PROBCAST

A Web-Based Portal to Mesoscale Probabilistic Forecasts

BY CLIFFORD MASS, SUSAN JOSLYN, JOHN PYLE, PATRICK TEWSON, TILMANN GNEITING, ADRIAN RAFTERY, JEFF BAARS, J. M. SLOUGHTER, DAVID JONES, AND CHRIS FRALEY

An experimental Web site tests innovative approaches for displaying probabilistic weather predictions based on postprocessed high-resolution ensemble forecasts.

The transition from deterministic weather prediction to an environment in which probabilistic forecasts are generated, communicated, and applied is one of the great challenges before the meteorological community. Recently, both an official statement of the American Meteorological Society (AMS 2008) and the "Completing the Forecast" report of the National Academy of Sciences (NAS; National Research Council 2006) noted that widespread dissemination of probabilistic weather predictions would yield substantial economic and societal benefits and that considerable work is required to realize this potential. In particular, the National Academy of Science report recommended efforts that would improve the generation and postprocessing of mesoscale, short-range ensembles. The NAS report also noted current deficiencies in the communication of uncertainty information and that social and behavioral scientists can help determine the best approaches.

The construction of ensemble-based mesoscale probabilistic prediction systems that produce reliable and sharp probability density functions (PDFs) is an area of active research, but one in which substantial progress has been made during the past 10 yr (e.g., Grimit and Mass 2002; Eckel and Mass 2005; Raftery et al. 2005; Tewson and Raftery 2006). The challenges begin with building a sufficiently large, diverse, and high-resolution mesoscale ensemble that sufficiently samples the underlying mesoscale uncertainty. Next, postprocessing is required to remove bias, properly weight the various ensemble members, and adjust the resulting PDF to be as reliable and sharp as possible.[1] However, even if excellent probabilistic guidance is produced through modeling and postprocessing, an even greater challenge looms: disseminating such information in a way that allows users to make better decisions. This task is daunting. Probabilistic prediction systems produce huge amounts of information, and the general public and even educated users are unfamiliar with the interpretation, benefits, and

AFFILIATIONS: MASS AND BAARS—Department of Atmospheric Sciences, University of Washington, Seattle, Washington; JOSLYN—Department of Psychology, University of Washington, Seattle, Washington; PYLE, TEWSON, JONES—Applied Physics Laboratory, University of Washington, Seattle, Washington; GNEITING, RAFTERY, SLOUGHTER, AND FRALEY— Department of Statistics, University of Washington, Seattle, Washington
CORRESPONDING AUTHOR: Professor Clifford F. Mass, Department of Atmospheric Sciences, Box 351640, University of Washington, Seattle, WA 98195
E-mail: cliff@atmos.washington.edu

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/2009BAMS2775.1

In final form 8 January 2009

[1] A reliable (or calibrated) PDF is one in which if a quantity is predicted to occur with a probability of X%, on average it verifies X% of the time. A sharp PDF is one in which prediction intervals (say, between the 10% and 90% values of the PDF) are as narrow as possible and certainly less than those obtained from climatology. Climatological forecasts are reliable but not sharp.

application of uncertainty guidance. Research and experimentation are acutely needed to explore and perfect the best approaches for communicating probabilistic weather forecasts, including the development of new terminology, displays, and interfaces.

For nearly 10 yr, an interdisciplinary group at the University of Washington (UW) has worked to develop an end-to-end probabilistic mesoscale prediction system. This effort, sponsored by the Department of Defense, the National Weather Service, and the National Science Foundation, has brought together the Departments of Atmospheric Sciences, Statistics, and Psychology at UW, as well as colleagues from the UW Applied Physics Laboratory, to develop a complex mesoscale probabilistic prediction system that encompasses mesoscale ensembles, grid-based bias correction, Bayesian postprocessing, and a Web page dissemination portal that takes advantage of research on how people interpret and use probabilistic information. This portal, the University of Washington Probability Forecast (PROBCAST), and the infrastructure and research underlying it, are described in this paper. PROBCAST, a real-time system available for approximately two years, is generally used by hundreds of unique visitors per day and has been the subject of a front-page story in the *Seattle Times* (available online at www.stat.washington.edu/MURI/seattleTimes/times.html). As discussed later, PROBCAST is intended as a prototype and test bed for exploring the best approaches for communicating high-resolution uncertainty information to a large and varied user community.

PROBCAST INFRASTRUCTURE. The centerpiece of the UW PROBCAST project is the Web site (online at www.PROBCAST.com or PROBCAST.washington.edu; see Fig. 1). This Web site is the front end of a sophisticated modeling and postprocessing data system (Fig. 2) that begins with the University of Washington mesoscale ensemble system (UWME). In operation since 2000, UWME is a real-time prediction system based on an ensemble of 17 fifth-generation Pennsylvania State University (PSU)–National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5) forecasts run over domains with 36- and 12-km grid spacing, the latter encompassing the Pacific Northwest (Washington, Oregon, Idaho, and southern British Columbia, Canada). The UWME includes a subset of eight twice-daily 72-h forecasts, with the various runs driven by the initializations and boundary conditions from major numerical weather prediction centers around the world [see Grimit and Mass (2002) for a complete description of UWME]. The output of UWME is then postprocessed using the Bayesian model averaging (BMA) methodology developed by the UW Department of Statistics (Raftery et al. 2005; Sloughter et al. 2007). BMA improves the PDF produced by the raw ensemble in a number of ways. First, it provides bias correction, which is important because model surface parameters often have large systematic biases. Second, it applies an assumed PDF (Gaussian for temperature and gamma distribution for precipitation) to each member and then weights each member by its relative performance over the past

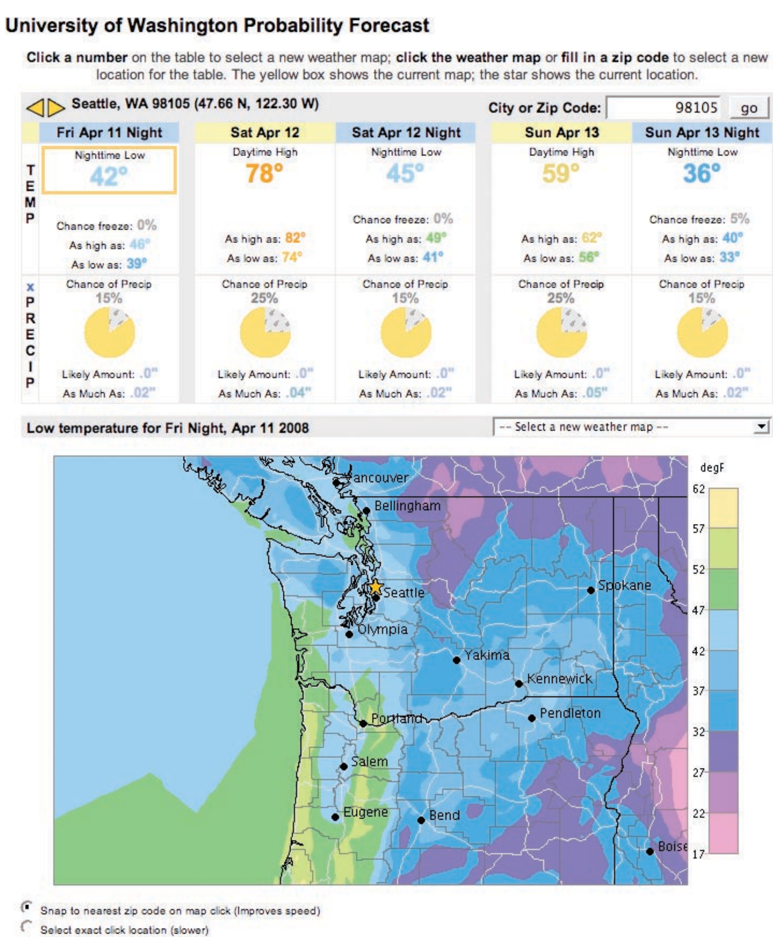

Fig. 1. Example of the PROBCAST Web page, shown for 11 Apr 2008.

few weeks. The controlling parameters for BMA are determined during this multiweek training period to produce PDFs that are reliable and sharp. For temperature, the UW system uses local BMA, in which the weights and control parameters vary spatially across the domain, while for precipitation global BMA is applied, with the same weights and control parameters everywhere. The result of the UWME–BMA system are PDFs at all surface grid points over the domain, in contrast to conventional approaches such as traditional model output statistics (MOS) that produce probabilistic information only at observation locations.[2]

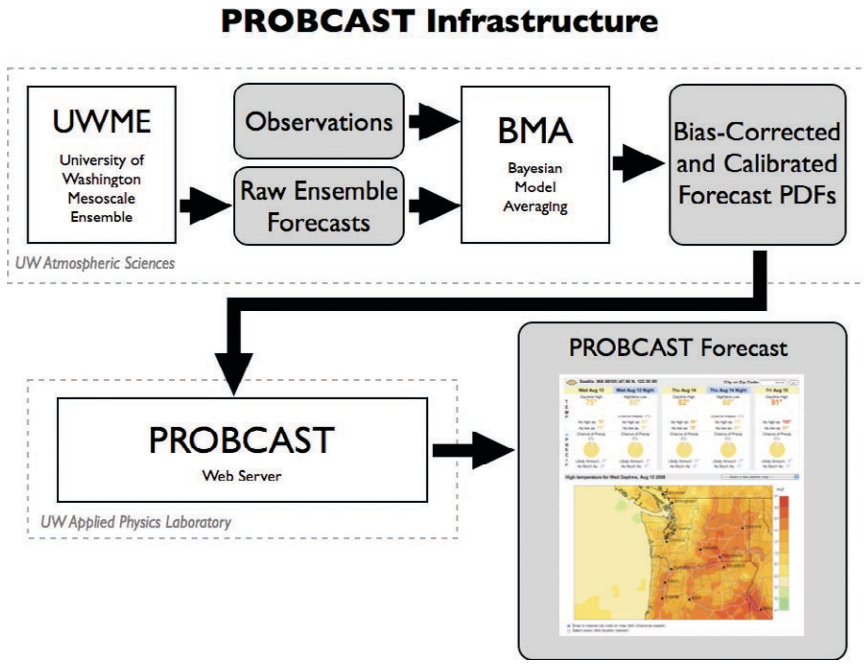

Fig. 2. PROBCAST information flow diagram.

The PROBCAST Web server runs a Web application that provides an easy-to-interpret interface for the UWME–BMA probabilistic forecasts. The full BMA output is available online at http://bma.apl.washington.edu/. At this site the complete PDF information at each location is accessible as well as the weightings of the various forecasts used in the ensemble.

Based on the location provided by the user, the application determines the relevant forecast PDFs and composes a Web page depicting probabilistic temperature and precipitation forecasts for that location. In addition, the software produces a variety of forecast maps for the entire region using the forecast PDFs at the grid points. These features are described in a later section.

Important issues in the design of the PROBCAST Web site include determining the best way to communicate the probability of precipitation and potential extreme values, both based on the forecast PDF. The next section will describe some of the research underlying PROBCAST's design.

EVALUATING USER INTERPRETATION OF PROBABILISTIC INFORMATION. According to Web-based surveys completed during the early stages of this project and the recent results of Morss et al. (2008), the general public applies implicit weather probability thresholds for normal tasks, such as clothing selection and travel plans. These results suggest that users need direct access to uncertainty information so they do not have to rely on deterministic forecasts and their own subjective estimates of uncertainty. Explicit uncertainty information can also increase users' trust in weather forecasts. When the observed value falls within the predicted range of a probabilistic forecast, the prediction is perceived as accurate, while a single-valued deterministic forecast is considered wrong when it does not verify.

A survey of UW college students (S. Joslyn and S. Savelli 2008, personal communication) indicates that expected forecast uncertainty is quite large, about twice the typical 80% predictive interval from calibrated probabilistic prediction systems. In fact, a number of students estimated a 12° range for a 2-day temperature forecast, and most expected less accuracy for more extreme forecasts. Thus, the availability of calibrated uncertainty forecasts allow users to make decisions based on their own tolerance for risk but without subjective, and often problematic, estimation of the underlying uncertainty.

A major thrust of the psychological research underlying PROBCAST was to design icons that communicate the probability of precipitation accurately. Several studies suggest that people routinely misunderstand probability of precipitation (PoP; Murphy et al. 1980; Gigerenzer et al. 2005; Morss et al. 2008).

[2] An experimental gridded MOS is now being tested by the National Weather Service.

For example, some believe that the percentage refers to the proportion of area or time over which rain will be observed. UW psychologists, investigating this issue in three separate studies among over 500 participants (Joslyn et al. 2009), found that approximately one-third of the participants gave answers suggesting the time or area misconceptions. Other studies (e.g., National Research Council 2006) have found that currently used precipitation icons are generally inconsistent and confusing, and have been devised without any human subjects research.

After extensive testing of several visual presentation formats, we found that interpretation errors were reduced when the chance of *no* rain was made explicit, either in writing (there is a 25% chance of rain *and a 75% chance of no rain*) or graphically, using a pie icon that illustrates a clear section as well as a section with rain imagery (Joslyn et al. 2009). Although it is clear that a 25% chance of rain implies a 75% chance of no rain to a person with the correct understanding, specifying the chance of no rain either in numbers or graphically appears to be helpful to many subjects. After testing a variety of icons in user surveys (e.g., Fig. 3), the pie icon was applied to display the PoP forecast in PROBCAST.

Another major issue for any probabilistic prediction system is communicating the probability of extremes. Such information is, of course, inherent in the PDFs produced by the BMA; however, most members of a lay or general audience are not prepared to interpret or apply such raw statistical information. PROBCAST provides the 90th percentile as the upper bound ("as high as") and the 10th percentile as the lower bound ("as low as") from the BMA probability distribution for temperature (an 80% predictive interval). The interpretation of these two bounds would seem to be straightforward; for example, the lower bound of the 80% predictive interval is the temperature for which there is a 10% chance that the observed temperature will be lower. Notice, however, it could also be defined as the temperature for which there is a 90% chance that the observed temperature would be higher. Both definitions describe the same situation, although the first tends to emphasize the unlikely outcome and can lead people to think it is more likely than it really is. This is a psychological effect known as "framing" (Kuhberger 1998; Levin et al. 1998), which is a trick that is well known to advertisers ("save 20%," not "pay only 80%"). The lower bound could also be defined in terms of frequency (1-time-in-10 lower; 9-times-in-10 higher) rather than a probability. Some have argued that frequency expressions of uncertainty are easier for people to understand (Gigerenzer and Hoffrage 1995; Cosmides and Tooby 1996).

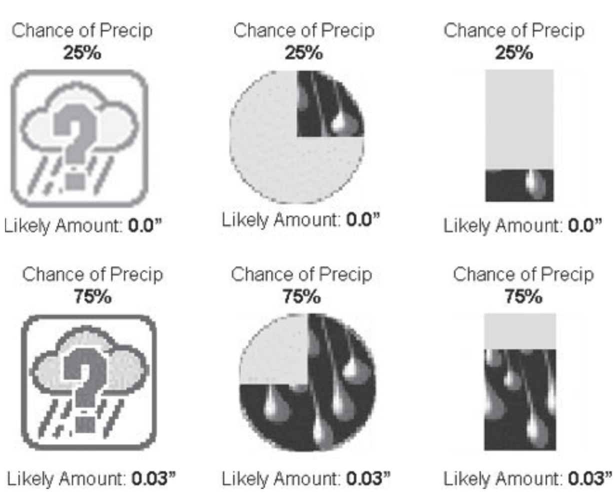

Fig. 3. Three examples of precipitation icons surveyed before choosing the pie-chart approach for PROBCAST.

All of the above approaches were tested in the context of a freeze-warning task (Joslyn et al. 2009). Three-hundred and thirteen participants were told that a warning should be issued at 32°F, which was at the lower bound of the 80% predictive interval. The forecast incorporated one of the four definitions for the lower bound described above (less than and more than, both in percentage and frequency). Surprisingly, neither frequency nor framing made a difference to people's understanding or to the quality of decisions made. What mattered was the compatibility between the expression of uncertainty and the task the participants were assigned. Reasoning errors occurred when there was a mismatch between the expression (e.g., 9 times in 10, or a 90% chance that the temperature will be *greater than* 32°F) and the relevant threshold (post a warning when temperatures will be *less than* 32°F). Errors were reduced when both the uncertainty expression (1 time in 10, or a 10% chance that the temperature will be *less*) and the threshold (temperature *less* than 32°F) matched, both describing "temperatures less." The effect was also observed in a study with 290 participants making a high-wind warning decision. They made fewer errors when *both* the forecast problem and the guidance were related to "winds greater." Based on these studies, it was concluded that the most important factor in facilitating understanding of uncertainty information is the match between the expression of uncertainty and the overall task goals. Using this guideline and assuming that people's interest in the upper bound is driven by concerns for high temperature extremes, PROBCAST provides the "greater than" definition for the upper bound. Similarly, assuming that interest in the lower bound arises from concerns about the low temperature extreme, PROBCAST provides the "less than" definition for the lower bound.

THE PROBCAST WEB PORTAL. The PROBCAST interface (Fig. 1) was designed to look familiar and unintimidating, yet it was able to provide easy access to probabilistic information furnished by the BMA-postprocessed UWME ensemble. The upper portion displays forecasts of maximum and minimum temperatures and precipitation in a way that is superficially similar to presentations seen in the media, while the lower portion presents forecast maps of key meteorological parameters. An important feature of the PROBCAST interface is the ability to view probabilistic predictions at any location in the domain, either by clicking at a point on the lower map or by entering a zip code or city name in the upper-right corner. The upper section is divided into five 12-h time periods, beginning at either 0000 UTC [1600 Pacific Standard Time (PST)] or 1200 UTC (0400 PST). For temperature, each time division reflects conditions over an 18-h period, since maximum and minimum temperatures can occur after 1700 or 0500 PST, respectively. In contrast, precipitation is shown for the relevant 12-h period only.

Although the numbers in the upper portion of the page appear similar to typical deterministic forecasts displayed on television and in newspapers, they reflect the output of the probabilistic prediction system. Minimum and maximum temperatures are taken from the means of the corresponding BMA PDFs. The probability of temperature dropping to or below freezing at the selected location, again based on the temperature PDF, is shown. Finally, the probabilistic forecasts are used to indicate possible extreme values during the 18-h period. PROBCAST indicates an "as high as" temperature, that will only be exceeded 10% of the time, and an "as low as" value, for which the temperature will drop below only 1 time in 10, on average. Thus, a user secures some of the benefits of the probabilistic information without having to deal with PDFs or other statistical terminology. As noted below, the PROBCAST group is currently investigating whether other approaches might be superior to the "as high as" and "as low as" terminology. Immediately above the precipitation icon is the numeric probability of measurable 12-h precipitation based on the BMA-produced precipitation PDF. The icon provides the same information in graphical form and, as noted earlier, has been tested to reduce confusion for a nonexpert audience. Below the icon is found the median or 50th percentile of the BMA precipitation PDF, which is termed "the most likely" amount, and beneath that an "as much as" value is displayed, which provides an amount that will only be exceeded on average once in every 10 forecasts.

The lower section of PROBCAST provides maps of probabilistic predictions based on the UWME–BMA system. Map displays are available for all of the parameters noted in the upper section, such as high, low, extreme, and freezing temperatures; the probability of freezing; or the probability of precipitation and its extreme and most likely amounts. It is possible to select these maps by either using the drop-down selector or by clicking on the corresponding parameter in the upper panel, which also allows one to select the period of interest.

A variety of help features are built into the PROBCAST interface. For example, scrolling over any of the numbers provides a detailed explanation of what it signifies. In addition, a help page can be accessed in the upper-right corner.

FUTURE ENHANCEMENTS. PROBCAST is both a real-time probabilistic forecast dissemination system and a test bed for the generation and display of probabilistic weather information. A variety of extensions are planned in the future. As noted above, the current system is based on local BMA for temperature, where the weights and control parameters for the various ensembles vary across the domain, while precipitation forecasts are made using the global BMA, where the same weights and parameters are used everywhere. Local BMA is clearly superior for temperature, and local BMA for precipitation is currently being perfected and tested. During the next year, local BMA for precipitation will be used in the real-time PROBCAST system.

PROBCAST will also be extended to other parameters, such as wind speed and direction. Another extension will be more user-friendly measures of extremes, because our current research suggests there is considerable confusion with the "as high as" and "as low as" terminology. Integration of verification statistics into PROBCAST products is another goal. Such verification measures will help users determine the confidence they should place in the forecasts for each parameter. The obvious challenge, beyond the creation of sharp and reliable probabilistic information, will be to make additional information available in a way that does not clutter the interface or make the system too complex for lay users. All of the interface changes will be informed and guided by comprehensive testing by the PROBCAST team from the UW Psychology Department and the UW Applied Physics Laboratory.

SUMMARY AND CONCLUSIONS. It is clear that the meteorological profession needs to move toward high-resolution probabilistic weather prediction if it is to supply society with the meteorological information it requires. During the past few decades, progress toward applying probabilistic prediction has been slow due to the lack of sharp, reliable mesoscale probabilistic predictions and insufficient understanding of how to communicate such information to users in a way they would find useful. The PROBCAST project is directed toward both of these problems by starting with a high-resolution mesoscale ensemble system, postprocessing the output using Bayesian model averaging to produce sharp and reliable probability density functions, and applying new approaches to presenting this information on the Web, guided by sound psychological research.

PROBCAST should be seen as an early step down a long path—one that explores how typical users could acquire useful probabilistic weather prediction information in an accessible format. Many of the technologies and approaches developed for PROBCAST are now being transferred to the Joint Ensemble Forecast System (JEFS) program, a synoptic/mesoscale probabilistic prediction system being developed jointly by the U.S. Navy and Air Force. It is hoped that the PROBCAST effort will stimulate the National Weather Service, and other major operational prediction groups, to invest in high-resolution ensemble prediction, comprehensive postprocessing, and interfaces that will allow the public to profitably use uncertainty information. Investments are also necessary in psychological and human–interface research to learn the best approaches for disseminating and applying probabilistic weather predictions.

ACKNOWLEDGMENTS. This research has been supported by the Department of Defense MURI program, the National Science Foundation (Grant ATM 0724721), the Department of Defense Joint Ensemble Forecast System (JEFS) project, and the National Weather Service CSTAR program.

REFERENCES

AMS, cited 2008: Enhancing weather information with probability forecasts. [Available online at www.ametsoc.org/policy/2008enhancingweatherinformation_amsstatement.pdf.]

Cosmides, L., and J. Tooby, 1996: Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature of judgment under uncertainty. *Cognition,* **58,** 1–73.

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

Fraley, C., A. E. Raftery, T. Gneiting, and J. M. Sloughter, 2007: *EnsembleBMA:* An *R* package for probabilistic forecasting using ensembles and Bayesian model averaging. Department of Statistics, University of Washington, Tech. Rep. 516R, 17 pp.

Gigerenzer, G., and U. Hoffrage, 1995: How to improve Bayesian reasoning without instruction: Frequency formats. *Psychol. Rev.,* **102,** 684–704.

——, 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 Anal.,* **25,** 3623–3629.

Grimit, E. P., and C. F. Mass, 2002: Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Wea. Forecasting,* **17,** 192–205.

Joslyn, S., L. Nadav-Greenberg, and R. M. Nichols, 2009: Probability of precipitation: Assessment and enhancement of end-user understanding. *Bull. Amer. Meteor. Soc.,* **90,** 185–193.

Kuhberger, A., 1998: The influence of framing on risky decisions: A meta-analysis. *Organ. Behav. Hum. Decis. Processes,* **75,** 23–55.

Levin, I. P., S. L. Schneider, and G. J. Gaeth, 1998: All frames are not created equal: A typology and critical analysis of framing effects. *Organ. Behav. Hum. Decis. Processes,* **76,** 149–188.

Morss, R. E., J. Demuth, and J. K. Lazo, 2008: Communicating uncertainty in weather forecasts: A survey of the U.S. public. *Wea. Forecasting,* **23,** 974–991.

Murphy, A. H., S. Lichtenstein, B. Fischhoff, and R. L. Winkler, 1980: Misinterpretations of precipitation probability forecasts. *Bull. Amer. Meteor. Soc.,* **61,** 695–701.

National Research Council, 2006: Completing the forecast: Characterizing and communicating uncertainty for better decisions using weather and climate forecasts. National Academies Press, 124 pp.

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

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.

Tewson, P., and A. E. Raftery, 2006: Real-time calibrated probabilistic forecasting website. *Bull. Amer. Meteor. Soc.,* **87,** 880–882.