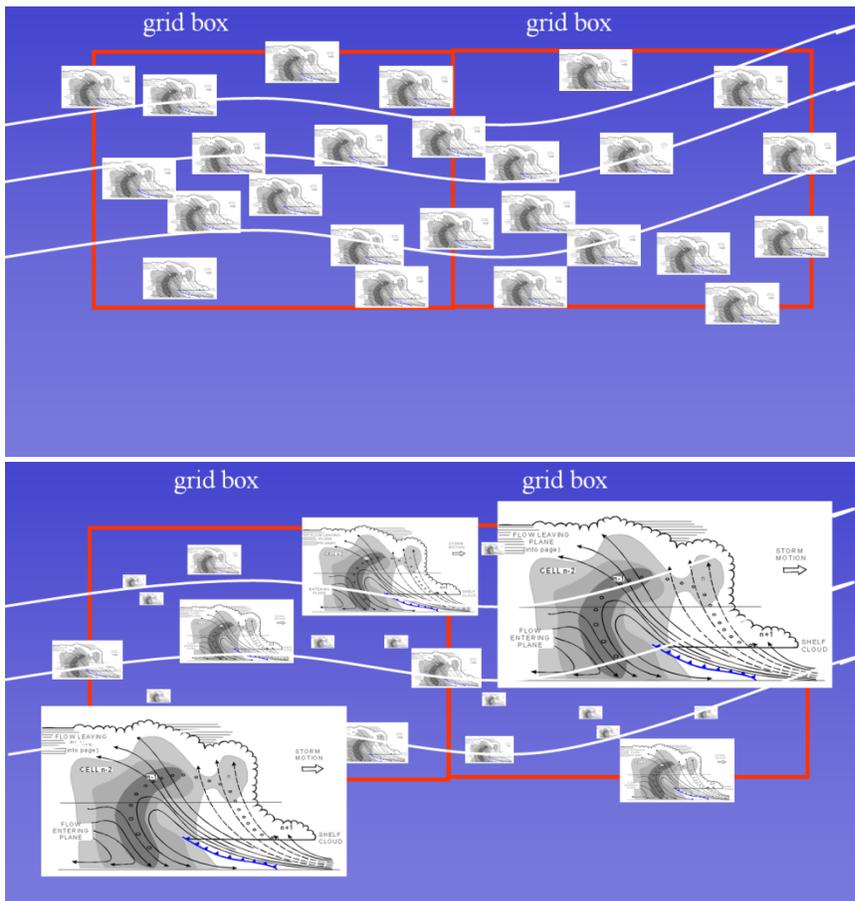


Figure 1. a) Schematic showing clear scale separation between resolved flow and sub grid-scale convection. b) schematic of a more realistic situation where there is no scale separation. From Palmer (2012).

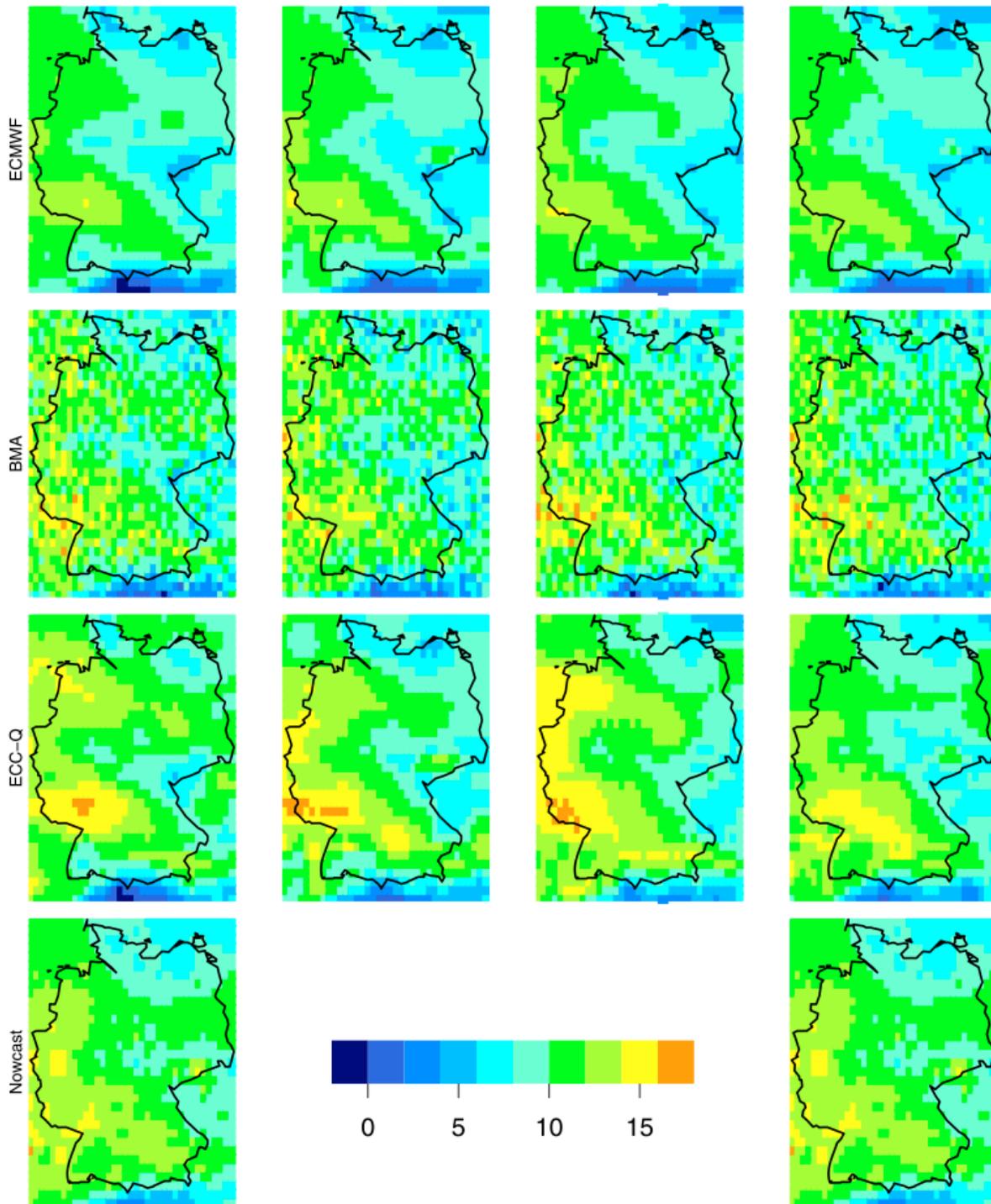


Figure 2. Demonstration of the ability of ECC to recover spatial structure in calibrated ensembles. Top row shows raw temperature forecasts (degrees C) from different ECMWF ensemble members. Second row shows fields generated after ECMWF forecasts were post-processed with BMA. Draws for each grid point represent a random selection from the PDF. Third row shows the samples of random forecasts after ECC processing. The bottom picture

shows the corresponding nowcast temperature field. Figure reprinted from Schefzik et al., (2013).

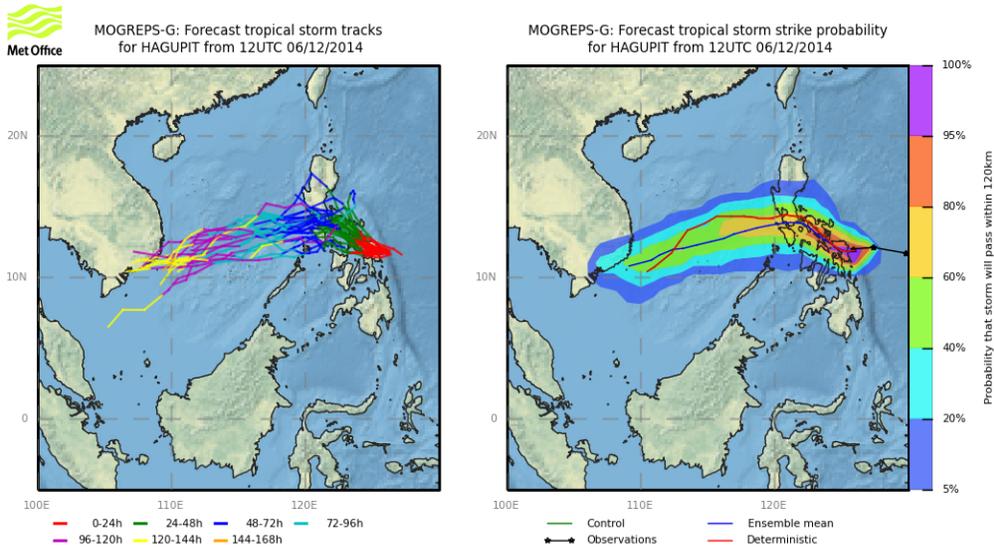


Figure 3: Example of tropical cyclone forecast products for a forecast of typhoon Hagupit from the initial time of 12 UTC 6 December 2014. Left plot: individual track forecasts from each ensemble member. Right plot: summary plot showing strike probability key tracks.

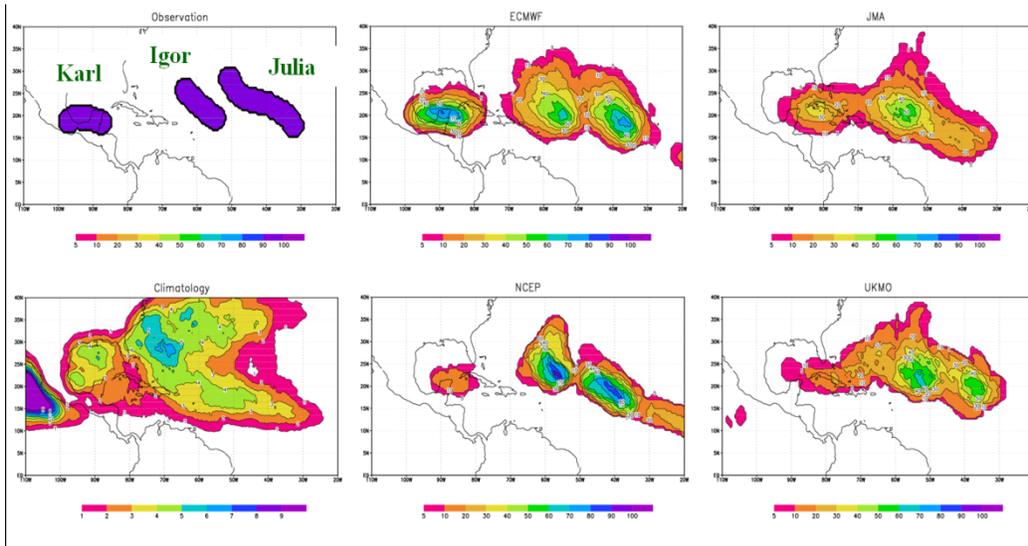


Figure 4. Ensemble forecasts of tropical cyclone activity at a range of 6-9 days, from initial conditions at 12UTC on 9th September 2010. Three tropical cyclones occurred in this period: hurricane Igor formed about a day before the start of forecast, Julia about 3 days after the start and Karl about 5 days afterwards. Top left plot: areas affected by the three hurricanes. Bottom left plot: climatological hurricane activity. Remaining plots: forecasts of activity from four ensemble systems, ECMWF, JMA, NCEP and UKMO. From Yamaguchi (2014).