Completing the forecast: assessing and communicating forecast uncertainty

Zoltan Toth[1], Paul Schultz[2], Steve Mullen[3] Julie Demargne[4] and Yuejian Zhu[1]

[1] EMC/NCEP/NWS/NOAA),
[2] GSD/ESRL/OAR/NOAA)
[3] University of Arizona)
[4] OHD/NWS/NOAA)

"The wise person knows what s/he does not know" - Socrates

1. Summary

Over the years, uncertainty in weather, water, and climate forecasts has been continuously reduced through the use of ever improved forecast technology, but uncertainty, due to the chaotic nature of the atmosphere-land-ocean coupled system, can never be eliminated from environmental forecasts. For optimal decision making, users need to consider all possible future weather, water, and climate scenarios, not just the most likely outcome. A clear need, articulated by the "Completing the Forecast" report by the National Research Council of the US Academy of Sciences (NRC, 2006), exists to broaden the scope of environmental forecasting with the inclusion of a new dimension, forecast uncertainty. Assessing and communicating forecast uncertainty is important not only for users but also from a scientific perspective. Imperfect forecasts, as all weather, water, and climate predictions are, are statistically reliable (i.e., consistent with observations) only if given in a probabilistic format.

Forecast uncertainty can be assessed either through traditional approaches based on error statistics of a single forecast, or recently developed dynamical approaches which use an ensemble of forecasts. The traditional forecast process focuses on the estimation of the expected value of the forecast distribution. A single scenario is developed for the evolution of the future weather, water, or climate conditions, often via the use of a single numerical forecast. Forecast uncertainty estimates are based on a statistical analysis of errors in past forecasts. Limitations of this approach include (a) a sub-optimal estimate of the "most likely" forecast scenario due to a neglect or inappropriate consideration of alternative future scenarios; (b) a sub-optimal estimate or neglect of case-to-case variations in forecast uncertainty due to the exclusive use of statistical methods in assessing forecast uncertainty; and (c) difficulty in presenting dynamically consistent alternative scenarios for use in decision support systems. The first two issues act as factors limiting forecast skill, whereas all three curtail the utility of forecasts.

The new alternative approach to assessing forecast uncertainty is based on an ensemble of numerical forecasts and is designed to mitigate the limitations of the traditional forecast process. Ensemble methods are known to reduce errors in the expected value of a forecast distribution (i.e., best single value estimate), but in addition, case-dependent variations in forecast uncertainty are captured. Both of these advantages arise from the systematic consideration of alternative forecast scenarios. Case dependent forecast uncertainty estimates are generated via statistical post-processing (i.e., a combination of the dynamical and statistical methods). Ensemble forecasts are the centerpiece of a proposed new forecast process where uncertainty is assessed, propagated, and conveyed throughout the entire end-to-end forecast process in a self-consistent manner. The systematic use of ensembles, supported by extensive verification results, can maximize both the skill and

utility of the forecasts, as well as facilitate the adaptive configuration of the entire forecast process to accommodate time-varying user requirements in high impact events. The transition from the old to the new forecast process requires concerted efforts on the parts of both the producers and users of the forecasts.

2. The importance of forecast uncertainty information

Weather, water, and climate forecasts are imperfect in the sense that they exhibit errors when compared to the verifying observations. Typically, the longer the lead time, the more uncertainty there is. This is primarily due to the chaotic nature of the atmosphere and the coupled atmosphere-land-ocean system. Errors in the initial state of a chaotic system and of the predictor algorithm (i.e., numerical model) are bound to accumulate and grow in time until all useful forecast information is lost.

Naturally, users always prefer weather, water, and climate forecasts with reduced or possibly no uncertainty. The level of uncertainty in a forecast, however, cannot be reduced arbitrarily. The minimum attainable level of forecast uncertainty at any time depends on the objective quality of the forecast system used to make predictions. Traditional approaches to improving forecast systems have mainly focused on the reduction of forecast errors, but the user community is presenting new requirements to know about the level of uncertainty in the forecasts prepared for them.

In general, rational decisions regarding future actions are typically made by reviewing all possible scenarios. For optimal decisions a user must consider all possible future weather/water/climate scenarios, along with their likelihood and the incremental cost or value to that user associated with those possible outcomes (see, e.g., Zhu et al. 2002). The decision that leads to the most benefits, while incurring the lowest risk, is chosen. Losses can be incurred if a user does not prepare for any scenario other than what was predicted. Therefore a well defined need exists for making uncertainty information available to interested users.

From a scientific perspective, single value forecasts can be consistent with the distribution of verifying observations only if the forecasts have no error in them (i.e., perfect forecasts) and hence all verifying observations coincide with their corresponding forecast values. It follows that in the presence of uncertainty, probabilistic expression is required for statistical consistency.

STATISTICAL CONSISTENCY

Statistical consistency (or reliability) is one of two major attributes of forecast quality. In general terms, statistical consistency is defined as a forecast having the same form (i.e., the same distribution) as the distribution of observations that follow after the same forecast is issued over many occasions. Reliability means that what is being forecast over many occasions is observed in a statistical sense (for more details, see Toth et al., 2006). Reliability in itself does not guarantee informativeness. Predicting the climatological distribution all the time guarantees perfect reliability, yet provides no useful forecast information. Statistical consistency can be measured by a number of specific scores.

3. ASSESSING FORECAST UNCERTAINTY

Forecast uncertainty can be quantified in a variety of ways. The methods can be grouped into statistical and dynamical approaches. Today, most weather, water, and climate predictions are based on numerical forecasts. Traditionally, a single numerical integration is carried out. These single value forecasts have errors, and are thus statistically inconsistent with the observations. However, they can be made reliable and forecast uncertainty can be provided via statistical methods related to bias correction.

STATISTICAL APPROACH TO ASSESSING FORECAST UNCERTAINTY

The statistical approach to assessing forecast uncertainty is based on an evaluation of the relationship between identical (or very similar) forecasts and the statistical distribution of the ensuing verifying truth. The method involves the construction of a statistical frequency distribution based on verifying observations after the same (or similar) forecast was issued many times. The form of the original forecast is then replaced with the observed frequency distribution. Uncertainty in forecasts of any form (i.e., single value, categorical, or probabilistic) can be assessed using this method, given that a large enough stationary sample of forecasts and corresponding proxy for truth exists. In practice, observations, or information derived from observations is used as a proxy for truth.

A numerical forecast used in the traditional forecast process represents a single realization out of countless other equally plausible scenarios. Unlike statistical methods that assess uncertainty in an average sense, dynamical methods capture variations in forecast uncertainty that are intrinsically related to the natural processes in the coupled atmosphere-land surface-ocean system. For example, forecast uncertainty is increased in the presence of instabilities such as around fronts and convective systems.

DYNAMICAL APPROACH TO ASSESSING FORECAST UNCERTAINTY

To quantitatively assess forecast uncertainty, a number of numerical forecasts (i.e., an ensemble) can be generated as a supplement to the traditional single integration, by slightly varying the initial condition and the model itself to reflect uncertainty in them (see Buizza et al., 2005). Each member of an ensemble of forecasts represents a plausible outcome, and as a set, they offer a sample of all possible scenarios. Case to case variations in the behavior of the ensemble members (e.g., the spread among the ensemble members) provide an indication of variations in forecast uncertainty (i.e., expected forecast error). A well constructed ensemble thus naturally captures forecast uncertainty. As of now, no other scientific method is known to provide similar quality forecast information for complex systems such as the atmosphere.

The reliability of ensembles, just like single value forecasts, can be enhanced by statistical bias correction methods.

IMPERFECT ENSEMBLES

Probability forecasts derived from ensembles are usually much more consistent with observations as compared to any single forecast. Nevertheless, probability forecasts derived by counting how ensemble many members predict a certain event are not necessarily consistent with truth. This is due to initial and model related forecast errors, and also due to the finite size of any ensemble. Ensemble forecasts, however, can be made consistent with truth using statistical techniques similar to those used for single value forecasts. For example, the spread of the ensemble can be adjusted in a climatological sense (e.g., increased in case the ensemble is found under-dispersive) or the probability of events can be corrected in other ways. Such statistical adjustments will improve reliability, while at the same time retain the ensemble’s ability to distinguish between low and high uncertainty cases, although forecast resolution (see below) will not be improved.

4. BENEFITS OF ENSEMBLE APPROACH

The use of ensembles for weather, water, and climate prediction is a relatively new concept. It may be useful to consider changing from a single to a large ensemble of forecasts as a continuum. A single forecast can in fact be considered as a one-member ensemble. As more members are added, the quality and value of an ensemble of forecasts, including both its reliability and resolution, continually increases, eventually approaching a value characteristic of an ensemble with very large membership.

Derivation of most likely scenario. A narrow focus on finding a good single value estimate of future weather, water, or climate conditions can become a limitation to the quality of such forecasts themselves. The evolution of weather, climate, and associated water conditions exhibit strong nonlinearities at all temporal and spatial scales, so the consideration of all possible scenarios is critical for arriving at the best estimate of the future state of the system (e.g., NRC 2006). This notion is supported by the fact that the mean of well constructed ensemble forecasts provides a better estimate of truth than any single forecast alone. Neglect or improper use of ensemble information in the forecast process leads to suboptimal estimates of the first moment of the forecast distribution; i.e., the best guess. Information contained in the nonlinear evolution of a cloud of ensemble forecasts cannot be recovered by any other means (see Fig. 1). Statistical post-processing can change the format of a forecast from single value to probabilistic, and can make such forecasts statistically reliable. Since the error in single forecasts is higher than that in an ensemble mean, the calibrated probability distribution, however will be wider compared to that with a forecast based on a well constructed ensemble, resulting in diminished predictive skill, easily detectable by measures of statistical resolution.

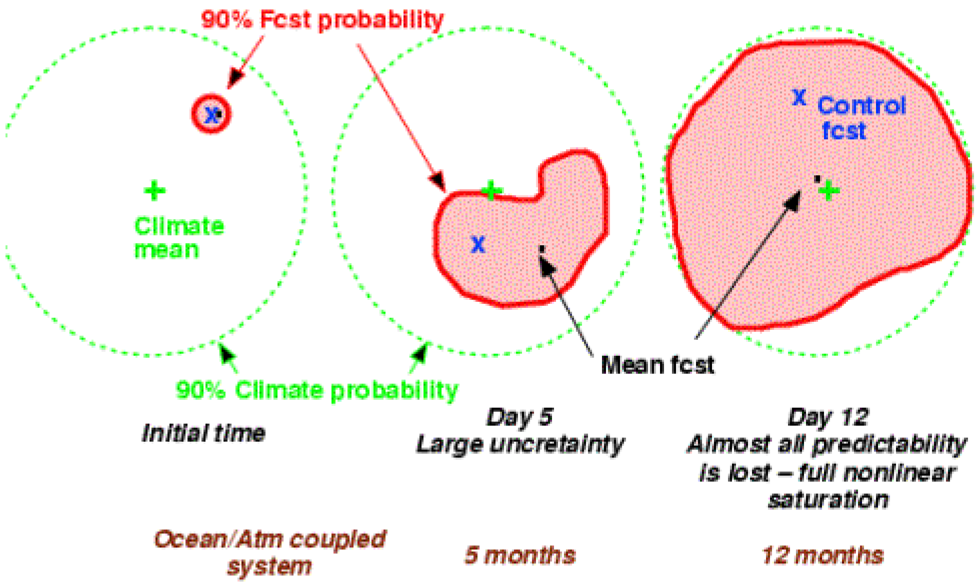

Figure 1: Schematic of a forecast of a chaotic system. The 90-percent probability contour and the mean of the climatological attractor in 2 dimensions are shown as a green circle and a plus sign, respectively. The 90-percent probability contour and the mean of an ensemble of dynamical forecasts at initial (left circle), and two later lead times (middle and right circle) are shown as red contours and a black dot, while a single control forecast from the best estimate of the initial condition are shown as a blue cross. Note that the mean of the ensemble cloud in general do not overlap with the control forecast.

STATISTICAL RESOLUTION

Statistical resolution is the second major attribute of forecast quality. Statistical resolution is the ability to discriminate, ahead of time, among cases that lead to different future outcomes. It can also be thought of as the degree to which different forecasts lead to unique outcomes. An ensemble mean forecast, having an error lower than a single forecast, would also have higher resolution because the ensemble as a predictor better sorts future outcomes than a single forecast does. Unlike statistical reliability, resolution cannot be improved by statistical bias correction. Statistical resolution captures the inherent predictive power of forecast systems, and can be measured using a variety of scores. Note that statistical reliability (statistical similarity between the form of a forecast and the distribution of verifying observations following such forecasts) and resolution (predictive power of a forecast system) are independent attributes of forecast systems (see, e.g., Toth et al. 2006).

Determination of forecast uncertainty. Using a single forecast followed by a statistical approach, forecast uncertainty is evaluated over a sample of forecasts in an average sense, and therefore the uncertainty estimate does not vary from case to case. The size of the available sample and the choice of the statistical method determine whether and to what extent regime-dependent variations in forecast uncertainty can be identified as opposed to providing just climatological forecast uncertainty. Extraction of maximum forecast uncertainty information requires case-dependent methods. The use of climatological (or even conditional climatological) average uncertainty estimates, as done traditionally, will lower predictive skill, detectable by measures of statistical resolution.

Consider a single forecast whose climatological uncertainty is assessed using a statistical method. Every time a particular forecast value is predicted, it is replaced with a probability distribution of a fixed width (i.e., the observed frequency distribution of the event predicted, given the forecast value). An ensemble that has skill in distinguishing between situations with low (small spread) and high forecast uncertainty (large spread) can be used to enhance the single forecasts by providing case dependent probability distributions for the single forecast. Such a system would, in general, have more value to the user community.

EXAMPLE

A construction company can lay concrete if the temperature stays above 3°C. Let us assume that every time the low temperature forecast is 5°C, the probability of the temperature dropping below 3°C is 20% in a climatological sense, based on the statistical approach described above. The 20% probability may pose too high a risk for the company, comparing the cost (C) they incur by losing profit for that day by skipping work, and the potential loss (L) if their concrete is ruined by low temperature. To maximize her/his profit, a user with a cost / loss ratio of say 15% will have to refrain from concrete laying work every time the single value forecast indicates a 20% probability (i.e., 5°C single value forecast). This is because to maximize the value of forecast information, users must use their C/L ratio as a decision probability threshold when deciding whether to act or not on a particular forecast (see, e.g., Zhu et al. 2002). If an ensemble method can separate those cases of 5°C forecasts when the probability of below 3°C temperature is below 15% (low uncertainty cases) from other cases when it is above (high uncertainty case), the user can economically benefit by carrying out their work when the probability of adverse weather is lower than average, given the 5°C single value forecast.

Forecast format. As the example above illustrates, a forecast in the typical format of a single value ignores the users' need for multiple decision thresholds depending on their cost / loss situation. A single value forecast is not suited for optimal decision making that depends on the specifics of each user's application. An ensemble of forecasts, on the other hand, can be used to formulate multiple probabilistic thresholds, allowing each user to select information related to their optimum decision level. In addition, ensembles can also be directly fed into automated decision support systems to determine the best course of action given all available information about the future environment.

There is also a less-obvious but very important aspect of expressing forecasts with quantitative uncertainty information included. Although current forecast products provided by the National Weather Service generally describe a single future scenario, forecasters usually know much more about the range of possible outcomes. This places the forecaster in a decision making role, answering specifically and perhaps uniquely to each user who is impacted enough to seek additional information from the forecaster. As forecasters are not adequately trained with respect to each user's specific decision situation, they may become distracted, potentially leading to reduced forecast quality.

SEPARATING FORECAST AND USER DECISIONS

The use of single value as the basic forecast format necessarily blurs the distinction between forecast and user decisions, leaving both the forecasting and user communities at unease. Providing only a single value (or categorical) format feeds some users' unrealistic expectation that uncertainty from forecasts can be entirely eliminated and that user decisions can essentially be transferred to the meteorologist who with the choice of a single value effectively determines user actions, too. By entirely focusing on a single scenario, forecasters are also kept away from a careful evaluation of alternative scenarios, essential for optimum decisions from the users' perspective. The situation can be ameliorated by the use of a probabilistic forecast expression. The designation of probabilistic information as the basic, internal forecast format requires the forecaster to formally evaluate and quantitatively convey forecast uncertainty; and it allows sophisticated users to formally assess the possible impact of various forecast scenarios on their operations. Forecasters and users can refocus their efforts on their respective areas of expertise, potentially enhancing the productivity of both communities. Less sophisticated users can, as before, be served with single value or other specific information they require, all derived from the same basic probabilistic information. This process ensures full consistency among various types of forecast information in an equitable way for all users.

5. Traditional forecast paradigm

To contrast the new probabilistic forecast paradigm described next with current practices, a brief synopsis of the traditional forecast paradigm is given first. The traditional forecast process focuses on the reduction of forecast uncertainty in the first moment of the predicted variable (see Fig. 2). The entire process is driven by the ultimate goal of producing the best single value estimate of the future weather, water, and climate conditions, often resulting in the formation of a single weather, water, or climate scenario. Evaluation metrics, performance measures, and importantly, official requirements are chosen accordingly and usually forecast uncertainty is neglected.

Through the various steps in the forecast process, information from observations to the end users flows in one direction, and usually a single value, related to the best estimate of the first moment of the variable in question, is transmitted. At the end of the chain, a single scenario, often called the most likely evolution of weather, water, or climate conditions, is established as an end product. The single forecasts are often interpreted in a binary way (e.g., if the forecast is 5 C, the temperature will not be below 3 C, so concrete can be laid), instead of a more informed decision making process involving both multiple weather scenarios and the user’s specific cost / loss situation. The result is limited utility to society.

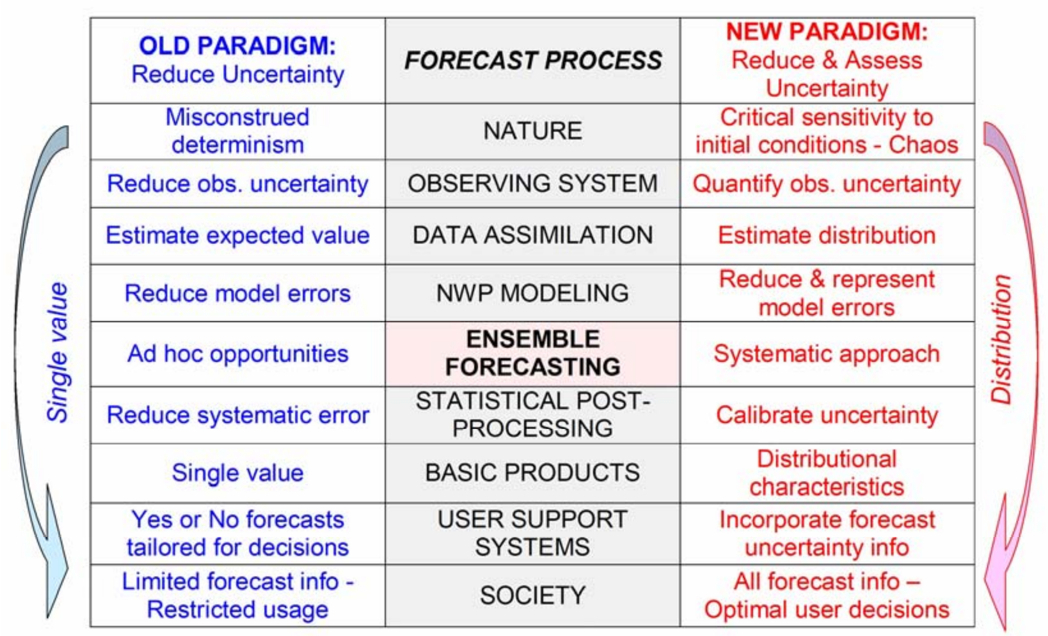

Figure 2: Schematic indicating how the various steps in the forecast process differ under the traditional vs. the new forecast paradigm.

To mitigate user concerns, forecast uncertainty appears as a second thought in the traditional forecast process. Once a decision is made regarding the single scenario that is often called the “official forecast”, the uncertainty may be assessed using statistical methods. An example is the “cone of uncertainty” attached to the official tropical storm track forecasts issued by the Tropical Prediction Center. The “cone” is based on statistics of past official track forecasts and therefore does not capture variations in forecast uncertainty from one case to another (i.e., the same cone is used in every case); and is attached to the latest official forecast (i.e., a single value forecast).

6. NEW FORECAST PARADIGM

The systematic use of ensemble information in the production of multiple forecast scenarios is necessary for the optimal derivation of both the best estimate of the state, and the associated uncertainty. The two pieces of information, put together, are conveniently and completely provided by probabilistic expression of forecast information.

Although various elements of the new forecast concept have been developed and may be in use to fulfill various responsibilities in the National Weather Service, its full introduction amounts to a major paradigm shift. Concerted changes to most aspects of forecast operations are needed for consistency and good functionality. As past efforts demonstrate, changes limited to certain aspects of the forecast process can have only a limited effect.

The new process for weather, water, and climate forecasting is based on general principles related to chaos, probability, and decision theory. Once a general framework for the new forecast process is developed and agreed upon the various forecast system components can be designed.

The rest of this document attempts to identify critical elements where the fully implemented new process must differ from the traditional forecast process, with suggestions for intermediate steps where appropriate.

7. COMPONENTS OF THE NEW FORECAST PROCESS

Ultimate goal. In general, the new forecast process can be considered as not a replacement, but rather an expansion to the traditional forecast process. The ultimate goal of the traditional process is the reduction of forecast uncertainty. This goal is supplemented (and not replaced) by the additional, new goal of also providing forecast uncertainty. All other changes in the forecast process are related to or a consequence of this broadening of its goals.

Concept of nature. The new forecast paradigm, as the old one, considers nature as a deterministic system. This gives ground to predictability. The new paradigm also recognizes that the atmosphere and the coupled atmosphere – land surface - ocean system is chaotic. Forecast evolution is sensitive to minor changes caused either by external forcing (in nature), or by errors introduced in the initial condition or the numerical model in the forecast process. Chaos limits predictability, and its recognition allows for the quantification of the loss of predictability that is at the heart of the new forecast paradigm.

Forecast process. Under the old paradigm, information related to the best estimate of the predicted system is propagated through the entire forecast process, and throughout the forecast range. In the new paradigm, not only the best estimate, but also uncertainty (related to the entire distribution of possible scenarios) is propagated. To properly assess the uncertainty in the final probabilistic forecast, ideally all uncertainty (related to the observing, data assimilation, numerical modeling, statistical post-processing components) is accounted for at its source, and its effect is propagated via an ensemble of forecasts.

Observing system. Emphasis expands from the reduction of observational uncertainty (old paradigm) to include a quantification of uncertainty. All observations must be accompanied by estimates of random instrument and representativeness error variance, as well as estimates of systematic errors. For properly propagating uncertainty information, observational errors must be provided as input in the next, data assimilation step.

Data assimilation. 3- and 4-dimensional variational and other traditional methods (old paradigm) focus on the reduction of analysis error. Aims of new methods that are under development (e.g., various ensemble-based data assimilation methods, or the use of ensemble information in variational methods) include at the same time an assessment of uncertainty in the analysis. This information is critical input for the generation of initial ensemble perturbations.

Numerical modeling. Traditionally, numerical modeling efforts have focused on reducing systematic and random errors related to model formulation. These efforts have typically led to models that exhibit a higher level of fidelity when compared to nature. The new forecast paradigm requires a major refocusing of efforts to include, beyond their reduction, a quantitative assessment and simulation of model related random and systematic errors. This is critical in ensemble forecast applications of numerical modeling, where model errors need to be represented as part of an ensemble.

Ensemble forecasting. In the new forecast process, ensemble forecasting occupies a central place in the entire process following the observing, data assimilation, and numerical modeling components, and prior to the statistical post-processing, product generation, and user support components (see Fig. 1). Ensemble forecasting involves the generation of multiple forecast trajectories, (i) initialized from a sample of plausible initial states provided by data assimilation, and (ii) integrated with a numerical modeling system that can represent model-related uncertainty. This is in contrast with traditional forecasting that is often based on a single numerical integration. Ensembles play a critical role and are used in a systematic way to connect all components of the forecast process together. They are used to assess, represent, and propagate forecast uncertainty information from the observing system through the end users.

TYPES OF ENSEMBLES

In a broad sense, any collection of forecasts can be considered an ensemble. Ensemble forecasting had many variations and flavors over the past decades. Subjective forecasts prepared by different individuals, single value forecasts prepared by different Numerical Prediction Centers (NPCs), or a “synthetic” ensemble prepared by any of the producing NPCs are examples of ensembles. In addition, ensembles produced by different centers can also be combined to yield a multi-center ensemble. For maximum quality, care must be taken to ensure that ensembles represent both initial condition and model related forecast uncertainty. The quality of ensembles can be compared using objective measures of statistical reliability and resolution. The better an ensemble can estimate the expected forecast distribution, and the better it can capture case dependent variations in forecast uncertainty, the higher the overall scores and its expected utility are. Note that the practical value of ensembles depends on both reliability and resolution.

Statistical post-processing. In the traditional forecast process, emphasis is placed on estimating and correcting systematic errors in the first moment of the forecast distribution (i.e., expected value, single forecast value). The new forecast paradigm requires a broader approach, where characteristics of the entire distribution are considered and if possible statistically corrected. A number of such techniques have been recently proposed or are being developed. Since systematic errors in numerical forecasts are a function of lead time, systematic errors are estimated and corrected separately for each lead time. Two independent facets of statistical post-processing are (a) the sample size of truth – forecast data pairs used in the statistical process, and (b) the choice of the particular statistical technique used. Relevant characteristics of bias correction algorithms are (i) bias free statistical estimation of systematic errors, and (ii) convergence rate of bias estimate as a function of increasing sample size; faster convergence rates lead to better estimates from relatively small samples.

Downscaling. Information of interest to users (e.g., some esoteric variables at a particular point in space and time) is different from the information generated by numerical models (i.e., a gridbox average of standard NWP variables). Therefore numerical model output, even if bias free, typically requires some sort of “downscaling” to maximize utility. Downscaling is defined here as the process of deriving information from a numerical forecast system (analysis or forecast) on a finer spatio-temporal resolution regarding the same or possibly new variables. Downscaling can be done in a variety of ways, including use of high resolution (but possibly simplified) numerical models or statistical methods. While methods used in the traditional forecast process often suppress realistic variability on fine temporal and spatial scales for the sake of reduced errors in a single forecast, the inclusion of such variability in downscaled ensemble forecasts (while ensuring the grid-scale characteristics are unchanged) is desirable and can improve the realism, skill and utility of forecasts. If downscaling is done using statistical methods, bias correction of lead time dependent forecast errors and downscaling the numerical forecast information to user variables can be done either in one single step, or as two separate steps.

Product generation. An ensemble of bias-corrected and downscaled forecasts constitutes the basis for generating guidance products for forecasters, and official forecast products for end users (derived products). Statistical tools can be used to derive the probability of any event based on an ensemble considered as a sample of all possible future scenarios. Events can be defined based on a single or multiple variables of interest. Alternatively, information from ensemble forecast systems can be directly used as input in decision support systems. Temporal, spatial, and cross-variable correlations are naturally captured in all basic and derived products represented in the ensemble. In contrast, the generation of probabilistic products based on a single forecast involves more statistical considerations and may be characterized by less realistic and less skilful temporal, spatial, and cross-variable correlations. Despite a higher level of complication, such methods generally do not offer the flexibility in product generation that the new, ensemble-based forecast process does.

Role of human forecaster. In the traditional forecast process, the forecaster's role is limited to the identification of the most likely forecast scenario. In the new process, this role is expanded to include the identification of alternate scenarios. This requires a broader perspective and also some new tools that allow the forecaster to manipulate entire ensemble distributions just as forecasters are able to change single value forecasts today. Due to the shear volume of data, forecasters will focus their attention on high impact events in terms of manual modifications to an automated, bias corrected and downscaled numerical forecast guidance that will be available based on the best ensemble forecast system. The primary role of the forecaster will evolve from the routine preparation of forecasts to directing and quality controlling the forecast process, saving valuable time for interactions with users and other constituents.

Decision support systems. Decision support systems (DSS) provide a link between forecast information and user applications. The sophistication of DSS can vary greatly. Single value forecasts severely limit the utility of weather, water, or climate forecast information as they allow decisions made at a single level only (yes or no, based on whether an event is forecast or not). In contrast, the multiple scenarios in an ensemble support decisions to be made at multiple levels of certainty or probability, depending on the users' cost/loss considerations. More formal DSS systems can use either derived ensemble products (e.g., probability of an event), or the basic ensemble solutions as input. Through case study, users can find the optimum decision criterion (i.e., threshold probability of a forecast event) at which they must take action that best exploits weather forecast information for their application.

User feedback. Information in the traditional forecast process flows in one direction, from observations to the end users. In the new process, using the dynamical spatio-temporal structure inherent in a set of ensemble forecasts, information can also flow in the opposite direction from the users to the observing system. Through the use of ensembles, the new forecast process facilitates adaptations of the entire forecast process (I.e., taking adaptive observations, using special adaptive data assimilation methods, on demand numerical modeling, or ensemble applications) to case dependent user needs via propagating user requirement information backward throughout the forecast steps (for an example, see Fig. 3).

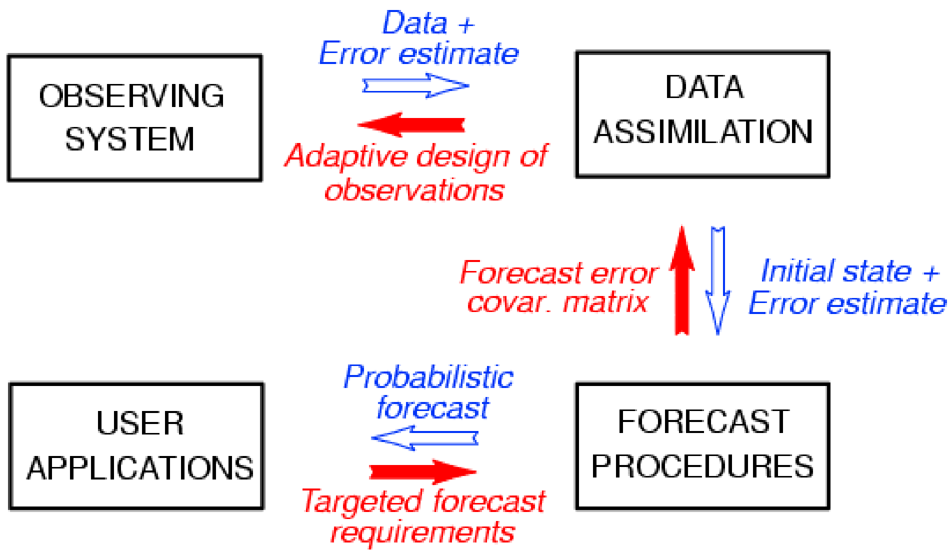

Figure 3: Schematic illustrating the two-way flow of information from observations to users and users to the observing system, as envisaged by the THORPEX Research Program sponsored by the World Weather Research Program of the World Meteorological Organization,. Adopted from Toth and Majumdar (2007).

Evaluation. In the traditional forecast process, forecast evaluation focuses on the verification of the best single value estimate. Under the new paradigm, probabilistic forecasts are evaluated using a variety of metrics to assess the reliability and resolution of forecasts in each stage of the process. Such statistics can be used to assess the value added at each step of the forecast process, and serve as a guide in terms of future improvements to the entire forecast system, as well as a reference for user expectations in various applications. Note that as each user's circumstances are unique, verification must also encompass user specific statistics, best computed at the user's end. Such statistics can inform decisions on the future development of the forecast system, as well as quantitatively influence adaptive decisions regarding the allocation of adaptively configurable observing and forecast resources.

Impact on society. The traditional forecast system has a limited set of outputs based on a single forecast and users have to trust or distrust it based on personal experience. In contrast, in the new forecast process, users can not only acquire the exact information they want to use on a case-by-case basis, but via the aggregate of all prioritized user requirements, can also potentially influence the entire forecast process (from the observing system, data assimilation, and numerical modeling / ensemble systems) to maximize user value for high impact event forecasts.

8. RECOMMENDATIONS

This section is devoted to the specific changes and new elements that need to be introduced in a well coordinated fashion into the existing system for the implementation of the new forecast paradigm, along with a list of existing related research and development activities, and with some specific recommendations.

GENERAL RECOMMENDATIONS

1) Define the end goal (5-10 years): Operational use of general probabilistic forecast approach across all forecast activities.

   Adopt the concept of a new, ensemble-based forecast process. This will allow the realization of maximum skill and utility on one hand, and the development and use of a common scientific, software, and technological infrastructure across all weather, water, and climate applications within an entire forecast organization on the other hand. Details of the new approach related to various applications etc. are to be developed later as part of scientific, technological, and workforce development activities.

2) Change official forecast requirements and performance measures from single value to probabilistic format.

   Keep the existing requirements for a transitional period; establish new probabilistic format requirements and associated performance measures that are consistent with long term goal. Establish a timeline for transition process for various applications. Allow various organizations some flexibility regarding the date of their compliance with new requirements. This will facilitate an orderly transition from the existing to the new forecast process.

3) Design, develop, and implement the new forecast process.

   Planning and execution must be consistent with the mission of NOAA and the long-term goals of the NWS. Leverage off related existing activities such as THORPEX research and development aimed at improved predictability, probabilistic forecasting, and adaptive forecast methods. For maximum efficiency, ensure adequate funding of ongoing and planned work through the existing framework of these activities.

4) Identify the first practical steps on the path to achieving the end goal for the next 2-3 years.

   Define required basic capability that is consistent with, and can evolve toward end goal, provides critical functionalities albeit with limitations, and achievable in 2-3 years. Identify critical components in the forecast process, with specific functionalities and performance measures, and well defined interfaces between adjacent components, to be developed in following years. A well laid out modular design will allow interested groups within and outside of a forecast organization to work independently but in a coordinated fashion on various components of the new system.

SPECIFIC RECOMMENDATIONS

1) Develop **new forecast requirements** consistent with general probabilistic approach
   a) Rewrite forecast requirements for each application (short-term)
   b) Define corresponding new objective performance measures (short-term)
2) Continue **development of new forecast process** with an emphasis on high impact events. Maximize the utilization of forecast resources by allowing case dependent adjustments through adaptive methods in
   a) Collection and use of observations (targeted observations)
   b) Data assimilation (e.g., flow dependent background error estimation)
   c) Numerical modeling (adaptive resolution models)
   d) Ensemble forecasting (case dependent variations in membership and composition of ensemble)
   e) Decision support systems (allowing flexible user actions depending on level of forecast probabilities etc)
3) Develop **bias correction and downscaling methods** that statistically adjust ensemble forecast data generated automatically on a course model grid and brings it to a finer spatial resolution user grid, e.g.,the National Digital Guidance Database (NDGD) of the US National Weather Service (NWS) This will provide both bias-free and downscaled ensemble trajectories and derived probabilistic products.
   a) Estimate lead-time dependent forecast model error (both first and higher moments of ensemble distribution) and correct all ensemble forecasts on model grid.
   b) Generate fine resolution ensemble data consistent with coarse resolution bias-corrected ensemble, including the addition of fine scale spatial and temporal variance missing from coarse resolution data
4) Contribute to the establishment of an organization-wide ***environmental data depository*** for the collection, organization, and exchange of observational, numerical guidance, and official forecast information accessible from both within and outside of a forecast organization.
   a) Generalize high resolution forecast databases (e.g., the National Digital Forecast Database (NDFD) and the NDGD of the NWS so they can hold forecast uncertainty information (summary statistics based on ensemble forecasts - short-term solution)
   b) Develop database capable of holding all ensemble trajectories (all members, relevant variables, lead times, full capability, long term solution)

5) Define **summary ensemble statistics** (e.g., 10, 50, and 90 percentile values of cumulative forecast distribution) that can be used to
    a) Collapse vast amount of ensemble data into smaller dataset suitable for inclusion into expanded NDFD/NDGD datasets (short-term)
    b) Manually inspect and modify ensemble forecast information (short-term)
6) Develop ***statistical interrogation, forecast modification, and product generation tools*** associated with guidance and forecast database that can be used to
    a) Derive summary statistics from full ensemble data (to populate new NDGD & NDFD uncertainty grids – short term)
    b) Manually modify automatically generated summary statistics, including uncertainty grids (short-term)
    c) Derive selected additional information from summary statistics stored in NDGD & NDFD grids (Product generation - short term)
    d) Automatically modify full ensemble trajectory data based on manually modified summary ensemble statistics (long term)
    e) Derive any forecast information from full ensemble trajectory data stored in future database (Product generation - long term)
7) Develop adequate **telecommunication facilities** that allow all offices participating in the forecast process and all users real time access to
    a) Summary forecast statistics and limited derived products (short term)
    b) All ensemble forecasts and any derived product (long term)
8) Develop general **probabilistic verification package** to be used across a forecast organization, and shared with user community, to
    a) Compute official performance measures
    b) Evaluate the value added at each step along the forecast chain
    c) Assess the value of newly developed techniques as compared to methods used in operations
9) Develop and implement **comprehensive training** to prepare forecasters for their new role in the forecast process that includes forecasting uncertainty, including:
    a) Statistical background
    b) Ensemble methods
    c) Best forecast practices related to capturing forecast uncertainty
    d) Applications of probabilistic and other uncertainty information
10) In partnership with the weather, climate, and water enterprise, develop and implement an **outreach program** related to the communication and use of uncertainty information by the user community, to:
    a) Determine best ways of communicating forecast uncertainty to various segments of the user community (see, e.g., Morss et al 2008)

b) Compile a sample of Decision Support Systems based on ensemble/probabilistic forecast information

c) Establish close partnership with a subset of users (possibly in the public sector in the emergency and/or water management communities) to demonstrate the use of ensemble/probabilistic forecast information in real life decision making situations

ACKNOWLEDGEMENTS:

This paper is based on presentations and discussions at and following the 3rd NCEP Ensemble User Workshop held in Laurel, MD, 31 Oct – 2 Nov 2006 (http://wwwt.emc.ncep.noaa.gov/gmb/ens/UserWkshop_Oct2006.html).

Contributions by all participants of the workshop are gratefully acknowledged. Comments by Tom Hamill, Roman Krzysztofowicz, Rebecca Morss, and John Schaake on earlier versions of this manuscript are also acknowledged.

REFERENCES:

Buizza, R., 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.

Morss, R., E., J. L. Demuth, and J. K. Lazo, 2008: Communicating Uncertainty in Weather Forecasts: A Survey of the U.S. Public. *Weather and Forecasting*, in press.

National Research Council of the National Academies (NRC), 2006: Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts. [Available at: http://www.nap.edu/]

Toth, Z. and S. J. Majumdar, 2007: Meeting Summary: NOAA THORPEX Science Workshop. *Bull. Amer. Meteor. Soc*. 88, 717-719.

Toth, Z., O. Talagrand, and Y. Zhu, 2006. The attributes of forecast systems: A general framework for the evaluation and calibration of weather forecasts. In *Palmer, T.N. and Hagedorn, R., editors, Predictability of Weather and Climate*. Cambridge University Press, 584-595.

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