AMS Ad Hoc Committee on Uncertainty in Forecasts (ACUF)

Update on

Enterprise Strategic Implementation Plan for Generating and Communicating Forecast Uncertainty

Tom Hamill

for

Paul Hirschberg and Elliot Abrams, Co-Chairs

17 September 2009, NCAR Mesoscale Ensemble Workshop
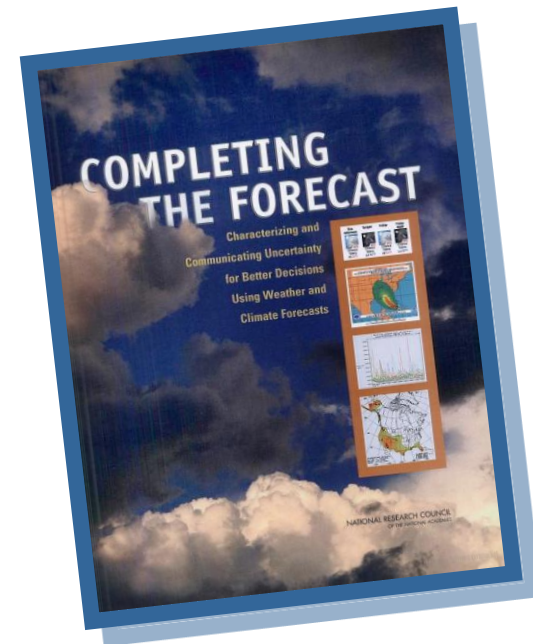
Background
ACUF Charge

- **ACUF formed by AMS Commission on Weather and Climate Enterprise, Board on Enterprise Communication**

- **Charged to formulate a cross-Enterprise plan to provide forecast uncertainty information to Nation**

- **2006 NRC Report - Completing the Forecast**

  "… confluence of compelling reasons for Enterprise to transition to a new paradigm … in which uncertainty is integral and essential component of all forecasts."

- **2008 AMS Statement - Enhancing Weather Information with Probability Forecasts**

  "Ideally, all weather forecasts would include information that accurately quantifies their uncertainty."

- **Purpose of ACUF Plan – guide for Enterprise partnership (following NRC Fair Weather Report) – not prescriptive program plan**
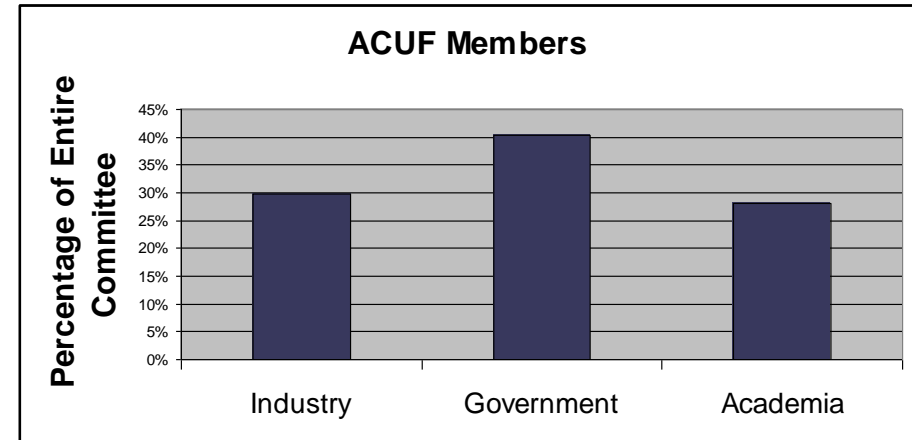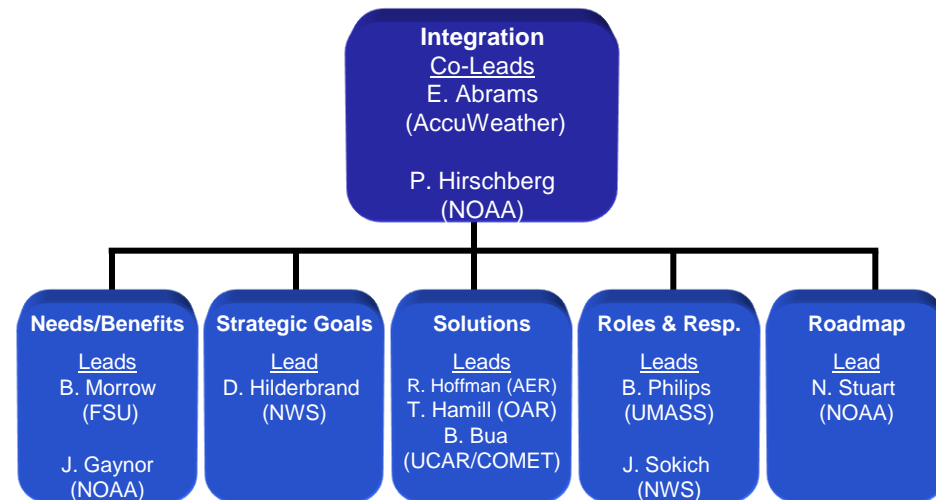
Background
ACUF Demographics and Organization

- **>60 volunteers from across Enterprise**

- **Organized into Topic Workgroups along plan components**

ACUF Members

Bar chart — Percentage of Entire Committee:
- Industry: 30%
- Government: ~40%
- Academia: ~28%

Integration
Co-Leads
E. Abrams (AccuWeather)

P. Hirschberg (NOAA)

Needs/Benefits	Strategic Goals	Solutions	Roles & Resp.	Roadmap
Leads	Lead	Leads	Leads	Lead
B. Morrow (FSU)	D. Hilderbrand (NWS)	R. Hoffman (AER)	B. Philips (UMASS)	N. Stuart (NOAA)
J. Gaynor (NOAA)		T. Hamill (OAR)	J. Sokich (NWS)	
		B. Bua (UCAR/COMET)		

Background
Not Starting from Scratch

- **NRC Report**

- **AMS Statement**

- **WMO Guidelines**

- **THORPEX results and plans**

- **Related literature -- physical science, socio-economic, behavioral, …**

- **Best Practices and Plans**

  - *National / International hydromet services*

  - *Industry*

  - *Academia and Labs*

Ten-day Temperature and Precipitation Probability Forecast for Trondheim, Norway

Temperature Spread

Precipitation Spread

Mean + Spread (50%+30%)
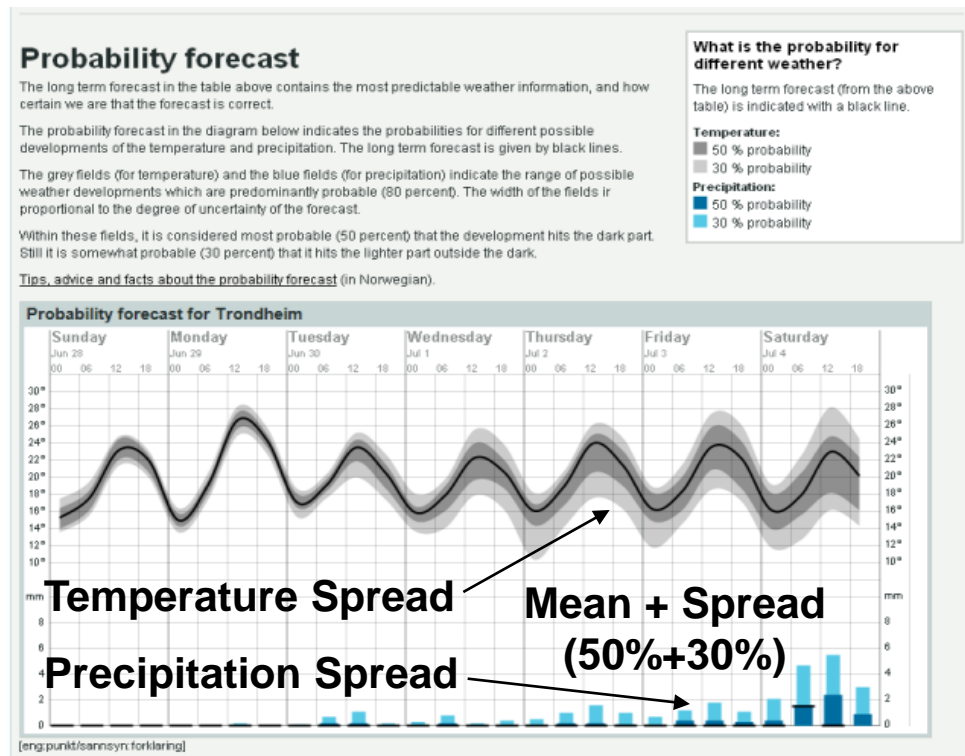
from Norwegian Meteorological Institute

Draft Plan
Organization

- **Use and Benefits of Forecast Uncertainty Information**

- **Vision and Strategic Goals**

- **Enterprise Partners' Roles and Responsibilities**

- **Implementation Roadmap**
  - *Objectives*
 - *Performance Measures*
  - *Tasks*

- **Monitoring Progress**

Benefits

Vision

Strategic Goals

Roles & Responsibilities

Objectives

Performance Measures & Targets

Tasks

Enterprise must build capabilities in four key, interrelated strategic areas to meet challenges of increasing focus on probabilistic forecasting

- **Understanding** nature of forecast uncertainty -- how to quantify it; how societal and human factors influence communication and use of uncertainty information

- **Generating** a reliable foundational suite of uncertainty data, products, services, and information needed by user communities;

- **Communicating** uncertainty information effectively, and **Assisting** users in interpreting and applying information in decision making

- **Enabling** development, acquisition, and operation of forecast uncertainty processing and communication systems with necessary computational, telecommunications, and other types of infrastructure.
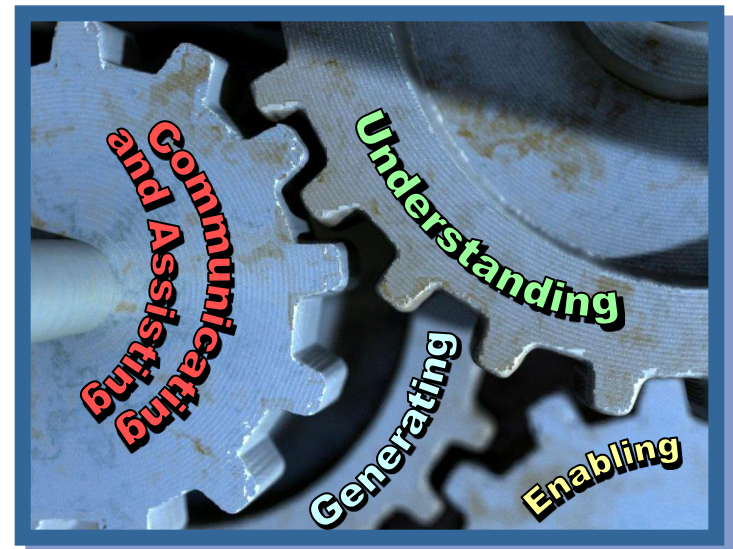
Draft Plan
Recommended Enterprise Partners' Roles and Responsibilities

- **Enterprise consists of four primary sectors:**

    1) Government (local, state, and federal governments)

    2) Commercial (companies and media)

    3) Academia (universities and associated research institutions)

    4) Non-Government Organizations (NGOs)

- **Distinct leadership roles for each partner group to fill**

- **But, expertise and resources of all sectors need to be leveraged to bring uncertainty into mainstream of decision making**

Draft Plan
Roadmap Overview

- **Roadmap contains set of objectives to accomplish four strategic goals**

- **Each objective includes:**
  - *Background*
  - *Need*
  - *Current capabilities and gaps*
  - *Performance measures and targets*
  - *Proposed solution strategy, specific tasks and sector leads.*

- **Tasks include specific actions for short-, mid-, and long-term periods over next decade**

Draft Plan
Roadmap Objectives

Strategic Goal 1:
Understand Forecast Uncertainty

- **Predictability**
- **Theoretical basis and optimal design of uncertainty prediction systems**
- **Best methods for communicating uncertainty**

Strategic Goal 2:
Generate a Foundational Set of Uncertainty Data and Information

- **Ensemble prediction systems**
- **Probabilistic nowcasting systems**
- **Statistical and not-statistical post-processing techniques**
- **Probabilistic forecast preparation and management systems**
- **Forecaster and other service provider training**
- **Probabilistic verification systems**
- **Digital probabilistic forecasts in Weather Information Database**

Draft Plan
Roadmap Objectives

Strategic Goal 1:
Understand Forecast Uncertainty

- **Predictability**
- **Theoretical basis and optimal design of uncertainty prediction systems**
- **Best methods for communicating uncertainty**

Strategic Goal 2:
Generate a Foundational Set of Uncertainty Data and Information

- **Ensemble prediction systems**
- **Probabilistic nowcasting systems**
- **Statistical and not-statistical post-processing techniques**
- **Probabilistic forecast preparation and management systems**
- **Forecaster and other service provider training**
- **Probabilistic verification systems**
- **Digital probabilistic forecasts in Weather Information Database**

Draft Plan
Roadmap Objectives, (cont.)

Strategic Goal 3:
Communicate / Assist Users Interpret and Apply Uncertainty Information

- K-12, undergraduate and graduate education
- User education and outreach
- Presentation uncertainty forecast products
- Decision assistance tools and services
- User evaluation and feedback

Strategic Goal 4:
Enable with Supporting Infrastructure

- High performance computing
- Adequate telecommunications
- Comprehensive archive and data access
- Research-to-operations and applications policies and processes and test beds
- User infrastructure needs

Some relevant objectives

Objective 1.2: Develop the theoretical basis for and optimal design of uncertainty prediction systems

Background: There are two sources of error that must be accounted for properly when generating probabilistic forecasts, initial-condition error and model error. Methods for generating initial conditions for ensemble modeling systems are mature relative to methods for dealing with errors in the forecast model. For example, there is now a general theory on what are the appropriate properties of initial conditions for ensemble forecasts. Building and testing ensemble Kalman filter techniques for improving data assimilation and initializing ensemble forecasts reflects an attempt to move from more ad-hoc approaches of initializing ensembles to ones that are consistent with such theoretical principles. However, there is not a similar underpinning for the treatment of numerical model error in ensembles.

Need: A fuller understanding of the sources of forecast uncertainty are needed as well as efficient numerical methods for estimating uncertainty in prediction systems.

Current Capability: Current approaches to estimate the effects of uncertainty in ensemble prediction systems include: (1) design of improved methods for initializing ensembles, e.g., ensemble Kalman filters. (2) reducing model error via improving the numerical models used in ensembles and/or by increasing ensemble model resolution, (3) estimating the uncertainty introduced by model error by using multiple models, and (4) introducing random elements into the integration of the ensemble.

Capability Gaps: Current ensemble prediction systems do not yet incorporate effective, theoretically justifiable methods for quantifying the effects of model errors in ensemble systems. Further, the relative tradeoffs between applying extra computational resources to add more ensemble members vs. increase model resolution vs. improve model numerics is not well understood. For example, more ensemble members decreases the sampling error associated with spread estimates, but if those more members are from a lower-resolution model, there may be a systematic underestimate of the spread.

Performance Measures and Targets:
• Peer-reviewed research publications that provide useful guidance on advanced methods for addressing initial-condition and model error in probabilistic predictions.
• Experimental systems at academic institutions and government labs that demonstrate the effects of new prediction methods.

Solution Strategy: Perform research on the underlying theory and optimal design of probabilistic prediction systems. Perform systematic studies of the theoretical underpinnings of the treatment of numerical model error in ensembles. Develop and demonstrate techniques that address model error that is consistent with theory.

Short-term (0-2 years):
• Continue research supported by existing research grants programs, primarily. *(Lead, academia, sponsored by US Navy, NOAA, NSF, and NASA)*

Medium-term (2-6 years):
An expanded research program on improved numerical techniques for estimating uncertainty. For improving the generation of ensembles, the basic research may follow a general sequence, such as: (a) identify a particular deficiency in existing ensemble forecast systems; (b) generate a hypothesis for how this deficiency could be ameliorated; (c) develop new techniques in accordance with this hypothesis, (d) perform experiments, comparing the new technique alongside existing techniques to determine if the new technique is superior. (e) If the new technique is superior (i.e., hypothesis confirmed), work with operational model developers (see Obj 2.2) to implement the technique, otherwise refine the hypothesis and repeat. *(Lead, academia, supported by government grants)*

Long-term (6+ years):
• Sustain predictability studies using improved, higher-resolution models. *(Lead, academia, supported by government grants)*

Some relevant objectives

Objective 2.1: Improve the initialization of ensemble prediction systems.

Background: Ensemble prediction systems require an ensemble of initial conditions. Ideally, these comprise a sample from the distribution of possible analysis states, and reflect the flow-dependent uncertainties due to the synoptic conditions and the distribution of observations, past and present. Within the last few years, ensemble Kalman filter (EnKF) methods have been developed that, for the first time, provide an ensemble of initial conditions that are theoretically justifiable, unlike past methods such as "breeding" and "singular vectors." However, EnKF methods are just beginning to be tested in a quasi-operational setting.

Need: An ensemble of initial conditions that are accurate, that sample the range of possible true analysis states (there is inevitably uncertainty in the analyses, especially at the mesoscale), and that project upon growing forecast structures so that differences between member forecasts grow (appropriately) quickly.

Current Capability: NCEP currently runs an "Ensemble Transform" (ET) technique, an update to their "breeding" technique. Other centers still use breeding or singular vector techniques. The Canadians are the only ones running an operational ensemble Kalman filter.

Capability Gaps: The existing sets of initial conditions are typically designed primarily to grow quickly, but in doing so do not accurately reflect the flow-dependent analysis uncertainty. Sets of initial conditions that both grow quickly and which correctly sample analysis uncertainty are needed in order for ensemble prediction systems to produce more realistic uncertainty forecasts at all leads. While NCEP continues to pursue 4D-Var, the 4D-Var strategy will only produce a single best-guess analysis, not provide an ensemble.

Performance Measures and Targets:
(1) Reduced error of deterministic forecast from the ensemble-mean initial condition relative to deterministic forecast from current alternatives such as 3D-Var. Target is gain of at least 12-h lead time in precipitation and hurricane track forecasts by 2014.
(2) Improved uncertainty forecast scores (flatter rank histograms, sharper, more reliable forecasts, increased skill, especially for high-impact weather). Together with improvements provided by objective 2.2, the target is gain of at least 24-h lead time in skill of high-impact weather (heavy precipitation, hurricane track, etc.) by 2014.

Solution Strategy: Develop EnKF (and potentially 4D-Var/EnKF hybrid) ensemble data assimilation assimilation techniques, including improved methods for the treatment of model error in ensemble filters.

Short-term (0-2 years):
• Quasi-realtime tests during hurricane season of an EnKF using a global forecast model. EnKF data assimilation will be performed with a model with 30-km grid spacing using 60 members, and 10-member forecasts from a 15-km global model will be tested. Data will initially be evaluated for 1-week hurricane track and intensity forecasts and compared with operational GFS ensemble products. After the fact, this ensemble can be evaluated for its skill with other high-impact events, e.g., heavy precipitation.
• Continue R&D on improved methods for the treatment of model error and sampling error in ensemble filters.
• Explore methods for hybridization of variational and EnKF methods.

Medium-term (2-6 years):
• Implement EnKF into operations at NWP facilities.
• Evaluate EnKF relative to 4D-Var for its ability to produce reduced-error initial conditions.
• Develop hybrid 4D-Var / EnKF methods.

Long-term steps (6+ years):
• Test and implement hybrid 4D-Var/EnKF methods.

Some relevant objectives

Objective 2.2: Improve operational ensemble prediction systems

Background: Ensemble prediction systems are core technology for producing uncertainty forecasts. Multiple forecasts are run from slightly different initial conditions, and sometimes using different models or parameterizations. They produce an ensemble of possible future weather scenarios. They need to be improved.

Need: An ensemble of future weather scenarios that, without much statistical post-processing, provide a reasonably sharp and reliable probabilistic forecasts.

Current capabilities: Current-generation ensemble prediction systems in the US produce uncertainty forecasts that are biased and underestimate the forecast uncertainty. Partly this is because of the low-resolution of the forecast models, partly because of improper initial conditions (see objective 2.1) and partly because the ensemble prediction systems do not include effective treatements for the error introduced by model deficiencies. Currently, NCEP runs global and regional ensemble prediction systems. The global ensemble currently run at T126 (approximately 125 km grid spacing) with 20 members, 4x daily. For comparison, ECMWF runs a 50-member global ensemble at T399, greater than 3 times the resolution of the NCEP global ensemble. NOAA ESRL is currently evaluating a new global model on an icosahedral grid (the ṆHMÓ model). NCEP also runs a regional, short-range ensemble with 21-members, 87-h forecasts with a model with ~32-km grid spacing. NOAA has experimentally evaluated a 10-member small-domain ~4 km grid spacing ensemble for severe weather applications on the Great Plains. NOAA also has a short-range, higher-resolution lagged ensemble available over the CONUS from the Rapid Update Cycle, which generates a new forecast every 1h. By 2010 this system will use the WRF model with a 13-km grid spacing. The US Navy and Air Force also have their own suites of operational forecast models, and some university run quasi-operational ensemble systems (e.g., the University of Washington). Recent results from using the THORPEX Interactive, Grand Global Ensemble also indicates that more skillful, reliable forecasts may be possible through the sharing of global ensemble forecast data.

Capability Gaps: The raw output from existing ensemble forecast systems do not provide reliable probabilistic forecasts in all circumstances. There are several root causes, including: (1) systematic model errors, grid resolution, and parameterizations of sub-gridscale physical processes (deep convection, boundary layer, surface layer, land surface, microphysics) often provide unrealistic estimates of the effects of the unresolved scales of motion, and their uncertainty. (2) poor initialization of the ensemble with model states that do not properly represent the uncertainty in the analysis.

Performance Measures and Targets:
- Comparisons with the much higher resolution ECMWF ensemble forecasts suggest that such improvements should result in a general 1-day improvement in forecast lead (i.e., a 5-day forecast as skillful as a current 4-day forecast), with greater improvements for many severe-weather phenomena. Accordingly (see also objective 2.1) the target is gain of at least 24-h lead time in skill of high-impact weather (heavy precipitation, hurricane track, etc.) by 2014.

Solution Strategy: Refine ensemble model grid resolution dramatically in the next few years (followed by regular resolution improvements thereafter); incorporating research results from Objective 1.2 (basic research in ensemble prediction) into the operational systems; greater sharing of forecast data between operational facilities to create multi-model ensemble products; and adding a new, limited-area, high-resolution, high-impact event regional ensemble system.

Short-term(0-2 years):
- Share global ensemble forecast model output among Federal agencies (e.g., NOAA, Navy, Air Force) and governments (Canada and others) through projects such as the North American Ensemble Forecast System (NAEFS), National Unified Operational Prediction Capability (NUOPC) and TIGGE/GIFS (THORPEX Interactive Grand Global Ensemble/Global Integrated Forecast System). (Lead: Government)
- Develop and test higher-resolution global ensembles, with improved model physics and initialization techniques (Lead Government)
- Develop and test higher-resolution, improved physics short-range ensembles (Lead: government and Academia)
- Work with extramural basic researchers and lab partners that have developed improvements in experimental ensemble forecast systems to test these ideas in operational environments (Lead Government)
- Sustain development and testing of new ensemble initialization techniques (e.g., ensemble-based data assimilation methods, or the use of ensemble information in variational methods) include (Lead: Government and Academia)
- Develop and test improved hydrologic ensemble forecast system models (Lead: Government and Academia)
- Explore product development from lagged-average ensemble forecasts from the Rapid Update Cycle and the WRF Rapid Refresh (Lead: government and Academia)

Medium-term (2-6 years):
- Implement higher-resolution ensemble model systems. To provide state-of-the-art forecasts, U.S. Enterprise (NOAA, DoD) should increase resolution threefold by 2011-2012. (Lead: Government and Academia)
- Implement the best ensemble initialization technique from a range of competitors (Lead, Government)
- Increase SREF system resolution threefold, to 10-20 km grid spacing by 2011-2012 with hourly output. (Lead: Government and Academia)
- Develop an advanced method for perturbing the initial conditions that is consistent with the strategy for the GEFS but which also includes realistic mesoscale differences in the initial and lateral boundary conditions. (Lead: Government and Academia)
- Develop and implement a relocatable, high-resolution, explicit convection (~4-km grid spacing) limited-area ensemble forecast system for hurricanes, severe local storms, and fire weather that can be nested inside the SREF or GEFS systems, and appropriate methods for initialization. (Lead: Government and Academia)
- Explore the usage of a high-resolution, lagged ensemble forecast system

Some relevant objectives

Objective 2.3: Develop and Implement Probabilistic Nowcasting Systems

Background: The accuracy of the first few forecast hours of classical numerical model guidance, including ensemble guidance, is often reduced as these complex models develop internally consistent vertical motions. Beyond forecast leads of several hours, the need for alternate forecast techniques is obviated as this so-called model "spin-up" process completes and ensemble systems are able to provide useful forecast uncertainty information.

Need: New techniques are needed to generate reliable probabilistic forecast information for forecast lead times of zero to several-hours. This includes developing approaches for adding uncertainty estimates to current nowcast systems and variably weighting nowcast output with classical ensemble-based forecast products as a function of forecast lead.

Current Capability: Several research groups have developed observationally based tools for making nowcasts. For severe weather, these tools involve the detection of significant features using radar or other remotely sensed data; the extrapolation of this data using wind or derived motion vectors; and perhaps some development/decay mechanism (e.g., the NCAR/RAL AutoNowcaster). Another nowcast system in the western US makes short-range extreme precipitation forecasts using expected wind and moisture-flux information and thermodynamic stability, estimating how much precipitation may occur as a column interacts with the complex terrain along the US west coast. As demonstrated with these two examples, typically the nowcast tools may be application- and location-specific. However, little work has been performed to date to develop *probabilistic* nowcast tools; most have a more deterministic cast to them, though there is some scant literature on the application of Bayesian statistical methods to nowcasting.

Capability Gaps: The NRC report recommends the inclusion of uncertainty information throughout the spectrum of forecast products, including those at the shortest leads. In the near, term the techniques that are appropriate at 24-h lead may not be appropriate at 2-min or 2-h lead, so development of probabilistic nowcast techniques is required.

Performance Measures and Targets:
•

Solution Strategy: Explore the "dressing" of deterministic forecasts with samples of errors to produce an ensemble of realizations from which probabilities can be estimated. For example, Nehrkorn and Hoffman (2006) investigate a dressing technique based on phase errors, which might be particularly appropriate for nowcasting applications. Also develop enhanced nowcasting techniques that rely on recent observed data. Perhaps tools like the AutoNowcaster can be modified to permit other discrete realizations based on realistic changes to the estimated propagation speed or rate of development/dissipation. In this manner, an ensemble of possible future scenarios may be obtained. Perhaps statistical post-processing techniques (see Obj 2.4) can be used to remove systematic biases in the output. It is unlikely that there will be any one-size-fits-all nowcast scheme; perhaps different methods will be developed tailored to different customers and different locations. Develop tools for blending together the observationally-based nowcast and numerically-based NWP guidance as forecast lead increases. At the earliest leads, nowcasts would be heavily weighted, and at longer leads the NWP guidance would receive increasingly larger weight. Weights will be adjusted as NWP improves for short leads and become more useful than observationally-based approaches.

Short-term (0-2 years):
• Accumulate (what kind of?) error statistics and use to produce an ensemble of realizations from which probabilities can be estimated. *(Lead: Government)*
• Modify current deterministic nowcast tools to ??? *(Lead: ???)*

Medium-term (2-6 years):
• Develop enhanced nowcasting tools that rely heavily on recent observed data (Lead: ??).
• Develop tools for blending together the observationally-based nowcast and numerically-based NWP guidance as forecast lead increases. (Lead: ??)

Long-term steps (6+ years):
• Based on improvement of data assimilation and numerical weather prediction systems, decrease emphasis on separate nowcasting tools. (Lead: ??)

Objective 2.4: Improve statistical post-processing techniques.

Background: Statistical post-processing refers to the adjustment of the current operational forecast using relationships derived from past forecast and observed/analyzed data. These techniques are applied in order to ameliorate the deficiencies in the raw model output, which, in the case of ensemble prediction systems, are typically manifested in biased, overly sharp uncertainty forecasts. Statistical post-processing can also Ndownsale Ó the coarse-resolution output to the fine detail available at an observation location or to a high-resolution analysis.

Need: Uncertainty guidance must be reliable and skillful in order to be widely used and accepted. Uncertainty estimates from unadjusted ensemble models are liable to be sub-optimal due to the small ensemble size, the biases inherent in forecast model(s) and methods used to initialize the ensemble, and the comparatively coarse grid spacing used in current ensemble prediction systems. While improved ensemble prediction methods (see Objective 2.2) may reduce the need for post-processing, they will never completely eliminate it. A comprehensive program is thus needed to determine the optimal calibration techniques across the spectrum of high-impact applications, develop routine methods for computing the supporting data sets, and implementing the calibration techniques. Additionally, information of interest to users (e.g., some esoteric variables at a particular point in space and time) may be different from the information generated by numerical models (i.e., a grid-box average of standard NWP variables). Therefore numerical model output, even if bias free, typically requires some sort of NdownscalingÓ to maximize utility.

Current capabilities: Dissemination of statistically based guidance in the NWS has a long history as NModel Output StatisticsÓ(MOS). MOS uses regression relationships based on prior forecasts and observations to statistically adjust current numerical guidance. These regression relationships correct for model bias and can implicitly perform a statistical downscaling, adjusting the sub-gridscale weather to be colder and wetter inside the mountain grid box, for example. Most existing MOS products are deterministic, though implicitly the MOS regression analyses produce uncertainty forecast information, and this guidance could be disseminated with little additional effort as a baseline uncertainty product. More recently, with the advent of ensemble-prediction techniques, a variety of new extensions to the basic MOS techniques have been developed that leverage the extra information in the ensemble. So-called NEnsemble-MOSÓtechniques may improve estimates of the most likely outcome in situations where the ensemble mean provides a more accurate estimate than a deterministic forecast (this characteristic is common at leads longer than a day or two). If the ensemble spread provides information on the situation-dependent uncertainty (high spread is related to lower forecast skill, low spread to higher skill), then ensemble-MOS techniques may provide improved, weather-dependent uncertainty estimates as well. A variety of ensemble-based calibration techniques are currently being developed and tested at many NOAA facilities (for example, MDL, NCEP, ESRL, and OHD), and at many universities. NCEP has its own bias-correction technique running operationally on NAEFS forecasts.

For every-day weather such as surface temperature at short forecast leads, a variety of calibration techniques appear to perform relatively competitively, and small training samples appear to be adequate. For rare events and long-lead forecasts, more training samples (perhaps many years or even decades of NreforecastsÓfrom a stable model/data assimilation system) are needed, and the calibration technique of choice may depend upon the particular application and data at hand. The need for large training data sets from a stable model conflicts with the desire to rapidly implement improvements in ensemble forecast systems. There are two possible compromises: (1) maintain one model that is rapidly upgraded but unaccompanied by reforecasts and another model that is only

Solution Strategy: Continue ongoing efforts to improve statistical post-processing of ensemble model output and pursue systematic approaches to optimize ensemble calibration techniques.

Short-term (0-2 years):
- Define which observational data set(s) or time series of analyses should be used for testing techniques. (*Lead: Government*)
- Continue to test a variety of algorithms for operational correction of ensembles, including Ensemble Kernel Density MOS and stepwise calibration through bias correction at the large scales and a downscaling. (*Lead: Government*)
- Continue research into the optimum reforecast training size (*Lead: Government*)
- Develop roadmap for statistical post-processing. (*Lead: Government*)
- Implement a single-member reforecast into real-time ensemble forecast system, with reforecasts calculated over several decades and a temporal frequency of approximately once per week. (*Lead: Government*)

Medium-term (2-6 years):
- Develop the robust reforecast and observational/analysis training data sets to perform statistical post-processing studies, including high-resolution analyses. (*Lead: Government*)
- Survey the literature on existing calibration techniques for each of the parameters, as well as requirements the supporting data sets required. (*Lead: Academia*)
- Perform an objective comparison of the techniques for each forecast variable to determine which is the most appropriate. If none of the existing calibration techniques appear to be appropriate, develop and test new, more appropriate calibration techniques. If the performance of the calibration technique strongly depends on the training sample size, determine which compromises in reforecast sample size (e.g., shrinking the number of reforecast members, or skipping days between reforecasts) lead to the least degradation. (*Lead: Government*)
- Evaluate the prototypes of calibrated forecast products (see Obj 2.6) using standard verification tools (see Obj 2.4). (*Lead: Government*)
- Implement the calibration techniques. (*Lead: Government*)

Long-term steps (6+ years):
- Reforecast data sets will need to be computed regularly with the then operational model (*Lead: Government*)
- New customers may be expected to request new products for less standard variables, which may require the development of specific calibration approaches. (*Lead: Government*)

Some relevant objectives

Objective 2.8: Develop and implement probabilistic verification systems

Background: *The weather enterprise is used to evaluating models and weather forecast with deterministic metrics. Once forecasts are expressed with uncertainty information, they will need to be evaluated using metrics that quantify how well they do in making uncertainty forecasts.*

Need: The enterprise needs a comprehensive, agreed-upon set of standards and software algorithms for uncertainty verification. Currently, forecast verification methods focus on verifying the best single-value estimate. Probabilistic forecast verification techniques must be developed and/or applied that will assess the characteristics of uncertainty forecasts and provide quantitative feedback to ensemble developers, forecasters, service providers, and end users to aid in interpretation and decision-making. Statistics generated from these techniques are needed to serve as a reference for user expectations, guide future improvements, and assess the value added during each step of the forecast process.

Current capabilities: A variety of verification methodologies are currently used to assess forecast uncertainty products. These include measures of reliability such as the rank histogram and reliability diagrams, and measures that incorporate probabilistic forecast accuracy such as the Brier Score, Ranked Probability Skill Score, and the Relative Operating Characteristic, and end-user relevant metrics such as Potential Economic Value diagrams.

Capability Gaps: There is no universally agreed-upon set of metrics that provide a comprehensive diagnosis of forecast uncertainty. Intercomparisons are currently complicated by the use of different data sets by different model developers, verification on different grids, and comparisons of ensembles of different sizes. Further, while metrics have been developed that are appropriate for assessing some aspects of uncertainty forecasts, such as their reliability, other aspects such as event timing and forecast covariance in time and space are not standardized, or even developed. Another problem is that even for the basic uncertainty statistics, their availability varies widely. For operational forecasts, a common but not universal practice is to post verification scores to a web page. Verification information from non-governmental forecast producers typically is difficult to obtain.

Performance Measures and Targets:
- Publication of an uncertainty verification manual (2012).
- Version 1.0 of uncertainty forecast verification software library (2013)
- First results from verification clearing house (2014).

Solution Strategy: Form a group of experts to develop forecast uncertainty verification standards and best practices. Develop and freely disseminate a software library of verification and verification display procedures appropriate to forecast uncertainty products. Implement a "verification clearing house" where forecast uncertainty verification statistics are collated and displayed

Short-term (0-2 years):
- Continue ongoing research into new verification methods.
- Develop prototype probabilistic verification packages for particular applications, such as for aviation forecasts.
- Note: World Meteorological Organization activities may coalesce in the next two years around standards for the verification through its joint working group on verification and THORPEX/TIGGE working groups. *(Lead: Government)*

Medium-term (2-6 years):
- Form a team of uncertainty verification experts to: (i) identify a standard, scientifically defensible set of methodologies for the evaluation of uncertainty forecast products, (ii) evaluate new metrics and determine whether they should become part of the baseline, and (iii) based on (i) and (ii), develop and publish an "uncertainty verification manual" that specifically indicates how each metric is to be computed.
- Periodically thereafter, this team will meet to review new verification developments and determine whether the manual should be expanded or changed.
- Software developers will build a standardized library of routines based on the specifications in the uncertainty verification manual, as well as software for the convenient display of verification data. This software will be made publicly available as freeware.
- Relevant parts of the weather enterprise (model developers, forecasters, applications developers) use the toolkit for verification, and make the verification data publicly available.
- Institute a verification "clearing house" where the synthesized verification results from various components (raw ensemble guidance, objectively post-processed guidance, official forecasts, and end products) are all gathered in one place and made available. *(Lead: Government)*

Long-term steps (6+ years):
- Users are likely to continue to request the development and testing of new uncertainty products. *(Lead: Government)*
- Accordingly, new verification procedures specific to these products will be needed. *(Lead: Government)*.

Discussion
Next Steps

- **Complete Version 0.5 of Plan (Aug. 09)**

- **Full Committee Review – Ver. 1.0 (Sept. 09)**

- **Stakeholder Review – Ver. 2.0 (Nov. 09)**

- **AMS Town Hall (Jan. 10)**

- **Final draft (Feb. 10)**

- **AMS Leadership Review (Mar. 10)**

- **Publication (Apr. 10)**

Conclusion

- **AMS Ad-Hoc Committee on Uncertainty in Forecasts is in middle of the process for laying out an enterprise-wide implementation plan.**

- **SREF work fits nicely into this implementation plan.**

- **The plan encompasses many other areas of concern; training, education, basic research, communication, infrastructure.**

- **Your reviewing and suggesting changes to this plan will be very much welcome.**

To Get More Involved

Please contact Andrea Bleistein
andrea.bleistein@noaa.gov