

The Effect of Probabilistic Information on Threshold Forecasts

SUSAN JOSLYN AND KARLA PAK

Department of Psychology, University of Washington, Seattle, Washington

DAVID JONES

Applied Physics Laboratory, University of Washington, Seattle, Washington

JOHN PYLES

Department of Cognitive Sciences, University of California, Irvine, Irvine, California

EARL HUNT

Department of Psychology, University of Washington, Seattle, Washington

(Manuscript received 3 November 2005, in final form 17 October 2006)

ABSTRACT

The study reported here asks whether the use of probabilistic information indicating forecast uncertainty improves the quality of deterministic weather decisions. Participants made realistic wind speed forecasts based on historical information in a controlled laboratory setting. They also decided whether it was appropriate to post an advisory for winds greater than 20 kt (10.29 m s^{-1}) during the same time intervals and in the same geographic locations. On half of the forecasts each participant also read a color-coded chart showing the probability of winds greater than 20 kt. Participants had a general tendency to post too many advisories in the low probability situations (0%–10%) and too few advisories in very high probability situations (90%–100%). However, the probability product attenuated these biases. When participants used the probability product, they posted fewer advisories when the probability of high winds was low and they posted more advisories when the probability of high winds was high. The difference was due to the probability product alone because the within-subjects design and counterbalancing of forecast dates ruled out alternative explanations. The data suggest that the probability product improved threshold forecast decisions.

1. Introduction

Modern-day weather forecasters rely heavily on numerical weather and climate models that make weather predictions by transforming present into future weather conditions according to the known principles of atmospheric physics. Because of uncertainties in the initial state of the atmosphere and the computational representation of the equations of motion, model predictions are accompanied by varying amounts of uncertainty. It is now possible, with ensemble forecasts in which multiple simulations of the atmosphere are made, to esti-

mate and quantify the amount of uncertainty in the model prediction (Anderson 1996; Gritit and Mass 2002). In theory, this is very useful information to both weather forecasters and to the general public. However, with the exception of the probability of precipitation, forecast uncertainty is not usually communicated to the public. At present, most forecasts remain deterministic (National Research Council 2006). In part, this is because of a question about whether or not people can successfully make use of uncertainty information to improve deterministic forecast decisions.

There is very little research that addresses this issue directly. There is some empirical evidence that people can make use of uncertainty information in simulated tasks to increase rewards and reduce exposure to risk (Roulston et al. 2006). However, claims that uncertainty information enhances weather forecast value are

Corresponding author address: Susan Joslyn, Dept. of Psychology, University of Washington, Box 351525, Seattle, WA 98125.
E-mail: susanj@u.washington.edu

generally based on normative or prescriptive decision-making models (Richardson 2002; Palmer 2002). By contrast, there is ample research suggesting that experienced forecasters understand and reliably estimate credible intervals and probability forecasts (Murphy and Winkler 1974a,b, 1977). However, some recent evidence suggests an overforecasting bias when safety was an issue (Keith 2003). In sum, there is no research of which we are aware that attempts to gauge the impact of uncertainty information on realistic deterministic forecasts.

The study presented here was conducted to determine whether wind speed or high wind advisory [wind speeds in excess of 20 kt (10.29 m s^{-1})] forecasts differ as a result of reviewing charts indicating the probability of wind speeds exceeding 20 kt. The participants made the forecasts for marine areas in which small boats would be affected. The hypothesis was that probabilistic information would impact a threshold warning forecast in this context. The reasoning was as follows: forecasters might decide to post an advisory even if the probability for high winds was small to reduce the risk of boating accidents. Reviewing explicit probabilistic information would alert them to such situations.

2. Method

a. Participants

Ten University of Washington atmospheric science students participated in this study. All participants had completed basic instruction in forecasting. Three of the participants were undergraduate students who had completed a course in atmospheric structure and analysis. Seven participants were graduate students. The participants were paid \$40 for participating in the two-session study. They were paid \$20 after their first session and \$20 after their second session.

b. Task

The participants made four forecasts over two sessions. For each forecast they were asked to review historical information collected at approximately 2130 UTC (1330 local time) and to predict wind speed and direction for four locations in the Puget Sound region: Smith Island, Destruction Island, West Point, and Tatoosh Island. As part of a single forecast date, they predicted wind speed and direction, for each location, every 6 h over a 48-h period that began at 0000 UTC the next day. As a result, each forecast had a minimum lead time of 2.5 h and a maximum lead time of 50.5 h. This resulted in 36 wind speed forecasts in all (Table 1). The participants were also asked to decide whether

TABLE 1. Forecast dates and hours for participant 1. For each of the listed lead times, the participant made forecasts for Destruction Island, Tatoosh Island, Smith Island, and West Point. The use of the probability product was rotated between dates for each participant.

Forecast date	14 Feb	20 Feb	11 Mar	21 Mar
Probability product	No	Yes	No	Yes
Valid dates	15 Feb	21 Feb	12 Mar	22 Mar
	16 Feb	22 Feb	13 Mar	23 Mar
	17 Feb	23 Feb	14 Mar	24 Mar
Lead times (h)	2.5, 8.5, 14.5, 20.5, 26.5, 32.5, 38.5, 44.5, and 50.5			
Valid times (UTC)	Days 1–2: 0000, 0600, 1200, 1800			
	Day 3: 0000			

they would issue a high wind advisory, indicating that they expected wind speeds to exceed 20 kt, for any of the four forecast locations during the 48-h period. If so, they were asked to indicate the hours during which the advisory should be posted. The concept of the wind advisory, its purpose, and techniques for predicting it had been covered in the completed coursework. For the purposes of this exercise, a high wind advisory was defined to be winds greater than 20 kt. The participants were asked to disregard wave height, which is usually a factor in small craft advisories.

c. Materials

1) WEATHER DATA

The participants were provided with historic weather information that had been downloaded from the World Wide Web as data upon which to base the wind forecast. The selection of information sources was based on the results of an observational study of forecasters who produced a similar forecast at a nearby naval station. The participants in the present study had several products from prominent models, including the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5), the Nested Grid Model (NGM), and the Aviation Model (AVN). The sources also included satellite and radar imagery, regional terminal airdrome forecasts (TAFs), meteograms, and observations taken from buoys in locations near the forecast sites (see the appendix for a complete list).

In half of the trials participants were also provided with a probability product. It was a chart, color-coded for the probability of 10-m winds in excess of 20 kt (Fig. 1). The chart was based on the centroid mirroring ensemble (ACME) in the 12- and 36-km domains (Grimt and Mass 2003). ACME consists of 17 individual forecasts (called ensemble members) all using the MM5 with an identical physics package, but different bound-

nails, thumbnail pages, and image pages with navigation links. All product-image Web pages were connected to a simple main links page with a link that led to the main thumbnail page or information source. Thus, the user saw the main page with text links to individual products such as the satellite, radar, and model products. Clicking on a text link to a product opened a page of thumbnail images for that product, which could then be clicked on to display full-size images. The participants could then navigate through the images using the forward and backward buttons. An up arrow navigation button reconnected with the thumbnails page and a home button returned the browser to the main links page. TAFs, meteograms, and buoy observations opened in a separate window. When users closed the window, they returned to the main links page. The experimenter demonstrated these procedures to the participants.

3) ANSWER SHEETS

There were two answer sheets for each forecast: the wind speed and direction answer sheet and the wind advisory answer sheet. The wind speed and direction answer sheet provided four columns, each headed by the name of one of the four locations for which a forecast was required. There was a row designated for each of the nine forecast hours with a blank for the wind speed and direction. Forecasters were asked to record wind speed and direction to make the task realistic; however, there were no specific hypotheses concerning them. Nor were there any significant implications from a preliminary analysis of these variables. Therefore, they will not be discussed further.

The participants recorded the wind advisory on a separate sheet to reduce the influence of the wind speed forecast on the wind advisory. This procedure was used to discourage the participants from simply posting a wind advisory for time periods during which the deterministic forecast was over 20 kt. If the participants were not looking at their wind speed forecast, they might make a separate wind advisory decision, taking the probabilistic information into account. On the wind advisory answer sheet there were four questions asking whether participants would forecast a high wind advisory during the 48-h forecast period for each of the four locations and if so, to indicate during which hours.

4) MAP

Although the forecast locations were not marked on the probability charts, the participants were provided with a separate map showing the four forecast locations

as well as the locations of nearby airfields for which they had the TAFs.

d. Procedure

The participants were tested individually in two approximately 1.5-h sessions. In the first session, after informed consent procedures, the experimenter explained the forecasting task and how to fill out the answer sheets. The forecast locations were pointed out on the map that was posted above the workstation. The experimenter demonstrated how to access information on the computer and explained that the participants could use whichever sources of information they wished.

Then the experimenter introduced the probability product and informed the participants that it would also be available on some forecasts. The experimenter demonstrated how to read the probability chart and how it was generated. To ensure that the participants read and understood the probability information in trials with the probability product, they were required to record the probability ranges for each of the four locations at each of the eight forecast times for which the product was available (beginning at 0600 UTC). This procedure was initiated after several of the pilot participants ignored the probability product altogether, commenting that it was not useful for the forecast they were required to make. Because the impact of the probability product was the focus, it was necessary to ensure that the participants had encoded the information.

When all of the participants' questions had been answered, they made a practice forecast with the probability product to familiarize themselves with the procedure and the interface. Upon finishing the practice forecast, the experimenter checked the answer sheets for completeness. Unless there were further questions, the participant then made two test forecasts: one with and one without the probability product. The session was not time limited but generally took between 1.5 and 2 h. The participants returned for a second session less than 1 week later to complete the remaining two weather forecasts.

e. Design

Probability information was manipulated within the participants. Each participant had two trials *with* the probability product and two trials *without* the probability product. Half of the participants began the session with the probability product and half did not. Weather data date order was similarly counterbalanced. The weather data dates were rotated through conditions so that each day was used in a forecast with the probability

product on half of the trials and a forecast without it on the other half. Rotation ensured that weather conditions were equivalent across conditions; that is, the same 4 days were used in the conditions with and without the probability product. This was important because the weather on some dates might have been easier to forecast accurately than on other dates. However, no participant saw data from the same date twice. In other words, an individual participant forecasted wind speed for different dates with and without the probability product so that each forecast was unfamiliar. Table 1 shows the forecast dates and hours for an example participant.

3. Results

The study was designed to investigate the effect of the probabilistic information (the probability of wind speeds exceeding 20 kt) on decisions to post a high wind advisory. First, however, the participants' ability to read the probability product was examined to ensure that the participants had encoded the probability information accurately. Then, the condition with the probability product was compared with the condition without the probability product to evaluate the wind advisory decisions.

a. Reading the probability product

Recall that, in the condition in which the participants were given the probability product, they were required to record the probability range indicated by the product on the answer sheet. The group as a whole recorded the probability ranges for all four forecast dates, although an individual participant used the probability product for only two dates. For each date, the participants recorded the range for the eight forecast periods for which the product was available, at each of the four geographic locations. We will refer to each of those forecast times and locations as a case. Thus, there were 128 cases in all (four days \times four locations \times eight times). To assess the consistency with which the participants read the probability of winds greater than 20 kt, the agreement between the participants was calculated.¹ Agreement was the percentage of participants who recorded the exact same range of probabilities for

¹ This is an estimate of reading accuracy, as the forecast locations were not marked on the probability product itself so there was no objective answer to the range of probabilities displayed. Only the cases in which more than half of the participants agreed on the interpretation of the probability product were included.

a given location at a given time period divided by the total number of participants reading the forecast for that date.² Summed across locations and time periods, the participants agreed in 80% (102/128) of the cases. In other words, the participants disagreed in the interpretation of probability product in 20% (26/128) of the cases. There are several possible explanations for this. The forecast sites were not labeled directly on the probability product, so locations were inferred using a separate map. In addition, some of the colored areas were small and the boundaries between them were difficult to distinguish in the graphic. All participants' responses to the 26 "disagreement" cases were omitted from subsequent analyses, on the assumption that the chart itself was difficult to read in those cases. As such, it would not be fair to compare performance with and without that chart for those times and locations (cases).

b. Wind advisory analyses

Next, the influence of the probability product on the wind advisory was examined. Accuracy for posting the wind advisory was defined in terms of the signal detection measure of sensitivity³ (Green and Swets 1966). Sensitivity is the degree to which the participant can discriminate between a high wind event and a non-high wind event, independent of the response bias. Response bias is the participant's overall willingness to post an advisory.⁴ A given response is a combination of these two factors, sensitivity and response bias. As such, the accuracy of a single wind advisory is not particularly informative. It could be due to a liberal response bias or to real sensitivity to the underlying conditions. In a signal detection analysis, sensitivity and response bias can be calculated separately. For a similar approach see Keith (2003) or Mason (1982, 2003). To compute sensitivity d' , the hours during which the participant posted an advisory were compared with the hours during which the observed wind speeds exceeded

² In some cases, even when directed to write down a range, the participants wrote down a single number. If the value could only be derived from one of the six possible ranges, it was assigned to that range. For instance if a participant wrote "0," he was given credit for "0-10." If the single value could be interpreted as part of more than one, e.g., "10," it was omitted from this analysis.

³ Here, $d' = z\text{Hits} - z\text{FAs}$, where z is the proportion of hits and FAs transformed into standard deviation units (distance from the mean in a standard normal distribution for that score). Normal deviates can be derived from normal tables, or the NORMDIST function in Microsoft's Excel software program.

⁴ Here, $C = 0.05(z\text{Hits} + z\text{FAs})$.

TABLE 2. Hits, FAs, misses, and correct rejections for each forecaster. These were ascertained by comparing the hours for which the forecaster posted an advisory with the hours during which observed winds exceeded 20 kt. Slight differences in the total number of hours occur because of a few missing verification data points.

Forecaster	1	2	3	4	5	6	7	8	9	10	Tot
With probability product											
Correct rejection	31	27	26	31	23	21	27	22	14	46	268
FA	6	6	7	6	9	12	10	10	6	4	76
Hit	13	8	9	16	6	9	15	18	13	0	107
Miss	8	4	3	3	7	3	5	1	19	1	54
Without probability product											
Correct rejection	29	35	20	22	19	14	8	17	31	10	205
FA	4	2	17	8	18	23	25	20	18	11	146
Hit	5	8	18	13	19	21	8	11	1	18	122
Miss	7	13	3	2	2	0	4	4	1	13	49

20 kt.⁵ Then, forecast hours were divided into the following four categories. Hits were defined as cases in which the winds were greater than 20 kt and the participant posted a wind advisory. Misses were cases in which the winds were greater than 20 kt but the participant did not post an advisory. False alarms (FAs) were cases in which the participant posted an advisory and the winds were less than 20 kt. Correct rejections were cases in which the winds were less than 20 kt and the participant did not post an advisory (Table 2). For the sensitivity measure (d'), higher scores indicate a greater ability to discriminate between a high wind event and a non-high wind event. The mean d' was greater in the condition with the probability product ($d'_{\text{with}} = 1.25$) than in the condition without the probability product ($d'_{\text{without}} = 0.92$).

For the response bias C , a value greater than 0 indicates a conservative approach (i.e., a reluctance to post advisories) and a value less than 0 indicates a liberal bias. When using the probability product, the participants had a more conservative response bias ($C = 0.11$) than they did without it ($C = -0.19$). That is, contrary to our original prediction, the participants tended to post *fewer* advisories with the probability product (38% of the time) than without it (45% of the time) with no reduction in sensitivity. Although none of these differences quite reached significance, the implications were intriguing.

To further investigate the impact of the probability product on the frequency of posting a wind advisory, the percentage of times the participants posted an advisory for each of the probability ranges displayed in

the product (0%–10%, 10%–30%, 30%–50%, 50%–70%, 70%–90%, and 90%–100%) was calculated. The number of cases in which the participants issued an advisory in a given range was divided by the total number of cases in which that range was identified, to determine the percent advisories issued. Then, the percent advisories issued in the conditions with and without the probability product were plotted over probability ranges (Fig. 2). For reference, a line matching the probability to frequency (perfect match) is included. The latter can be interpreted as the hypothetical pattern of responses that perfectly reflects the probability product's forecast.

There were some similarities between the forecasts with and without the probability product. In both conditions the participants posted more advisories as the probability of high winds increased. This is not surprising in that the participants had the model-produced deterministic prediction for all forecasts. In general, model-predicted wind speeds increase as the likelihood of high wind increases. When comparing the human forecasters with the model (perfect match response), note that the participants tended to post advisories in a larger percentage of cases than was indicated in the lower ranges (0%–30%). However, this bias was reversed when the probability of high winds was very high. In the very highest probability category (90%–100%) the participants issued a smaller percentage of advisories than was indicated. In these ranges the probability product was particularly well calibrated when compared with the actual occurrences of high winds. High winds were observed in about 2% of the cases for the 0%–10% range, 32% of the cases for the 10%–30% ranges, and 100% for the 90%–100% ranges. Thus, the human forecasters were too liberal in their willingness to issue a wind advisory when the likelihood was low and too conservative when the likelihood was high.

⁵ We examined only the 120 h per participant for which a frame of the probability product was provided and during which participants agreed on the value represented by the product for that location.

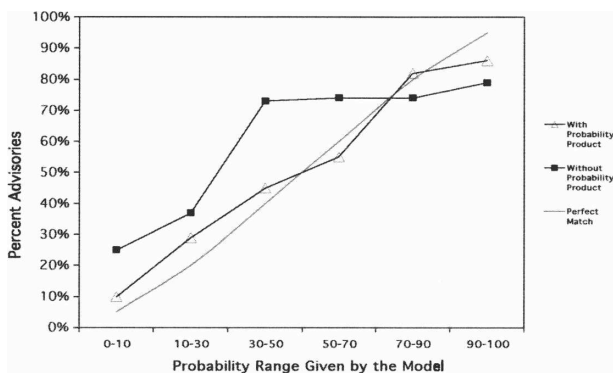


FIG. 2. Percent of advisories posted with and without the probability product, for each probability range given by the model. The diagonal represents the hypothetical response frequency that perfectly matches the probability displayed by the product.

Importantly, these two tendencies, a liberal bias at lower probabilities and conservative bias at higher probabilities, were attenuated by the probability product. With the probability product, the participants posted fewer advisories than without it in the lower ranges (0%–30%) and more advisories than without it in the very highest range (90%–100%). This effect was statistically significant in an analysis of variance (ANOVA). Forecasts were divided into two comparable categories, one in which the probability of high winds was high (90%–100%) and one in which the probability of high winds was low (0%–10%).⁶ Then, within each category, they were further subdivided into forecasts conducted with and without the probability product. Finally a 2 (probability of high winds: high versus low) \times 2 (probability product: with versus without) repeated-measures ANOVA was conducted to determine whether the differences in mean percent advisories posted with and without the product were statistically significant. The ANOVA yielded a main effect for the probability of high winds, the ratio of mean square treatment and mean square error (MSE) $F(1, 9) = 329.33$, $MSE = 4.52$, and probability $p < 0.0001$. This means that, not surprisingly, people posted significantly more advisories when the likelihood of high winds was high [$M = 85\%$, standard deviation (SD) = 17%], regardless of whether they had the probability product, than when it was low ($M = 17\%$, $SD = 2\%$). Importantly, the interaction of the probability of high winds and the use of the product was also significant, $F(1, 9) = 6.9$, $MSE = 0.09$, and $p < 0.05$. The probability product had a significantly different effect on posting deci-

sions, depending on the likelihood of high winds. People posted fewer advisories with the product ($M = 12\%$, $SD = 21\%$) than without the product ($M = 23\%$, $SD = 17\%$) when the likelihood of high winds was below 10%. The opposite pattern was observed when the likelihood of high winds was above 90%. The participants posted more advisories with the product ($M = 88\%$, $SD = 15\%$) than without it ($M = 81\%$, $SD = 17\%$). This suggests that the probability product discouraged the participants from posting advisories when the likelihood of high winds was low and encouraged them to post more advisories when likelihood was high.

4. Conclusions

These results suggest that the probability product improved the threshold forecast: posting high wind advisories. It appears to have had its effect by counteracting the natural biases in high and low likelihood situations. It has long been known that people do not treat probability linearly (Kahneman and Tversky 1979; Gonzalez and Wu 1999). In this study, the participants had a liberal bias in the lower-probability ranges and a conservative bias in the very highest range. A similar pattern was observed in the probability estimates of experienced forecasters over extended forecast periods (Baars et al. 2004) and when safety is an issue (Keith 2003).

The tendencies observed in the study reported here may be related to the warning task assigned to the participants. The participants may have been sensitive to different errors when the likelihood of high winds was very high than they were when the likelihood was low. Perhaps the participants attempted to minimize their misses in the low probability situations, leading them to post more advisories than were warranted. In high-probability situations they may have shifted their focus to FAs, causing a reduction in the number of advisories posted. Although this is mere speculation in the context of the present data, there is evidence that the severity of an outcome and the sensitivity to loss affect the interpretation of even precisely quantified uncertainty (Weber 1994; Windschitl and Weber 1999).

From a practical standpoint, a liberal bias makes sense in the context of the high wind warning task studied here. The purpose of the wind advisory was to prevent boaters from setting sail in conditions of dangerously high winds. The participants may have chosen to err on the side of caution in the lower-probability ranges by posting an advisory even when the chance of high winds was small. However, in real-life situations an overly liberal bias could lead to problems. Boaters may begin to disregard the advisory if it proves to be wrong too often and high winds fail to materialize. The user's

⁶ Two missing data points for two participants in one category range (90%–100%) were estimated by calculating the average percent advisories posted for that group.

FA tolerance is thought to be critical to the success of such warnings (Roulston and Smith 2004). Thus, for these situations the use of probabilistic information by the forecaster may be especially important. In the study reported here, the participants posted fewer advisories in the lower probability ranges when using the product than they did without it, reducing the overall number of FAs (15% FAs with the probability product versus 28% FAs without). This improvement could be critical in real-life situations in which trust in the advisory system is crucial for boater safety.

The participants were reluctant to post advisories as often as was warranted in the very highest category (90%–100%). This tendency is also problematic in a real-life situation in which small boaters could be endangered by setting sail in high wind situations of which they were not warned. Again, the probability product attenuated this effect. When the participants used the probability product, they posted more advisories when high winds were very likely than they did when they did not use it.

It is important to remember that the same participants and the same weather data were used in both conditions. The only difference between the conditions was the probability product itself. Thus, the probability product had an important positive impact upon counteracting two problematic biases and improving the threshold forecast decisions. There is now strong evidence that probabilistic information is beneficial for a realistic deterministic forecast decision among forecasters with a moderate level of experience. Because the biases counteracted by the probability product have been observed among forecasters with greater experience and on different threshold decisions (Keith 2003; Baars et al. 2004), we believe that it is likely that probabilistic information such as this will be beneficial for a wider range of tasks and populations as well.

Acknowledgments. This research was supported by the DOD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research under Grant N00014-01-10745. Special thanks to Meng Taing for editing and preparing this document.

APPENDIX

Complete List of Products Available to Participants

MM5 (1200 UTC run)

- Sea level pressure (SLP), 10-m winds, topography or 925-mb temperature, 36-km domain (72 h, 25 frames)

- SLP, 10-m winds, topography or 925-mb temperature, 12-km domain (72 h, 25 frames)
- SLP, 10-m winds, topography or 925-mb temperature, 4-km domain (48 h, 17 frames)
- 850-mb heights, temperature, winds, 12-km domain (72 h, 25 frames)
- Subdomain SLP, 10-m winds, 925-mb temperature, 4-km domain (48 h, 17 frames)
- Surface wind speed, 4-km domain (48 h, 17 frames)
- 500-mb heights, temperatures, winds, 12-km domain (72 h, 25 frames)
- 3-h precipitation, 12-km domain (72 h, 23 frames)
- Meteograms
 - NWS Seattle
 - Port Angeles
 - Quillayute
 - Victoria, British Columbia, Canada

Buoys

- Smith Island
- Destruction Island
- Tatoosh Island
- West Point

Satellite imagery

- Enhanced 4 km
- Infrared 4 km
- Infrared enhanced 4 km
- Visible 4 km

TAFs and current aviation routine weather reports (METARs)

- Whidbey Island, Washington
- McChord AFB, Washington
- Seattle–Tacoma International Airport
- Portland, Oregon
- Hoquiam, Washington
- Bremerton, Washington
- Everett, Washington
- Bellingham, Washington
- Port Angeles, Washington
- Fairchild AFB, Washington
- Moses Lake, Washington
- Pasco, Washington
- Friday Harbor, Washington
- Victoria, British Columbia, Canada

Radar

- Base reflectivity elevation 1
- Base radial velocity elevation 1

AVN (0000 UTC run)

- 850-mb winds, heights, temperatures

NGM (0000 UTC run)

- 850-mb winds, heights, temperatures

Probability stimulus

- MM5 ensemble probability of winds greater than 21 kt
 - ACME 36 km (48 h, 8 frames)
 - ACME 12 km (48 h, 16 frames)
 - ACME core 36 km (48 h, 8 frames)
 - ACME core 12 km (48 h, 16 frames)

REFERENCES

- Anderson, O., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530.
- Baars, J. A., C. Mass, and M. Albright, 2004: Performance of National Weather Service forecasts versus model output statistics. Preprints, *20th Conf. on Weather Analysis and Forecasting and 16th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., CD-ROM, 20.5.
- Gonzalez, R., and G. Wu, 1999: On the form of the probability weighting function. *Cogn. Psychol.*, **38**, 129–166.
- Green, D. M., and J. Z. Swets, 1966: *Signal Detection Theory and Psychophysics*. Wiley, 455 pp.
- 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.
- , and —, 2003: Toward short-range ensemble prediction of mesoscale forecast skill. Preprints, *10th Conf. on Mesoscale Processes*, Portland, OR, Amer. Meteor. Soc., 22–27.
- Kahneman, D., and A. Tversky, 1979: Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263–291.
- Keith, R., 2003: Optimization of value of aerodrome forecasts. *Wea. Forecasting*, **18**, 808–824.
- Mason, I. B., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- , 2003: Binary events. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, I. T. Jolliffe and D. B. Stephenson, Eds., Wiley, 37–76.
- Murphy, A. H., and R. L. Winkler, 1974a: Credible interval temperature forecasting: Some experimental results. *Mon. Wea. Rev.*, **102**, 784–794.
- , and —, 1974b: Subjective probability forecasting experiments in meteorology: Some preliminary results. *Bull. Amer. Meteor. Soc.*, **55**, 1206–1216.
- , and —, 1977: Can weather forecasters formulate reliable probability forecasts of precipitation and temperature? *Natl. Wea. Dig.*, **2** (2), 2–9.
- National Research Council, 2006: *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts*. Committee on Estimating and Communicating Uncertainty in Weather and Climate Forecasts, The National Academies Press, 124 pp.
- Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quart. J. Roy. Meteor. Soc.*, **128**, 747–774.
- Richardson, D. S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Meteor. Soc.*, **127**, 2473–2489.
- Roulston, M. S., and L. A. Smith, 2004: The boy who cried wolf revisited: The impact of false alarm intolerance on cost-loss scenarios. *Wea. Forecasting*, **19**, 391–397.
- , G. E. Bolton, A. N. Kleit, and A. L. Sears-Collins, 2006: A laboratory study of the benefits of including uncertainty information in weather forecasts. *Wea. Forecasting*, **21**, 116–122.
- Weber, E. U., 1994: From subjective probabilities to decision weights: The effect of asymmetric loss function. *Psychol. Bull.*, **115**, 228–242.
- Windschitl, P. D., and E. Weber, 1999: The interpretation of “likely” depends on the context, but “70%” is 70%—Right? The influence of associative processes on perceived certainty. *J. Exp. Psychol. Learn. Mem. Cogn.*, **25**, 1514–1533.