

PROBCAST: A Web-Based Portal to Mesoscale Probabilistic Forecasts

Clifford Mass¹, Susan Joslyn, John Pyle, Patrick Tewson, Tilmann Gneiting, Adrian Raftery, Jeff Baars, J. M. Sloughter, David Jones and Chris Fraley

Submitted to the *Bulletin of the American Meteorological Society*

August 2008

Capsule Statement for BAMS:

An experimental website tests innovative approaches for displaying probabilistic weather predictions based on post-processed high-resolution ensemble forecasts.

¹ Corresponding Author
Professor Clifford F. Mass
Department of Atmospheric Sciences
Box 351640
University of Washington
Seattle, WA 98195
(206) 685-0910
cliff@atmos.washington.edu

Abstract

This paper describes PROBCAST, a web-based portal to probabilistic weather predictions over the Pacific Northwest. PROBCAST products are derived from the output of a mesoscale ensemble system run at the University of Washington, with the fields being post-processed using Bayesian Model Averaging to produce sharp and reliable probabilistic predictions of temperature and precipitation. Based on research by University of Washington psychologists and human-interface specialists, a web site has been constructed that allows access to key elements of the probabilistic information. The design approach of the PROBCAST system is explained in this paper, as well as some of the challenges for future development.

Introduction

The transition from deterministic weather prediction and application 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 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 report recommended efforts that would improve the generation and post-processing of mesoscale, short-range ensembles. The NAS report also noted current deficiencies in the communication of uncertainty information and that the expertise of social and behavioral scientists was required to determine the best approaches.

The construction of an ensemble-based mesoscale probabilistic prediction system that produces 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 ten years (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, post-processing is required to remove bias, properly weigh the various ensemble members, and to adjust the resulting PDF to be as reliable and sharp as possible². But even if excellent probabilistic guidance is produced through modeling and

² 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 narrower on average than those obtained from climatology. Climatological forecasts are reliable but not sharp.

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 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 ten years, 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 the 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 post-processing, and a web-page dissemination portal that takes advantage of research on how people interpret and use probabilistic information. This portal, known as PROBCAST, and the infrastructure and research underlying it, are described in this paper. Although not heavily advertised, PROBCAST, a real time system available for approximately two years, is generally used by roughly one hundred unique visitors per day and has been the subject of a front-page story in the *Seattle Times*³. 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.

³ <http://www.stat.washington.edu/MURI/seattleTimes/times.html>

PROBCAST Infrastructure

The centerpiece of the UW PROBCAST project is the web site, found at www.PROBCAST.com or PROBCAST.washington.edu (Figure 1). This web site is the front end of a sophisticated modeling and post-processing data system (Figure 2), which begins with the University of Washington Mesoscale Ensemble system (UWME). In existence since 2000, UWME is a real-time prediction system based on an ensemble of seventeen 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). The UWME applied in PROBCAST 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 Grit and Mass 2002 for a complete description of UWME). The output of UWME is then post-processed 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 probability density function (PDF) produced by the raw ensemble in a number of ways. First, it provides bias correction, which is important since model surface parameters often have large systematic biases. Second, it applies an assumed PDF (Gaussian for temperature, gamma for precipitation) to each member and then weighs each member by its relative performance over the past few weeks. The controlling parameters for BMA are determined during this multi-week training period to produce PDFs that are reliable and sharp. For temperature, the UW system uses local BMA, in which the weights and controlling parameters vary spatially across the domain, while for precipitation global BMA is applied, with the same weights and controlling 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 locations⁴.

The PROBCAST web server runs a custom web application that provides an easy-to-use interface for the UWME/BMA probabilistic forecasts. 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 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 the PROBCAST's design.

Evaluating User Interpretation of Probabilistic Information

According to web-based surveys completed during the early stages of this project, the general public uses implicit weather probability thresholds for normal tasks, such as clothing selection, travel plans and winterizing their homes. 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. A survey of UW college students (Joslyn and Savelli 2008, personal communication) indicates that they would not be surprised if the observed temperature was within a 12-degree range for a 2-day temperature forecast. Furthermore, the more extreme the forecast, the

⁴ An experimental gridded MOS is now being tested by the National Weather Service.

less accurate people expect it to be. However, the forecast uncertainty that people expect is quite large, about twice that of the typical 80% predictive interval from calibrated probabilistic prediction systems. The availability of calibrated uncertainty forecasts allow users to make decisions based on their own tolerance for risk; furthermore, when the observed value falls within the predicted range the forecast can be regarded as accurate, in contrast to a single predicted value, which is considered wrong when it does not verify.

A major thrust of the psychological research underlying PROBCAST was to design icons that accurately communicate the probability of precipitation. Several studies suggest that people routinely misunderstand probability of precipitation (PoP) (Murphy et al. 1980, Gigerenzer et al. 2005). For example, some believe that the percentage refers to the proportion of area or time over which rain will be observed. UW psychologists investigated this issue in three separate studies among over 500 participants (Joslyn, Nadav-Greenberg and Nicholls 2008), finding that approximately one-third of the participants gave answers suggesting the time or area misconceptions. Other studies (e.g., NAS 2006) have found that current precipitation icons are often inconsistent and confusing, and have been devised without any human subjects research.

After testing several visual presentation formats (Figure 3), 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, Nadav-Greenberg and Nicholls 2008). 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 helps those used to thinking of it as percent area. As a result of this research, PROBCAST uses the pie icon to display the PoP forecast.

Another major issue for any probabilistic prediction system is communicating the probability of extremes. Such information is, of course, inherent in the probability density functions (PDFs) produced by 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 and the 10th percentile as the lower bound from the probability distribution for temperature (an 80% predictive interval). The interpretation of these two bounds seems 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 might lead people to think it is more likely than it really is. This is a psychological effect known as a “framing” (Kuhberger 1998; Levin, Schneider, and Gaeth 1998), 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).

All of the above approaches were tested in the context of a freeze warning task (Joslyn, Nadav-Greenberg, and Nichols 2008). 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 90% chance 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 times in 10 or 10% chance the temperature will be *less*) and the threshold matched (temperature *less than* 32°), 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 (Figure 1) was designed to look familiar and unthreatening, yet 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 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 (4 PM Pacific Standard Time, PST) or 1200 UTC (4 AM PST). For temperature, each time division

reflects conditions over an 18-h period, since maximum and minimum temperatures often occur after 5 PM or 5 AM PST. 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 one time in ten, 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 measureable 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 non-expert audience. Below the icon is found the median or 50th percentile of the BMA precipitation PDF, and beneath that an extreme “as much as” value is displayed, which provides an amount that will only be exceeded on average once in every ten forecasts.

The lower section of PROBCAST provides maps of probabilistic predictions based on the UWME-BMA system. Map displays are available for all the parameters noted in the upper section, such as high, low, extreme and freezing temperatures or the probability of precipitation and the 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 useful 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 controlling parameters for the various ensembles vary across the domain, while precipitation forecasts are made using global BMA, where the same weights and controlling 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 improved and more user-friendly measures of extremes, since 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 makes 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 Lab.

Summary and Conclusions

This paper describes PROBCAST, an end-to-end probabilistic prediction system designed to serve as an experimental prototype for the mesoscale uncertainty forecasting system of the future. It is clear that the meteorological profession needs to move towards high-resolution probabilistic weather prediction if it is to supply society with the meteorological information it requires. During the past few decades, progress towards 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 will find useful. The PROBCAST project is directed towards both of these problems by starting with a high-resolution, mesoscale ensemble system, post-processing 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 average users can acquire useful probabilistic weather prediction information in an accessible and clear-cut 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 make the necessary investments in high-resolution ensemble prediction, comprehensive post-processing, 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.

Acknowledgements

This research has been supported by the Department of Defense MURI program, the National Science Foundation (Grant No. ATM 0724721), the Department of Defense Joint Ensemble Forecast System (JEFS) project, and the National Weather Service CSTAR program.

References

American Meteorological Society, 2008: Enhancing weather information with probability forecasts. Available from

www.ametsoc.org/policy/2008enhancingweatherinformation_amsstatement.pdf

Cosmides, L. and Tooby, J. (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., Raftery, A.E., Gneiting, T., and J.M. Sloughter, 2007: *ensembleBMA*: An R Package for probabilistic forecasting using ensembles and Bayesian Model Averaging, Tech. Rep. No. 516R, Department of Statistics, University of Washington (revised July 2008).

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

Gigerenzer, G., R. Hertwig, E. van den Broek, F. Fasolo, and K. V. Katsikopoulos, 2005: A 30% chance of rain tomorrow: How does the public understand probabilistic weather forecasts? *Risk Analysis*, **25**, 3 623-629.

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., Nadav-Greenberg, L. and Nicholls, R. M., 2008. Probability of Precipitation: Assessment and Enhancement of End-User Understanding. Accepted in *Weather and Forecasting*

Kuhberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes*, **75**, 23–55.

Levin, I. P., Schneider, S. L. and Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, **76**, 149-188.

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

Nadav-Greenberg, L., Joslyn, S., and Taing, M. U. (2008) The effect of weather forecast uncertainty visualization on decision-making. *Journal of Cognitive Engineering and Decision Making*, in press.

NAS, 2006: Completing the forecast: characterizing and communicating uncertainty for better decisions using weather and climate forecasts. National Research Council, Washington, D.C., 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 Raftery, A.E., 2006. Real-Time Calibrated Probabilistic Forecasting Website. *Bull. Amer. Meteor. Soc.*, **7**, 880-882.

Captions

Figure 1. Example of the PROBCAST web page, shown for April 11, 2008

Figure 2. PROBCAST information flow diagram.

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

University of Washington Probability Forecast

Click a number on the table to select a new weather map; click the weather map or fill in a zip code to select a new location for the table. The yellow box shows the current map; the star shows the current location.

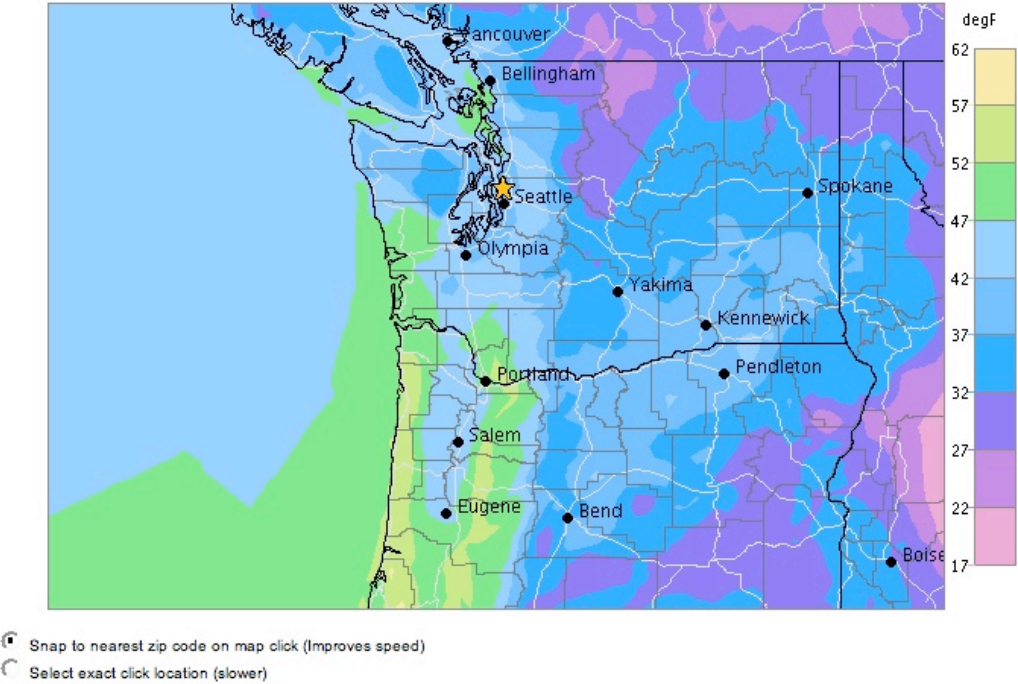
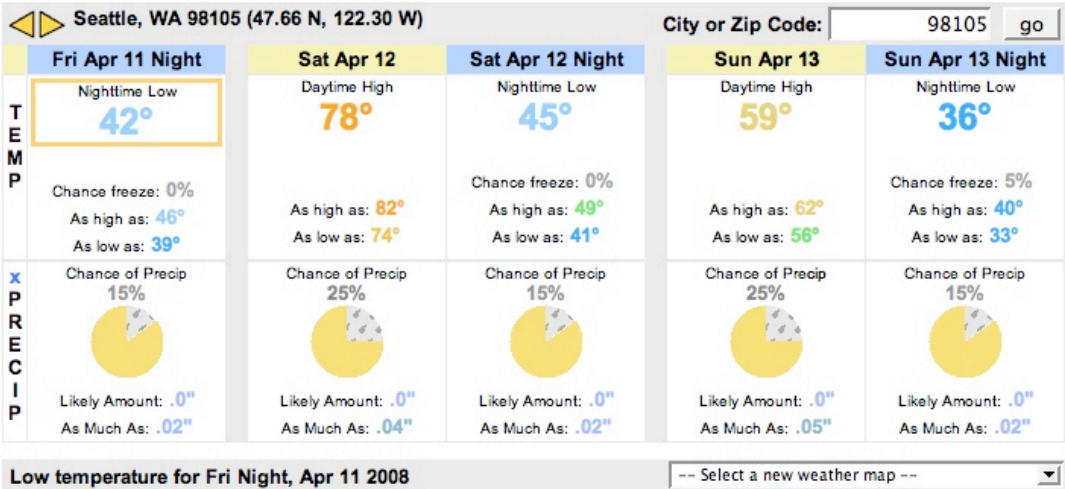


Figure 1. Example of the PROBCAST web page, shown for April 11, 2008.

PROBCAST Infrastructure

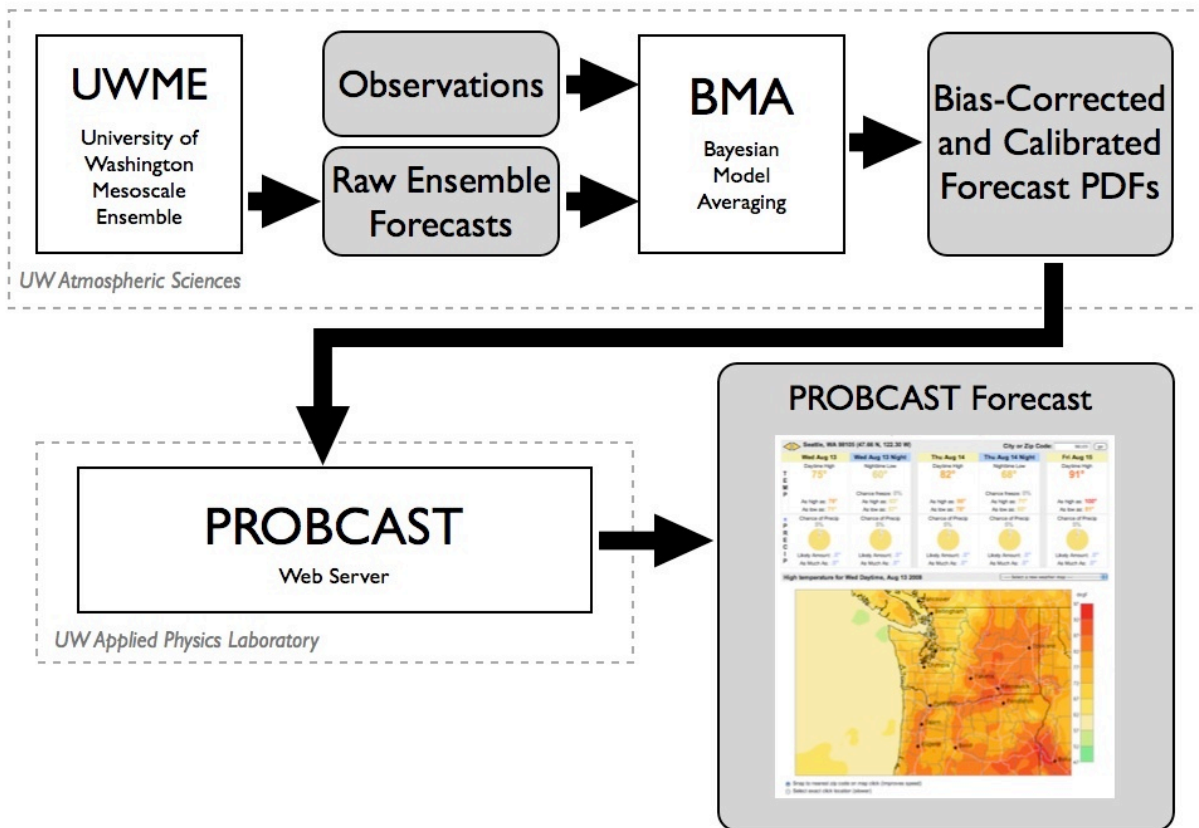


Figure 2. PROBCAST information flow diagram.

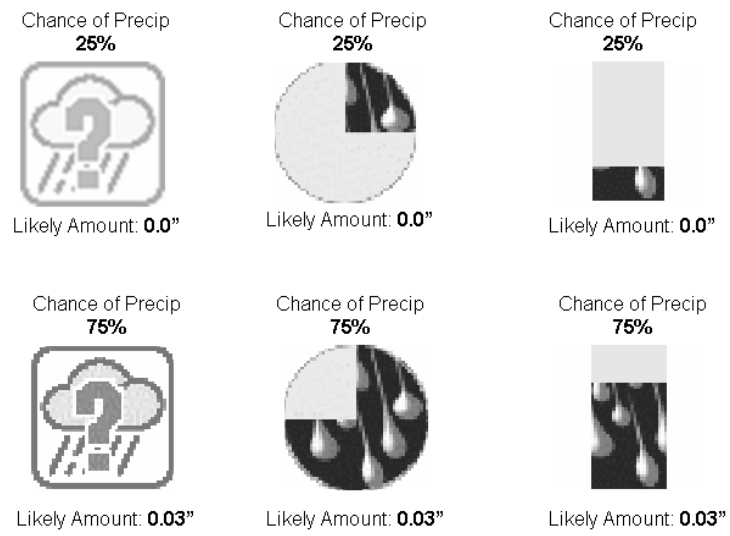


Figure 3: Three examples of precipitation icons surveyed before choosing the pie-chart approach for PROBCAST.