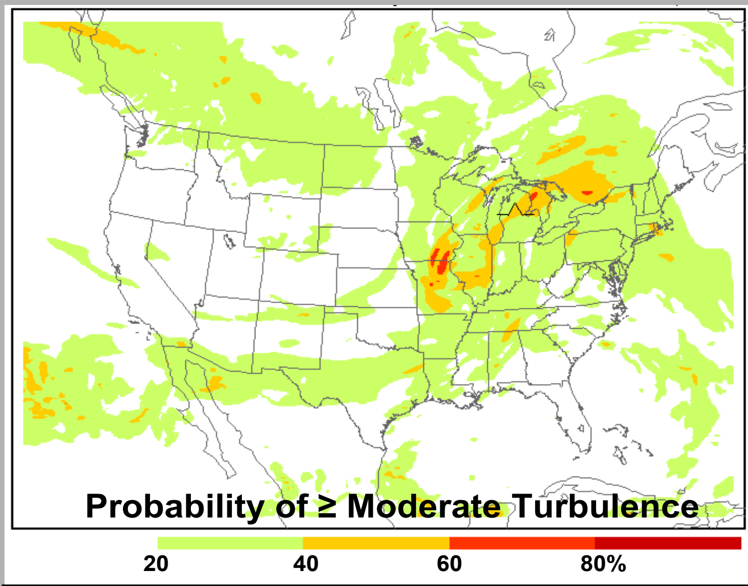
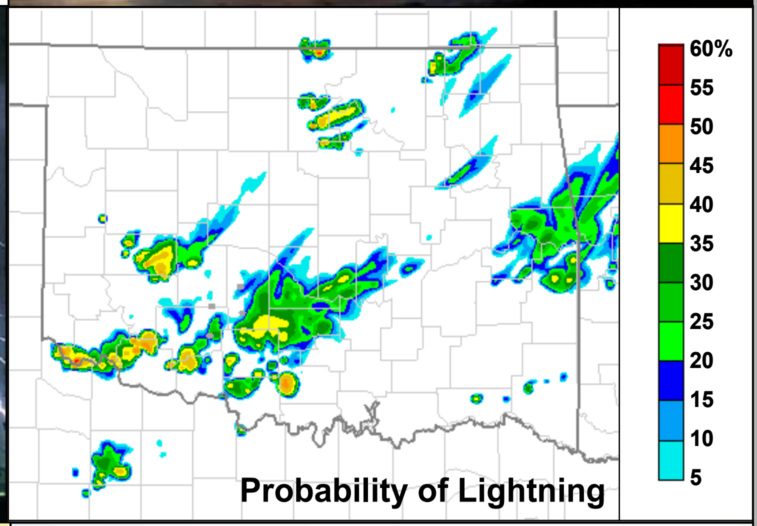
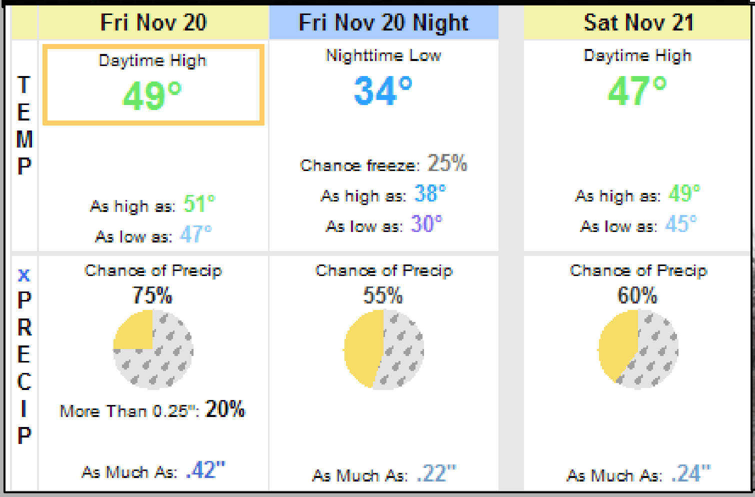
National Mesoscale Probabilistic Prediction: Status and the Way Forward

A White-Paper Report from the *National Workshop on Mesoscale Probabilistic Prediction*, 23-24 September 2009







May 28, 2010

Contents

Authors

F. Anthony Eckel, NWS Office of Science and Technology

Harry R. Glahn, NWS Office of Science and Technology

Thomas M. Hamill, NOAA Earth System Research Lab

Susan Joslyn, University of Washington

William M. Lapenta, NWS National Centers for Environmental Prediction

Clifford F. Mass, University of Washington

Foreword

United States citizens, government decision makers, businesspersons, and industry desire higher quality weather information. The United States has some of the most dangerous weather in the world and 30% of the gross national product is weather sensitive. We have the knowledge and technology to improve US global competitiveness by producing and applying skilled probability forecasts. Such high utility weather information will attract industry to develop and implement innovative methods for integrating weather into business decisions, potentially saving the US economy tens of billions a year in the aviation and renewable energy sector alone.

Fortunately, leaders and policy makers within the weather enterprise are laying groundwork for a concerted, national effort to synchronize national resources. The Hurricane Forecast Improvement Program (HFIP) has bred energy and teamwork into solving difficult scientific challenges associated with improving tropical storm track and intensity forecasts. Lessons learned through HFIP have rallied the data assimilation community around the 4D-VAR/Ensemble Kalman Filter hybrid concept. The National Center for Environmental Prediction (NCEP), in concert with NASA, is currently revamping the data assimilation strategy. HFIP also funded testing of a suite of advanced global models that has sparked a NOAA/Navy partnership dedicated to fielding the next generation global model. NCEP, in collaboration with Environment Canada and the National Meteorological Service of Mexico, introduced the first operational global multi-center ensemble, the North American Ensemble Forecast System (NAEFS). These efforts have lead to the National Unified Operational Prediction Capability (NUOPC) initiative between the DOD and NOAA that seeks to improve forecast accuracy and provide reliable probabilistic information for effective global military operations and civilian aviation. Additionally, today, NOAA researchers are working to operationalize the High Resolution Rapid Refresh forecast system, which is vital to improving 0-6 hour high-impact weather prediction.

One key area that still needs to be addressed is production of highly skilled probability forecasts on the mesoscale. Over the last two decades, the science and technology of ensemble-based probabilistic weather prediction has rapidly advanced, providing improved decision input to a wide range of users. The National Research Council (2006) and the American Meteorological Society (2008) clearly described that probabilistic hydrometeorological information is extraordinarily valuable to society, with the potential to bring substantial economic benefits and improved protection of life and property. Emerging societal needs, such as renewable energy systems, require a robust mesoscale probabilistic weather prediction capability. Additionally, NOAA is tasked to produce a 4D datacube with probabilistic forecast information for the Next Generation Air Transportation System (NextGen) by 2016. Current US operational probabilistic prediction is executed at relatively coarse resolution and uses cost efficient methodologies to make best use of limited computing resources. Advancement is also hampered by under-utilization of the considerable knowledge, talent, and experience in this field due to challenges associated with transitioning science from across the weather enterprise into operations. The end result of the current situation is suboptimal decision making across a broad spectrum of weather information users, including government agencies, private businesses, and even individual citizens.

Reflecting both the compelling national requirements for probabilistic prediction and the need to more effectively harness the efforts of the US research and operational communities, a

National Workshop on Mesoscale Probabilistic Prediction was held at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado, on 23-24 September 2009, sponsored by NCAR and the National Weather Service (NWS). The purpose of this meeting was to examine the current state of mesoscale probabilistic prediction in the US and discuss how to expedite progress. Nearly one hundred attendees from government, industry, and academia participated, including many leaders in probabilistic forecasting. The meeting reviewed current efforts, and then participants engaged in break-out discussions of the key challenges and recommendations for future directions. There was agreement that the transition to probabilistic prediction and application is a vital, but formidable task that extends beyond the generation of uncertainty information to issues such as communication and user education. This white paper presents proposals for achieving the following goal: ***By 2015, the US will implement a radically upgraded national capability for mesoscale probabilistic prediction to support current and future decision-making needs, helping return the US to a world-leadership role in numerical weather prediction.***

Reaching this goal is only possible with an inclusive, collaborative, enterprise-wide R&D effort. NOAA is taking the concept of inclusiveness seriously, as plans are underway to expand the Developmental Testbed Center's capability to test, validate, and maintain code for future cutting-edge ensemble techniques. Of course, success requires not only good science but also resources. Hence, success is predicated on development of a good business case to justify the significant R&D and high performance computing resources needed to meet the stated goal. Lastly, the recommendations herein are intended as complementary to the AMS Ad-hoc Committee on Uncertainty Forecasts (AMS 2010) activity.

Don Berchoff
Director, NWS Office of Science and Technology

Louis Uccellini
Director, National Centers for Environmental Prediction

Steve Koch
Director, ESRL Global Systems Division

Discussion

Effective production and application of mesoscale probabilistic forecasts requires a high-quality *ensemble prediction system* (EPS). When considering EPS design, attention is normally focused on aspects such as the number of members, ensemble initial conditions, and treatment of model uncertainty. A more broad view of an EPS, as presented in Figure 1, considers everything necessary to meet the objective of supporting optimal decision making by the user. The many parts of the system are intertwined and must work together to produce the best result.

The term *foundation* here is used to depict the traditional, deterministic (i.e., single value) numerical weather prediction (NWP) forecast system. It is actually the most critical component of the EPS since the foundation's quality sets the level of deterministic forecast error (or amount of forecast uncertainty), which ultimately dictates the value of the EPS output (Buizza et al. 2005). A better analysis (denser observations, better data assimilation, etc.) and/or better model (higher resolution, better physics, etc.) leads to less error.

The *production* encompasses everything necessary to generate consistently reliable information on the possible future states of the atmosphere. This includes what is typically called the ensemble [i.e., multiple NWP forecasts using various initial conditions (ICs) and model formulations] as well as postprocessing to calibrate for model biases and limitations of the ensemble. Continuing from above, a better EPS foundation means that there is less forecast uncertainty, which leads to production of sharper forecasts (i.e., probabilities closer to 0 or 100%) of greater value in decision making.

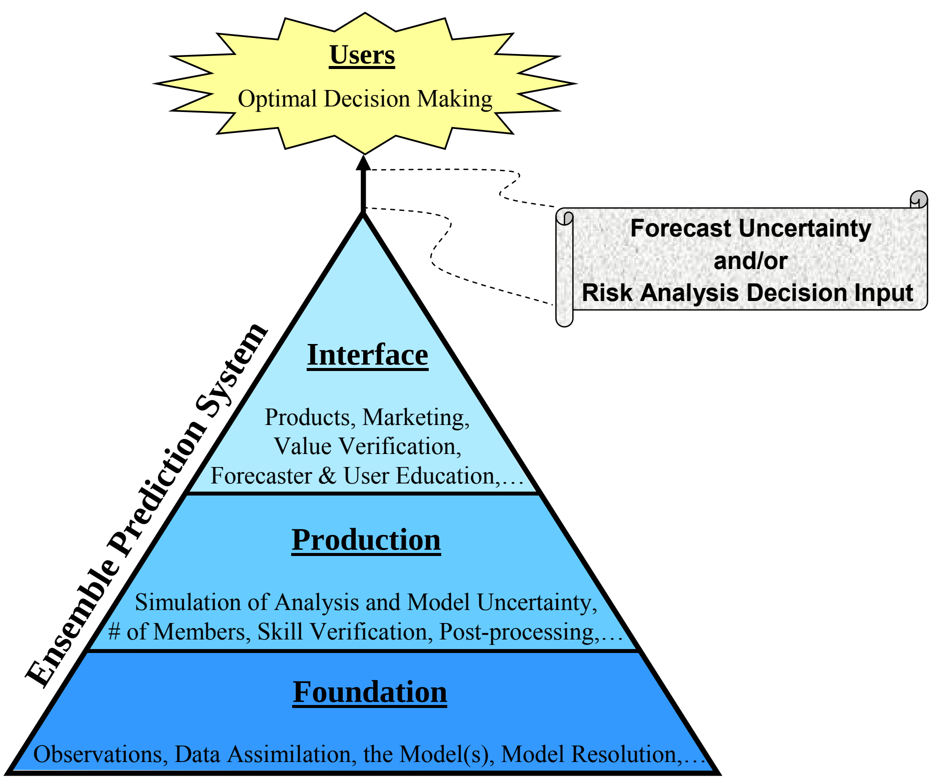


Figure 1. Schematic of an ensemble prediction system and its focus on supporting user decision making.

Lastly, the *interface* of the EPS is where the forecast information gets interpreted, tailored, and communicated to support optimal decision making via risk analysis for users sensitive to weather conditions. The final output, which may come in a variety of forms from a probability forecast to a weather warning based on a user's risk tolerance, encapsulates a flow-dependent estimate of the distribution of likely future states.

The EPS is only as good as its weakest part. For instance, we can put tremendous effort into designing a highly accurate analysis/model and an ensemble that thoroughly simulates the existing uncertainty, but users would not be able to apply the information without products tailored to their operations and education on how to use the information. Furthermore, the parts of the system have to be designed and configured to work well together. For example, the postprocessing technique must be tuned to correct the specific systematic errors in both the foundation and the ensemble.

A philosophical topic brought up at the workshop was the need to shift the paradigm of what we call and think of as THE forecast. When discussing ensemble output, even proponents of ensembles often refer to a single forecast and its uncertainty, thus failing to embrace the true nature of forecasting. The scientifically proper, and best way to think of *the forecast* for decision making, is as the complete probability density function (PDF) for the forecast variable, or as a probability when considering a specific event (e.g., visibility $\leq$ 500 m). Shifting toward such thinking helps focus and frame ideas when considering EPS design. For example, accounting for uncertainty in a parameterization should not be approached as perturbations about a "best-guess" value, but rather as thorough representation of possible parameter values.

The following sections cover the workshop's four break-out groups, which discussed key aspects of the status of EPS capabilities and recommendations on the way forward.

A. The Ensemble

This section covers the scientific and technical issues involved in producing the raw ensemble forecast data. While the workshop focused on short-range, regional ensemble forecasting over the US, discussion extended to global ensemble forecasting upon which the regional ensemble is dependent for lateral boundary condition (LBC) updates.

1. Status

a) Current Capabilities

Over the past decade, the National Centers for Environmental Prediction (NCEP) has made notable advancements in operational ensemble forecast production (Toth and Kalnay 1997; Toth et al. 2001; Wei et al. 2008), but progress has been constrained by limited resources and inadequate collaboration within the enterprise. Both the Global Ensemble Forecast System (GEFS) and the Short Range Ensemble Forecast (SREF) are coarse in resolution relative to international peers and must use cost effective, proven methodologies. Table 1 outlines some of today's leading operational ensembles. Increasing resolution of NCEP's ensembles to a comparable level is an expensive undertaking due to the computing demand, particularly for regional modeling since NWS requires a large area of coverage (i.e., CONUS and adjacent coastal regions).

GEFS currently includes 21 members at T190 (~75 km grid spacing) out to 384 h with four cycles per day. GEFS uses a form of the ensemble transform technique called the Ensemble Transform with Rescaling (Wei et al. 2008) to generate ICs and a stochastic

Table 1. Brief details on the current premier operational ensembles. Further details on the design of these ensembles are given in text below. Approximate equivalent grid spacing for the spectral models is calculated at 45° latitude and assumes the smallest resolvable feature covers two grid lengths.

Name	Center	Domain	Grid Spacing	Forecast Length (h)	# of Members	Future Plans
			Regional Ensembles			
MOGREPS	UK Met (United Kingdom)	N. Atlantic and Europe	24 km	54	24	**2010**→ 18 km main domain **2012**→ 1.5 km UK domain
COSMO-LEPS	COSMO Consortium (led by Italy)	Europe	7 km	132	16	**2010**→ 20 members
SRNWP-PEPS	DWD (Germany)	Europe	Variable (avg. ~7 km)	48	10-23	
			Global Ensembles			
ECMWF EPS	ECMWF	Global	T639 (~22 km)	360	51	**2015**→ T1000 (~14 km)
JMA EPS	JMA (Japan)	Global	T319 (~44 km)	216	51	

physics technique to simulate model uncertainty (Hou et al. 2006). SREF runs 21 members with a model grid of 32-35 km, covering a large North American domain out to 87 h with four cycles per day. To generate ICs, SREF currently relies on the regional breeding (for 11 members) and Global Ensemble Transform (for 10 members) techniques. Model uncertainty is simulated with the multi-model technique, employing both different models (i.e., WRF-NMM, WRF-ARW, Eta and RSM) and different physics packages (Du et al. 2009).

b) Design Factors and Computing Requirements

Ensemble design factors that dominate the computer resource demand for an operational EPS include:

- Model resolution (horizontal and vertical)
- Number of members
- Method of initializing the ensemble
- Method(s) for simulating model uncertainty during the forecast integration
- Forecast length
- Domain size
- Domain interaction (1-way or 2-way nesting)
- Forecast update frequency
- Forecast timeliness

Computational resources will always be a strong constraint in ensemble design. Balancing the effort devoted to each factor should, of course, be accomplished in a way that best serves the users' decision-making needs. For instance, the expense of finer model resolution could be compensated for by greatly reducing the number of members, but any gains might be overcome by the degradation to skill from undersampling.

A key advantage of the ensembles in Table 1 is high model resolution. Among the major influences on ensemble forecast quality (e.g., number of members, correlation between ICs and analysis-error statistics, and the appropriate treatment of model uncertainties), model resolution plays a very large role, particularly for sensible weather phenomena. An ensemble can only portray forecast uncertainty information on scales represented in the model. Decreasing ensemble members' grid spacing increases spread via increased interactions between a greater number of scales of motion and also decreases error (i.e., lower forecast uncertainty), thus improving statistical consistency (Szunyogh and Toth 2002; Eckel and Mass 2005; Clark et al. 2009). Statistical postprocessing can ameliorate some ensemble deficiencies, but the evidence suggests that improving the raw ensemble forecast, especially its resolution, can contribute greatly to higher forecast skill even after postprocessing (Hagedorn et al. 2008; Hamill et al. 2008).

Another key design factor is the number of members. The general consensus is that consistently skillful probability forecasts requires 20-30 members. Talagrand et al. (1999) showed that there is little benefit in going beyond ~ 30 members when evaluating the ensemble with common probabilistic metrics. However, in the absence of postprocessing, reliably estimating the probability of rare events (i.e., forecast scenarios in the tails of the forecast PDF) may require a greater number of samples. The difficulty is in verifying (and calibrating) for such events because of their rarity, so spending the large additional cost to ensure sampling of rare events may not be easily justifiable.

c) *Accounting for Analysis and Model Uncertainty*

Of all the ensemble challenges, the most progress has been made on methods for generating ensemble ICs. Early methods such as singular vectors (Buizza 1997) and bred modes (Toth and Kalnay 1993, 1997) provided sets of initial conditions that grew rapidly. However, the underlying theory of ensemble prediction indicates that the ICs should be sampled from the distribution of analysis uncertainty (Ehrendorfer and Tribbia 1997). As implemented, neither bred modes nor singular vectors satisfy this criterion.

The Ensemble Kalman filter (EnKF) technique, a data assimilation method alternative to variational approaches, does provide samples of analysis uncertainty (Wang and Bishop 2003; Descamps and Talagrand 2007) and is now used for ensemble initialization at the Canadian Meteorological Center. Methods based on the Ensemble Transform (ET), a generalization of the breeding technique, can also sample actual analysis error but with higher computationally efficiency (Bishop and Toth 1999; Wei et al. 2008). Approximations such as the Local Ensemble Transform Kalman Filter (Kuhl et al. 2007) are also used operationally, e.g., at the UK Met Office (Bowler et al. 2008). While there are still some remaining issues with these methods, such as the potential for diminished spread growth for EnKF (Hamill and Whitaker 2009), this technology is likely to be more broadly adopted in the coming years given its theoretical and practical appeal.

Accounting for forecast uncertainty due to model error is still a significant challenge mainly because model error is so poorly understood. The inability to thoroughly account for model error in an ensemble is likely a major contributor to the commonly observed problem of underdispersion among ensemble members. Nonetheless, several techniques are currently in use and/or being researched:

1) **Multi-model** – Different models and/or different physics schemes among the members (e.g., Stensrud et al. 2000).
2) **Stochastic Physics** – Perturbations (which may be formulated with spatial/temporal structures or other dependencies) to state variables' tendency during model integration (e.g., Buizza et al. 1999).
3) **Stochastic Backscatter** – Return dissipated energy via scale-dependent perturbations to wind field (e.g., Berner et al. 2009).
4) **Random Parameters** – Random perturbations to physics parameters (e.g., entrainment rate), which may be fixed prior to model integration or varied during model integration (e.g., Bowler et al. 2008).
5) **Perturbed Surface Parameters** – Perturbations to surface temperature, albedo, roughness length, etc., which may be fixed before model integration or varied during model integration (e.g., Eckel and Mass 2005).
6) **Stochastic Parameterizations** – Explicit modeling of the stochastic nature of subgrid-scale processes (e.g., Teixeira and Reynolds 2008).
7) **Coupling to Ocean/LSM Ensemble** – Explicit modeling of the considerable uncertainty from the surface boundary (e.g., Holt et al. 2009).

For the most part, these techniques do not have the theoretical underpinning that exists with the new EnKF methods for generating ICs. The general approach has been to find logical ways to add "good spread" – an increase in spread accompanied by an increase in forecast skill. The costs involved to apply each technique also vary considerably. Coupling to an ocean or LSM ensemble has a very high, real-time computational cost. Techniques such as random parameters have research costs associated with determining appropriate perturbations. The multi-model technique involves maintenance of a large amount of code, which may make it impractical for an operational EPS. Additionally, it can create non-physical clustering of forecasts due to similarities between members, which may result in unrealistic forecast PDFs even after calibration (which itself becomes more challenging as each member has unique systemic errors). The state-of-the-art ensembles, ECMWF EPS for global modeling and MOGREPS for regional modeling, do not use the multi-model technique, which suggests that we consider concentrating our efforts into a single model configuration.

A separate source of model uncertainty unique to regional ensembles is from the use of periodic lateral boundary condition (LBC) updates. It is difficult for the ensemble to represent the errors from coarse (spatial and temporal) LBC updates. Nutter et al. (2004a) showed that even when using a global ensemble to drive a regional ensemble, the perturbed LBCs fail to capture the full uncertainty and thus constrain ensemble dispersion. The smaller the model domain and the longer the forecast lead time, the more severe the problem (Du and Tracton, 1999). While Nutter et al. (2004b) proposed a possible solution to this issue involving adding structured noise to the LBC updates, it has not been tested for operational implementation.

2. The Way Forward

There is a clear potential for the US to produce greatly improved probabilistic forecasts, but how good do such forecasts need to be, and how do we justify the necessary R&D and computational costs to get there? One route is the top-down approach where the intricate decision processes of the myriad of mesoscale forecast users are analyzed to determine requirements and benefits, and thus establish EPS design standards with a thorough business case analysis. However, this would be an incredibly complex and time consuming task involving help from decision analysts. A more practical approach is to work bottom-up. Using our knowledge of NWP and experience as forecasters, the target is an EPS that likely meets the needs of the majority of users.

NCEP currently has several planned ensemble upgrades through 2014. In 2011, GEFS will be improved by running at T254 (~55 km). Also in 2011, the SREF will be improved to a 22-km grid and addition of the NEMS-NMMB model with stochastic parameterizations. Additionally, to boost ensemble size, SREF may be combined with the Canadian regional ensemble system (33 km, 20 members). In 2013, SREF will be further improved to a 12-km grid (which still does not permit explicit depiction of most mesoscale features) and use of the ensemble transform technique to generate ICs for all members. SREF will still be multi-model system but use only NEMS-NMMB and NEMS-ARW, both with stochastic parameterizations. In 2014, NCEP plans to introduce a North America Rapid Refresh Ensemble (NARRE) designed for near-term (24 h) forecasts focused on aviation support. NARRE will be similar in design to SREF except with an hourly update cycle, a 3-km grid, and only 6 members. To support high-impact weather forecasting, NCEP will at the same time introduce a 6-member High-Resolution Rapid Refresh Ensemble (HRRRE) that will have nested domains with 1-km grid spacing within CONUS and Alaska.

a) *Design Targets*

NCEP's planned ensemble upgrades represent valuable and aggressive improvements over current capabilities. However, the upgrades' success is dependent on the timely purchase of the next NOAA operational computer system and successful ensemble design and associated testing. Even under the best case scenario in terms of acquisition of sufficient computational resources between now and 2015, NCEP will need the assistance of the enterprise to conduct systematic testing of high-resolution ensemble designs in order to achieve goals identified at the workshop.

Table 2 lists recommended design specifications for NCEP ensembles to be operational by 2015. No attempt was made here to optimally balance the various design factors. GEFS requires significant additional upgrades to provide high quality, extended forecasts as well as to provide robust LBC updates (explained below) to ensure the quality of the short-range, regional ensemble effort. SREF and NARRE also need increased model resolution. The resolutions and forecast lengths of the ensembles is designed around the principle that predictability is lost starting from small scales working upward as lead time increases, thus longer lead times have coarser model resolution and less frequent updates.

A challenging area of research is system optimization, that is, investigation into trade-offs in the design configuration (model resolution, domain size, number of members, etc.). The number of design parameters combined with their ranges of possible settings

Table 2. NCEP ensembles target specifications for implementation by 2015. A relocatable ensemble, which may be necessary for tropical cyclone forecasting if output from GEFS and SREF does not suffice, is left for further consideration by the working groups.

			Resolution		Forecast	Update	# of
Name	**Model**	**Domain**	**Grid**	**Levels**	**Length (h)**	**Frequency (h)**	**Members**
GEFS	GFS	Global	T1000 (~14 km)	90	84	6	20
			T399 (~35 km)	40	84-360	12	20
SREF	NMMB	N. America	4 km	55	48	3	20
NARRE	NMMB	CONUS and AK	1 km	90	12	1	20
?	cyclone?	Relocatable	TBD	TBD	TBD	TBD	TBD

creates a seemingly endless number of permutations to test. Furthermore, the answer for the best settings may vary greatly depending upon how results are measured. This research question will require careful attention and considerable effort by the enterprise.

Model resolution is a critical design factor since it greatly influences both forecast value and the costs of producing the information. It was agreed that explicit modeling of convection is key since it frequently and severely impacts most users with phenomena such as damaging winds, lightning, severe turbulence, and flash flooding. Using a grid spacing of 10 km or more, an ensemble using parameterized convection can be designed to accurately simulate the forecast uncertainty (i.e., achieve statistical consistency) and provide reliable forecast probability. However, relative to a convection-resolving ensemble, the forecasts would be less sharp due to the larger forecast uncertainty. Using grid spacing of 4 km or less would bring the convection into focus and greatly reduce the uncertainty, resulting in significantly higher value in decision making.

b) Accounting for Analysis and Model Uncertainty

Extensive testing is needed to decide on the best choice for generating ICs. NCEP and NOAA/ESRL have already embarked on systematic comparison of the ET and EnKF methods as well as exploration of ways to reduce EnKF costs. Besides robust ensemble ICs, a potential benefit of EnKF is improved data assimilation (i.e., reduced analysis uncertainty, thus sharper forecast PDF). For regional ensemble systems, care must be taken to ensure the IC generation method is compatible with the LBC updates coming from the global ensemble. A potentially critical ability for short-term forecasts may be assimilation of radar data (Xue et al. 2009).

There are many questions to investigate involving treatment of model uncertainty in the ensembles, but the biggest need is basic research into more physically based algorithms designed to sample model uncertainty. One example is comparison of solutions from cloud-resolving simulations to parameterized solutions, which may help develop more effective cumulus parameterizations as well as determine how to stochastically perturb the parameterizations.

In considering only current methods, testing is needed on whether to include the multi-model approach or only use some combination of the other techniques with a single, best-model. Does the value added by a multi-model ensemble outweigh the many detriments (discussed above)? Similarly for the other techniques, their costs to develop, implement, and maintain should be weighed against their value added. The most promising of all the methods may be stochastic parameterizations. An additional question for investigation is which methods can be applied together without duplicating simulation of a particular source of uncertainty. An obvious example would be the overlap between using perturbed surface parameters and coupling to an ocean/LSM ensemble. A more subtle conflict may come from using stochastic physics along with random parameters.

By design, GEFS can directly provide robust LBC updates that effectively capture and feed in the error growth external to SREF. Using too coarse a resolution for GEFS would require running additional regional grid(s) to nest down to the SREF domain and also require extending SREF forecast length to fill needs not met by GEFS. Additionally, a coarser GEFS would likely reduce LBC diversity and degrade SREF skill. Given the fairly short forecast lead times combined with the fairly large domains, SREF may be able to provide sufficient LBC updates for NARRE. Testing is needed to investigate whether the dispersion of SREF and NARRE is seriously constrained by this LBC strategy. If so, an approach following Nutter et al. (2004b) should be explored.

c) *Improving the Foundation*

Apart from model resolution discussed above, the workshop unfortunately devoted very little time to discussing needed improvements to the EPS foundation. Ironically, the foundation may be the EPS component containing the greatest potential for advancing the entire EPS. There are still significant weaknesses in analyses and forecast models whose improvement would drastically reduce the demands on the ensemble and result in a sharper forecast PDF and more value to the user.

Ensemble researchers need to work more closely with model developers and data assimilation experts to identify areas of the foundation to improve that would most benefit the EPS. Top priority areas would include phenomena highly sensitive to error (e.g., initiation of convection) and any source of forecast uncertainty extremely difficult to simulate in the ensemble (e.g., uncertainty from parameterization of cloud microphysics).

B. Postprocessing

Ensemble postprocessing (or "calibration") is integral to producing high-quality probability forecasts. Calibration can ameliorate the systematic deficiencies in ensemble predictions (i.e., biases in mean and spread), which may be caused by sampling error from too few members, by inadequate model resolution, by improper initial conditions, or by model errors. These deficiencies are especially pronounced for surface and sensible-weather elements such as heavy precipitation, severe weather occurrence, surface temperatures, and wind speeds (Eckel and Mass, 2005). Most postprocessing methods adjust the current forecast using statistical relationships between past forecasts and observations. The end result of ensemble postprocessing should be probabilistic predictions that are reliable (e.g., 20 percent event occurrence when 20 percent probability is forecast) while retaining as much sharpness as possible (Wilks 2006, Gneiting et al. 2007). This makes the forecasts much more useful to a wide range of users.

1. Status

For deterministic forecasts, postprocessing techniques were first brought to fruition in US operations through Model Output Statistics, or MOS techniques (Glahn and Lowry 1972). Since the late 1990s, the ensemble postprocessing literature has grown rapidly. A general review of the ensemble postprocessing literature through 2006 is provided in Hamill et al. (2006). Since then, there has been further development of the Bayesian Model Averaging techniques (Raftery et al. 2005; Wilson et al. 2007; Berrocal et al. 2007; Sloughter et al. 2007, 2009; Fraley et al. 2009); applications of and technique development with reforecast training data sets (Hamill and Whitaker, 2006, 2007; Wilks and Hamill 2007; Hagedorn et al. 2008; Hamill et al. 2008; Fundel et al. 2009a, 2009b); the Bayesian Processing of Forecasts algorithm (Krzysztofowicz and Evans 2008); and ensemble-MOS techniques (Gneiting et al 2005; Glahn et al. 2009; Unger et al. 2009).

A wide variety of postprocessing techniques have been shown to improve probabilistic forecasts for variables with more consistent biases and quasi-normally distributed errors, but there are still many questions. Only a small body of literature discusses which of the techniques may be preferable to use with a given ensemble design. Techniques designed for rare and high-impact events are less well developed. And lastly, how can postprocessing be designed to achieve seamless probabilistic forecasts from the shortest to the longest forecast leads when using an EPS design, as in Table 2, that involves blending ensemble forecasts with different resolutions, forecast lengths, and update frequencies? This is a critical question for populating the 4-D data cube being built by NOAA to support NextGen.

There are some generally agreed-upon principles for ensemble postprocessing. All things being equal, simple algorithms are preferable to complex ones, given the cost of code maintenance. Similarly, maintaining only one algorithm that works well for any variable (both raw model output and derived quantities) is preferable to maintaining multiple, specialized algorithms. The postprocessed output must be meteorologically consistent so that, for example, a high probability of temperatures below freezing is not be accompanied by a high conditional probability of rain. Postprocessing techniques need high-quality observational and/or analysis data. Training data should use the same forecast model as applied in the real-time forecasts, and furthermore, large reforecast training data sets are very beneficial to ensemble postprocessing, especially for calibration of rare events. However, a careful analysis is needed to compare the cost of generating large training datasets to the ensuing benefits.

2. The Way Forward

Raw ensemble output will improve in the coming years through the use of better models, increased model resolution, and enhanced techniques for simulating uncertainty, however, for the foreseeable future, imperfections will continue to be large enough that most users will benefit from postprocessed forecasts. As we envision a mesoscale EPS of the future, we have the opportunity to simultaneously develop and integrate postprocessing as a critical piece between the raw ensemble forecasts and the delivered information, thus achieving the highest possible quality support to the decision makers.

We need a development process that balances advancement of NWP and postprocessing techniques. Postprocessing produces best results occurs when the forecast model is accompanied by a long, consistent training data set, but the production of these data sets may slow down model development. The development of an improved EPS offers an opportunity to work collaboratively and perhaps find compromises.

The end goal many years hence is postprocessed guidance that is skillful (as sharp as possible while still being reliable), comprehensive, dynamically and internally consistent, and seamless in time. The system should be cost-effective to maintain and adaptable to new models and new observational data sets. To move toward this goal, we recommend the following process for generating calibrated probabilistic forecast data.

a) Define the Requirements

What form should the postprocessed output take? Should the output be adjusted ensemble members, or should it be probability density functions, or both? What metrics will be used for evaluating various postprocessing methods, and what criteria should be established for selecting methods? What data sets will be used to develop, evaluate, and compare procedures? Answering such questions is a necessary first step in structuring a postprocessing effort. Perhaps the NWS's Office of Climate, Weather, and Water Services can lead this effort, entraining the ensemble and social science expertise in the wider community.

b) Prepare for Testing

This step would involve collecting and perhaps developing the observational, reanalysis, and reforecast data sets that would be used in the postprocessing, as well as in developing a library of existing postprocessing routines and verification software. Since postprocessing technique development will occur in parallel with ensemble system development, it will be necessary to use forecast data from current-generation systems. To support users who wish to do their own postprocessing, the forecast, reforecast, and observational data, as well the data-generating software, must be made available to the enterprise.

c) Develop New Algorithms

Ideally, funding would be provided by NOAA, both internally and through external grants, and by NSF to develop new postprocessing methods and/or to refine the existing methods. An important research area is how to apply postprocessing algorithms in situations where no trustworthy observations or analyses are available (e.g., calibrating probabilistic dust-storm forecasts in desert regions will be difficult). Participants would be expected to use the data sets collected (step *b* above) and evaluate according to specified metrics (step *a* above).

d) Inter-compare and Recommend Algorithms for Technology Transition

An independent organization such as the Developmental Testbed Center (DTC) in Boulder, CO, would evaluate the algorithms according to objective criteria. They would recommend one or a few for technology transition, or perhaps recommend hybridization of techniques or further development. Recommendations would take into account the complexity of implementation as well as the degree of potential improvement.

e) Operational Model Development, Deployment, and Production

The NWS, through the Meteorological Development Laboratory (MDL) and/or NCEP would be given the sample research-quality algorithms (step *d* above) and would produce and test operational code then implement the postprocessing algorithms.

f) Monitor Quality

The operational forecast quality would be monitored, perhaps by the production facilities (in step *e* above) or by an independent organization such as the DTC. When deficiencies are noted, they could be fed back to refine the requirements (step *a* above).

C. The Interface

To be beneficial, forecast uncertainty information must provide an overall positive impact on users' decisions, whether it is obtained directly from the forecaster, filtered through an emergency manager, or incorporated in a decision algorithm. If the user does not understand the forecast, finds it too taxing to incorporate into the decision process, or does not trust it, the information will have no impact, regardless of its quality. This section discusses aspects of building an effective EPS interface that enables the forecast uncertainty information to be fully exploited.

1. Status

It is virtually undisputed that the theoretical and economic advantages of optimized application of probabilistic forecasts is superior to single-valued forecasts. The debate centers on the extent to which human decision makers can realize that advantage. Automated decision systems are of increasing interest, but people are the focus here since they are critically involved in the majority of key, weather-related decision making. Most people are not familiar with the mathematical principles that underlie the theoretical advantages nor do they tend to make decisions that reflect these principles (Tversky & Khanaman 1974, 1981, 1992). Even when explicit decision rules are calculated for them, people do not necessarily follow them. Especially difficult are low-probability, severe-weather situations in which precautionary action is required but ignored (Baker 1995; LeClerc & Joslyn 2009).

Even sophisticated users do not always recognize the advantage of uncertainty information. Uncertainty products are often ignored by forecasters (Joslyn et al. 2007; Joslyn and Jones 2008), commercial and agricultural enterprises, and those in resource management (Grimit 2009; Pulwarty and Redmond 1997; O'Connor et al. 2005; Changnon et al. 1995). Thus, even people whose needs and expertise position them to make the best use of uncertainty forecasts are not convinced of the value of incorporating the information into the decision process.

Uncertainty information is challenging to human information processing in part because it increases overall processing load. It requires the decision-maker to consider multiple potential outcomes and consequences as well as the corresponding levels of uncertainty. Particularly difficult are rapid decisions that rely solely on "working memory" since there are only so many things people can consider simultaneously (Miller 1956). To exacerbate the processing issue, the wide variety of forecast uncertainty products currently available (see NRC 2006 for examples) were primarily designed by atmospheric scientists with little consideration of human cognition.

Another challenge is that probability is an abstract construct whose meaning is debated even within the scientific community (deElia and Laprise 2005). The most generally accessible meaning may be the "frequentist" interpretation, e.g. an 80% chance of precipitation means that precipitation will occur 8 out of the 10 times it is forecast at 80%. However, this interpretation only makes sense for repeatable events and is difficult to apply to rare, extreme events for which the consequences can be weighty. Thus, understanding uncertainty may be most difficult when successful communication is most important.

In light of these challenges, one approach is to omit uncertainty from weather forecasts and then make the decision for the end user, but the problems associated with this strategy may be far worse (Freudenburg 1996; Slovic 1993). Giving a single-value forecast can appear patently wrong whenever the forecast fails to verify, decreasing trust and willingness to comply with future weather warnings. Additionally, users may attempt to second guess a recommended course of action, assuming that it may be appropriate for others but does not apply to their own specific situation.

Many of these challenges point to limited education, which may currently be the weakest aspect of the EPS interface. There is some basic training available for forecasters (e.g., COMET 2005), but it lacks material on understanding how to best support user decision making. On the user side, there is currently no formalized process or program for educating users on optimal application of weather information.

2. The Way Forward

Research on human decision makers' ability to realize the theoretical benefit of probability forecasts is as yet a largely unexplored topic. As such, rather than specify a development path, this section describes experimental studies needed to guide development of the EPS interface. Successful research in these areas is challenging to design and to conduct but is extremely important both in economic terms and in terms of human safety.

a) Marketing

Advancing from use of traditional to probabilistic forecasts can involve costs such as education of personnel, restructuring of decision processes, and infrastructure upgrades to handle the additional information, as well as the cognitive costs mentioned above. To sell users on the likely handsome returns, solid behavioral evidence is needed to convincingly demonstrate the advantages. There are now a handful of studies showing that explicit uncertainty information improves economic decision making among nonexperts (Roulston et al. 2006; Nadav-Greenberg and Joslyn 2009).

Research is required that investigates whether people make better decisions given uncertainty forecasts in a wide range of weather and decision contexts. Important questions are whether uncertainty forecasts provide an advantage for 1) one-time decisions, which may be regarded as fundamentally different from repeated decisions (Patt and Shrag 2003), 2) group as well as individual decision-making, and 3) decisions whose benefits are not easily quantifiable. The aim is to demonstrate a clear advantage in terms of realistic decision outcomes, but any situations discovered that lack clear advantages will allow focusing of efforts on the areas that do demonstrate benefit.

b) Methods for Communicating Forecast Uncertainty

Existing studies described by the NRC (2006) provide some direction for improving communication of uncertainty, but research needs extension to realistic weather-related decisions in such areas as:

- Use of categorical risk expressions (e.g., high/med/low), which can be problematic (Walsten et al. 1986). User reactions are often more appropriate when given a frequency expression (e.g., 1 in 10) rather than a probability (Gigerenzer and Hoffrage 1995). Do those findings apply to weather forecasts and any user?
- Risk communication in severe-weather situations having low probability and high potential loss. Does the inclusion of consequences to the user (e.g., "widespread electrical outages") help people to understand the importance of precautionary action (Chagnon et al. 1995)? Should the worst-case scenario be emphasized?
- Visualizations that help people understand uncertainty information. Formats compatible with the decision at hand and with user expectations may reduce the transformations required to incorporate such information (Joslyn et al. 2008).
- Dealing with misinterpretation of probability expressions. They are often believed to be deterministic quantities such as percent of the area that will be affected (Gigerenzer et al. 2005; Joslyn et al. 2009). Predictive intervals can be mistakenly interpreted as fluctuations over the course of the forecast period. We need to uncover these and other basic error tendencies and then develop strategies to overcome them.

c) User Trust

All the effort of an EPS is wasted if the user distrusts and chooses to ignore the forecast. Are there specific expressions of forecast uncertainty that improve/detract from trust? Uncertainty forecasts anticipate a range of outcomes so users do not generally lose trust by perceiving them as wrong, as is possible with a single-value forecast. On the other hand, to an uneducated user an uncertainty forecast may appear to be hedging, hence untrustworthy. Could inclusion of verification data restore user trust under some circumstances? Lastly, how serious is the loss of trust due to variability in sequential forecasts for the same event, and how can we mitigate this problem?

d) Forecaster Activities

Historically, in the NWS the official deterministic forecast has not come directly from either a model or postprocessed guidance but the synthesis of this information after examination and adjustment by human forecasters. Does the shift toward predominantly probabilistic forecasting require an adjustment to that paradigm? In what situations (e.g., very short-range, high-impact?) can a forecaster add value to the well-calibrated ensemble forecast data? Research has shown that biases such as overconfidence seriously impact forecaster-derived probability statements (Keith 2003), which could perhaps be alleviated by improved training. The impact of forecaster intervention needs careful evaluation to help design the forecasters' activities in supporting optimal decision making.

e) Decision Support Systems

In many specific situations where repeated, similar decisions are made (e.g., reroute aircraft to avoid potential severe thunderstorms), the best approach may be to incorporate uncertainty estimates into decision support systems. But, for people to put them to good use, the decision algorithms should be transparent to the user (Endsley et al. 2003). Research is needed to explore the design and function of such systems and their role in supporting human decision making.

f) Education

The issue of education is intertwined to all the above topics and requires significant consideration. Users need not only to be convinced of the benefits of uncertainty forecasts, but also need to be taught how to optimally apply the information. Considering the wide array of potential users, this may be a formidable task for which resources will need to be acquired. Forecasters will need to continue to learn the fundamentals of ensemble methods and probabilistic forecasting as well as understand how users intend to apply forecast uncertainty information.

g) Research Methods

Sound experimental research is crucial to fully understand the psychological issues outlined above. Because many of the cognitive processes involved in decision-making are not open to conscious awareness, we cannot rely entirely on what people think they need. Thus, well-designed research is needed to compare decisions based on deterministic forecasts and various forms of uncertainty information.

It is critical for good research to use realistic tasks and stimuli in investigating the practical application of uncertainty expressions, including procedures that elicit optimal performance from participants. Since uncertainty information requires additional cognitive effort, participants need motivation that mimics real world situations in order to reveal the true advantages from following uncertainty forecasts. Additionally, using realistic tasks is important because meaningful, well-known contexts (i.e., actual forecasts, observations and decisions) facilitate logical thinking (e.g., Griggs and Cox 1982). If the stimuli contradict prior experience, even subtly, it may have an unintended influence on people's responses and invalidate research results.

D. R&D

1. Status

Today there are many "centers of excellence" for mesoscale modeling and ensembles, primarily at government centers and at universities. Sophisticated models and associated data assimilation systems have been developed and are sometimes run in a quasi-operational mode over periods of many years. Fully operational systems at government centers require 24/7 computer and communications backup capabilities to meet the required 99.95% on-time product delivery. Additionally, although it is widely recognized that the forecast utility is considerably increased with postprocessing, full-bodied postprocessing systems have not yet been associated with such models.

NWP is and continues to be one of the greatest achievements of modern times. Ideas, theories, and techniques are shared through technical papers, journal articles, personal communication and at conferences. Collaboration has been essential for supporting overall progress in this and other countries. However, the transition of cutting-edge research into NOAA operational systems needs to be more efficient. To reach the goals of this workshop, the centers of excellence need the opportunity to play a more significant role in advancing the development of the national operational EPS.

The challenges to efficient implementation from research into operations are the complexity and demands of the operational environment (e.g., essential backup, 24/7 support, on-time delivery, coding requirements, and rigorous testing standards). Incremental improvements in skill must be weighed against the level of additional cost and risk associated with the system improvements. New ideas and methods must be rigorously tested in the operational system to ensure changes do not have unanticipated negative consequences. Outside organizations are naturally frustrated with what appears to be a slow, inefficient implementation process. However, operational implementation will always require rigorous testing to mitigate the risk of service degradation.

2. The Way Forward

Considerable investment of resources will be necessary to realize the advancement envisioned in this paper in the form of high-performance computing (for both R&D and operations), R&D, and transition into operations. However, the return on this investment will be tremendous.

a) Governance Structure

We recommend that a National Advisory Committee (NAC) be created to act as the standing governing body for R&D of the nation's probabilistic forecasting capability. It will foster and lead an effective collaborative effort that moves the US forward in probabilistic prediction. The NAC will consist of membership from NOAA, NCAR, DOD weather, academia, the private sector, and the FAA. Their initial charge is to work with NOAA to draft an implementation plan to accelerate the advancement of the operational mesoscale EPS at NCEP.

The NAC would: (1) advise the Assistant Administrators of NWS and OAR concerning the activities, requirements, and accomplishments of the probabilistic forecasting activity, (2) appoint working groups as necessary to carry out specific studies and make recommendations to the NAC, (3) work to establish a standing council under the NOAA Science Advisory Board, and (4) continually review the development progress and performance of national EPS capabilities. There was also general agreement at the workshop that there needs to be high-level support at NOAA to champion the national EPS and lead the effort to secure the necessary resources (funding for R&D and operational high-performance computing).

The working groups will be the heart and soul of the collaborative effort, where ideas and processes are collected, exchanged, and further developed. The groups must not simply be "paper committees" that only rarely meet and only discuss issues. They will be the main conduit to solicit input from the field and must stay actively engaged through regular meetings (both in person and via telecoms) to keep the momentum going for the entire effort. Working groups to be formed soon after initiation of the NAC are:

- ***Ensemble Design***: Address questions from section A. One high-priority task is to work closely with the Business Case group on building a reliable cost estimate for the necessary high-performance computers.
- ***Postprocessing***: Address questions from section B.
- ***EPS Interface***: Address questions from section C. This diverse area may logically be broken up into several working groups (e.g., product design, education, marketing, etc.)
- ***Testing***: Work with the DTC and NCEP EMC (see below) to make recommendations on the processes and procedures to be used in R&D of the EPS.
- ***Business Case***: Develop a business case for a strong and viable mesoscale EPS by establishing requirements of the key users and researching cost-benefit analysis of the proposed system. Cost considerations will include the necessary high-performance computers, infrastructure, R&D/transition to operations, and system sustainment.

b) Testing

R&D, optimization, implementation, and future evolution of the national mesoscale EPS is dependent upon solid testing and evaluation. Formal testing procedures must be agreed upon and well managed in order to fairly judge and make good decisions on EPS design. This is a big undertaking and the effort must not be underestimated. The DTC in Boulder is ideal for the purpose. Note that in February 2010, DTC started work on establishing the DTC Ensemble Testbed (DET), designed to aid this effort.

A systematic approach must be defined that will lead to a prioritization of techniques and methodologies identified by the working groups to be tested under this new framework. Potential contributions to the EPS will be assessed based on their expected or measured value to improve upon operational NCEP EPS and contribute to the overall objectives as defined by the NAC. Methods deemed promising will be tested for their potential inclusion using the current prototype EPS, maintained at the DTC. The working groups will use test results to support their recommendations. In judging the different methods, the important considerations are:

- Probabilistic forecast skill (i.e., reliability and resolution).
- Forecast value (i.e., user benefit) to demonstrate ability to support optimal decision making.
- Consistency of forecasts in time, space, among weather elements, and from model run to run (i.e., different initiation times)
- Computational efficiency and maintenance costs

Besides testing, the DTC will facilitate collaboration among government, academia, and the private sector for creation of the prototype EPS. The DTC will make the prototype EPS available to support independent research and inter-comparison of various methods. In order to streamline the transition into operations, the prototype EPS and all newly introduced algorithms will be designed to be compatible with NCEP operating frameworks. NCEP will play a critical role in defining testing procedures that will promote successful transition and implementation of new techniques.

Summary

This white paper encapsulates the National Workshop on Mesoscale Probabilistic Prediction held in Boulder, CO, in September 2009. It was agreed that US operational probabilistic prediction has not kept pace with the known science and available technology. The goal was set to implement a radically upgraded national capability for probabilistic prediction by 2015 to support current and future decision-making needs. Meeting this goal will require a significant infusion of resources for high-performance computing and research, as well as a coordinated collaborative effort by the whole enterprise. Ideas were discussed on the way forward for four critical areas of the overall ensemble prediction system: (1) the ensemble design, (2) postprocessing the ensemble model output, (3) the user interface, and (4) R&D.

A governance structure was proposed, to be led by a National Advisory Committee (NAC) as the standing governing body for R&D of the future national EPS, which will build upon the current and planned upgrades to the NCEP ensembles. The NAC will be responsible for establishing working groups to address the key development areas and for composing an implementation plan for the future EPS. It is recommended that testing for EPS design and configuration be led by the Developmental Testbed Center in Boulder.

Acronyms

4D-VAR	four dimensional variational data assimilation
ACUF	AMS Ad hoc committee on Uncertainty Forecasts
AMS	American Meteorological Society
ARW	Advanced Research WRF
COMET	Cooperative Program for Operational Meteorology
CONUS	Continuous United States
DOD	Department of Defense
DTC	Developmental Testbed Center
ECMWF	European Centre for Medium-Range Weather Forecasts
EMC	Environmental Modeling Center
EnKF	Ensemble Kalman Filter
EPS	ensemble prediction system
ET	ensemble transform
FAA	Federal Aviation Administration
GEFS	Global Ensemble Forecast System
GFS	Global Forecast System
ICs	initial conditions
LBC	lateral boundary condition
LSM	land surface model
MOGREPS	Met Office Global and Regional Ensemble Prediction System
MOS	Model Output Statistics
NAC	National Advisory Committee
NAM	North American Model
NARRE	North America Rapid Refresh Ensemble
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Center for Environmental Prediction
NextGen	Next Generation Air Transportation System
NMMB	Nonhydrostatic Multiscale Model on B-grid
NOAA	National Oceanographic and Atmospheric Administration
NRC	National Research Council
NWP	numerical weather prediction
NWS	National Weather Service
OAR	Office of Oceanic and Atmospheric Research
PDF	probability density function
R&D	research and development
RSM	Regional Spectral Model
SREF	Short-Range Ensemble Forecast system
WRF	Weather Research and Forecasting model

References

American Meteorological Society Commission on the Weather and Climate Enterprise, Board on Enterprise Communication, 2010: *A weather and climate enterprise strategic implementation plan for generating and communicating forecast uncertainty information*. Available at [http://www.ametsoc.org/boardpges/cwce/docs/BEC/ACUF/2010-01-Plan.pdf], 108 pp.

American Meteorological Society, 2008: Enhancing weather information with probability forecasts. Available at [http://www.ametsoc.org/policy/2008enhancingweatherinformation_amsstatement.html]

Baker, E. J., 1995: Public response to hurricane probability forecasts. *The Professional Geographer*, **47**, 137-147.

Berner, J., G. J. Shutts, M. Leutbecher, and T. N. 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.

Berrocal, V. J., A. E. Raftery, and T. Gneiting, 2007: Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Mon. Wea. Rev.*, **135**, 1386–1402.

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

Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **125**, 99–119.

-----, M. J. Miller, T. N. Palmer, 1999: Stochastic simulation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **125**, 2887–2908.

-----, P.L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, Y. Zhu, 2005: A comparison of the ECMWF, MSC and NCEP Global ensemble prediction systems. *Mon. Wea. Rev*. **133**, 1076-1097.

Changnon, S. A., J. M. Changnon, and D Changnon, 1995: Uses and applications of climate forecasts for power utilities. *Bull. Amer. Meteor. Soc.*, **76**, 711–720.

Clark, A. J., W. A. Gallus, 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.

Cooperative Program for Operational Meteorology, Education and Training (COMET), 2005: *Introduction to Ensemble Prediction*. University Corporation for Atmospheric Research. Available at [http://www.meted.ucar.edu/nwp/pcu1/ensemble_webcast/]

deElia, R. and R. Laprise, 2005: Diversity in interpretations of probability: Implications for weather forecasting. *Mon. Wea. Rev.,* **133**, 1129-1143.

Descamps, L., and O. Talagrand, 2007: On some aspects of the definition of initial conditions for ensemble prediction. *Mon. Wea. Rev.*, **135**, 3260–3272.

Du, J., G. DiMego, Z. Toth, D. Jovic, B. Zhou, J. Zhu, H. Chuang, J. Wang, H. Juang, E. Rogers, and Y. Lin, 2009: NCEP short-range ensemble forecast (SREF) system upgrade in 2009. 19th Conf. on Numerical Weather Prediction and 23rd Conf. on Weather Analysis and Forecasting, Omaha, Nebraska, Amer.Meteor. Soc., June 1-5, 2009, paper 4A.4 [available online at http://www.emc.ncep.noaa.gov/mmb/SREF/reference.html]

Du, J., and M. S. Tracton, 1999: Impact of lateral boundary conditions on regional-model ensemble prediction. Research Activities in Atmospheric and Oceanic Modelling (edited by H. Ritchie), Report 28, CAS/JSC Working Group Numerical Experimentation (WGNE), WMO/TD-No. 942, 6.7-6.8.

Eckel, F. A., and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350.

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

Endsley, M., D. Bolte, and D. Jones, 2003: *Designing for situation awareness: An approach to user-centered design*. London: Taylor & Francis. 312 pp.

Fraley, C., A. E. Raftery, and T. Gneiting, 2010: Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Mon. Wea. Rev*., 138, 190–202.

Freudenburg, W., 1996: Risky thinking: Irrational fears about risk and society. *Annals of the American Academy of Political and Social Science* **545**, 44-53.

Fundel, F., M. A. Liniger, C. Appenzeller, and A. Walser, 2009: The cost efficient calibration of a limited area ensemble prediction system using reforecasts. *Mon. Wea. Rev*., in review. Available from felix.fundel@meteoswiss.ch.

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

Gigerenzer, G., R. Hertwig, E. van den Broek, F. Fasolo, and K. V. Katsikopoulos, 2005: A 30% chance of rain tomorrow: How does the public understand probabilistic weather forecasts? *Risk Analysis*, **25**, 623-630.

Gigerenzer, G., and U. Hoffrage. 1995. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, **102**, 684-704.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor*., **11**, 1203–1211.

-----, 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.

Gneiting, T., A. E. Raftery, A. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.,* **133,** 1098-1118.

-----, F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration, and sharpness. *J. Roy. Stat. Soc*., **69B**, 243-268.

Griggs, R. A. and J. R. Cox, 1982: The elusive thematics material effect in Wason's selection task. *British Journal of Psychology*, **73**, 407-420.

Grimit, E., 2009: 3TIER: A renewable energy information services company. *National Workshop on Mesoscale Probabilistic Prediction*, Boulder, CO, September 23, 2009.

Hagedorn, R, T. M. Hamill, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble forecasts. Part I: 2-meter temperature. *Mon. Wea. Rev*., **136**, 2608-2619.

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.

-----, J. S. Whitaker, and S. L. Mullen, 2006: Reforecasting: improving skill using retrospective forecasts. *Bull. Amer. Meteor. Soc*., **27**, 33-46.

-----, and J. S. Whitaker, 2007: Ensemble calibration of 500 hPa geopotential height and 850 hPa and 2-meter temperatures using reforecasts. *Mon. Wea. Rev*., **135**, 3273-3280

-----, and J. S. Whitaker, 2010: What constrains spread growth in forecasts initialized from ensemble Kalman filters? *Mon. Wea. Rev*., conditionally accepted. Available from tom.hamill@noaa.gov

-----, R. Hagedorn, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble forecasts. Part II: precipitation. *Mon. Wea. Rev*., **136**, 2620-2632.

Holt, T., J. Pullen, and C. H. Bishop, 2009: Urban and ocean ensembles for improved meteorological and dispersion modeling of the central zone. *Tellus A*, **61**, 232-249

Hou, Dingchen, Zoltan Toth and Yuejian Zhu, 2006: A stochastic parameterization scheme within NCEP Global Ensemble Forecast System. 18th AMS conference on Probability and Statistics. Atlanta, GA, Jan 29-Feb 2, 2006.

Joslyn, S., K. Pak, D. Jones, J. Pyles, and E. Hunt, 2007: The Effect of probabilistic information on threshold forecasts. *Wea. Forecasting,* **22***,* 804–812.

-----, L. Nadav-Greenberg, and M. U. Taing, 2008: The Effects of wording on the understanding and use of uncertainty information in a threshold forecasting decision. *Applied Cognitive Psychology,* **23**, 55-72.

----- and D. Jones, 2008: Strategies in naturalistic decision-making: A cognitive task analysis of Naval weather forecasting. In J. M. Schraagen (ed.), *Naturalistic Decision Making and Macrocognition*, Ashgate Publishing.

-----, L. Nadav-Greenberg, and R.M. Nicholls, 2009: Probability of Precipitation: Assessment and Enhancement of End-User Understanding. *Bull. Amer. Meteor. Soc.,* **90***,* 185-193.

Keith, R., 2003: Optimization of value of aerodrome forecasts. *Wea. Forecasting*, **18**, 808–824.

Krzysztofowicz, R., and W. B. Evans, 2008: Probabilistic Forecasts from the National Digital Forecast Database. *Wea. Forecasting*, **23**, 270–289.

Kuhl, D., I. Szunyogh, E. J. Kostelich, G. Gyarmati, D. J. Patil, M. Oczkowski, B. R. Hunt, E. Kalnay, E. Ott, and J. A. Yorke, 2007: Assessing predictability with a local ensemble kalman filter. *J. Atmos. Sci.*, **64**, 1116–1140.

LeClerc, J. and S. Joslyn, 2009: *Role of Uncertainty Information and Forecast Error in Weather-Related Decision Making.* Presentation at the Annual Meeting of the Psychonomics Society, Boston, MA.

Miller, G. A., 1956: The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* **63**, 81-97.

Nadav-Greenberg, L. and S. Joslyn, 2009: Uncertainty forecasts improve decision-making among non-experts. *Journal of Cognitive Engineering and Decision-Making*, **2**, 24-47.

National Research Council (NRC), 2006: *Completing the forecast: Characterizing and communicating uncertainty for better decisions using weather and climate forecasts.* Washington, DC: National Academies Press.

Nutter, P., D. Stensrud, and M. Xue, 2004a: Effects of coarsely resolved and temporally interpolated lateral boundary conditions on the dispersion of limited-area ensemble forecasts. *Mon. Wea. Rev.*, **132**, 2358–2377.

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

O'Connor, R., B. Yarnal, K. Dow, C. S. Jocoy, and G. Carbone, 2005: Feeling at-risk matters: Water managers and the decision to use forecasts. *Risk Analysis*, **25**, 1265-1275.

Patt A. G. and D. Schrag, 2003: Using specific language to describe risk and probability. *Climate Change* **61**, 17–30.

Pulwarty, R. S., and K. T. Redmond, 1997: Climate and salmon restoration in the Columbia River basin: The role and usability of seasonal forecasts. *Bull. Amer. Meteor. Soc*, **78**, 381–397.

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

Roulston, M. S., G. E. Bolton, A. N. Kleit, and A. L. Sears-Collins, 2006: A laboratory study of the benefits of including uncertainty information in weather forecasts. *Wea. Forecasting*, **21**, 116-122.

Slovic, P., 1993: Perceived risk, trust, and democracy. *Risk Analysis*, **13(6)**, 675-682.

Sloughter, J. M., A. E. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. *Mon. Wea. Rev.*, **135**, 3209–3220.

-----, T. Gneiting, and A. E. Raftery, 2010: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *Journal of the American Statistical Association*, in press.

Stensrud, D. J., J. W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107.

Szunyogh, I., and Z. Toth, 2002: The effect of increased horizontal resolution on the NCEP global ensemble mean forecasts. *Mon. Wea. Rev*., **130**, 1125-1143.

Talagrand, O., R. Vautard, and B. Strauss, 1999: Evaluation of probabilistic prediction systems. *Proc. ECMWF Workshop on Predictability,* Shinfield Park, Reading, United Kingdom, ECMWF, 1–25.

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.

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

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

-----, Y. Zhu and T. Marchok, 2001: On the ability of ensembles to distinguish between forecasts with small and large uncertainty. *Wea. Forecasting*, **16**, 436-477.

Tversky, A. and D. Kahneman, 1974: Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124-1130.

Tversky, A. and D. Kahneman, 1981: The framing of decisions and the psychology of choice. *Science,* **211**, 453-458.

Tversky, A. and D. Kahneman, 1992: Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty*, **5**, 297-323.

Unger, D.A., H. van den Dool, E. O'Lenic, and D. Collins, 2009: Ensemble Regression. *Mon. Wea. Rev*., **137**, 2365–2379.

Wallsten, T. S., D. V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth., 1986: Measuring the vague meanings of probability terms. *Journal of Experimental Psychology*, **115**, 348-365.

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

Wei, M, Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the Ensemble Transform (ET) technique in the NCEP global operational forecast system. *Tellus*, **60A**, 62-79.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences* (2nd Edition). Cambridge Press, Academic Press, 627 pp.

Wilks, D. S., and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev*., **135**, 2379-2390.

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.

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, K. K. Droegemeier, X. Wang, J. Kain, S. Weiss, D. Bright, M. Coniglio, and J. Du, 2009: CAPS realtime multi-model convection-allowing ensemble and 1-km convection-resolving forecasts for the NOAA Hazardous Weather Testbed 2009 Spring Experiment. 23nd Conf. Wea. Anal. Forecasting/19th Conf. Num. Wea. Pred., Omaha, NB, Amer. Meteor. Soc., Paper 16A.2.