Climate Forecast Maps as a Communication and Decision-Support Tool:

An Empirical Test with Prospective Policy Makers

Toru Ishikawa[1], Anthony G. Barnston[2], Kim A. Kastens[1,3],

Patrick Louchouarn[1,3], and Chester F. Ropelewski[2]

[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964 (ishikawa@ldeo.columbia.edu, kastens@ldeo.columbia.edu)

[2]International Research Institute for Climate Prediction, Columbia University, Palisades, NY 10964 (tonyb@iri.columbia.edu, chet@iri.columbia.edu)

[3]Department of Earth and Environmental Sciences, Columbia University, New York, NY 10027 (pl2065@columbia.edu)

Corresponding author: Toru Ishikawa, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964. Phone 845-365-8568; Fax 845-365-8179; E-mail: ishikawa@ldeo.columbia.edu

Abstract

This paper reports an empirical study of communication issues concerning climate forecasts. Students in a professional master's degree program in environmental science and policy participated in the study as prospective policy makers. Participants viewed a set of currently issued precipitation forecast maps, and answered questions designed to assess, in the context relevant to agricultural and environmental decision making, their understanding and evaluation of the maps. Participants failed to understand some aspects of the information shown on the maps, in the current design, as the map makers intended. In particular, participants had difficulty with understanding probability forecast maps, and distinguishing probabilistic three-category forecasts and the amount of precipitation. Most participants evaluated the degree of agreement between the forecast and observation as *agree only slightly* or *agree somewhat*. More than half of the participants were not inclined to use the forecasts in agricultural decision making. Implications for improvement in design for better communication are discussed.

KEYWORDS: Cartographic communication, climate forecasts, probabilities; uncertainty, decision making, policy makers.

Introduction

Since antiquity, humans have been using maps to represent space. By doing so, one can "look at" and think about the represented space while not actually standing within that space. On a map, one can record and organize information about the represented space at a reduced scale, thus making it possible to think about regions impossible to view in their entirety from any single vantage point. And, importantly for this study, one can communicate information about the represented space to another human being.

In this paper, we examine communication issues concerning climate forecasts, using currently issued forecast maps targeted to a broad spectrum of decision makers, including agricultural and environmental policy makers. Climate forecast maps differ from the more-studied weather maps (Monmonier 2000), in that the forecast looks months to seasons into the future, rather than hours to days, and thus incorporates a greater degree of uncertainty. Climate has long been of major concern to humans due to its potential impact on their activities. Researchers have been attempting to offer seasonal or interannual climate forecasts several months ahead of time, so that people can make important decisions that are vulnerable to future climate conditions (e.g., grain production, water management, natural disaster control). Skills of climate prediction have advanced in the last few decades, especially since the success of an experimental, model-derived prediction of El Niño in the late 1980s (Cane and Zebiak 1986; see also Goddard et al. 2001 for a review of the current state of prediction efforts).

Current climate prediction models are generally based on the El Niño-Southern Oscillation (ENSO), which refers to shifts in sea surface temperatures in the equatorial Pacific and related shifts in barometric pressure gradients and wind patterns in the tropical Pacific. Researchers have shown that ENSO activity has a large global impact on interannual climate variability and, importantly for this study, that it is highly correlated with agricultural production. For example, Cane et al. (1994) showed that more than 60% of variance in maize yield in Zimbabwe was accounted for by an index of ENSO, and that model-derived predictions of ENSO provided fairly accurate forecasts of maize yield. Other researchers have also discussed potential benefits of climate forecasts to agriculture, at various parts of the world and at various scales (e.g., Hammer et al. 2001; Hansen 2002; Jones et al 2000).

Such potential benefits of climate forecasts, however, cannot be fully realized unless people understand and use climate forecasts appropriately. Some researchers discussed the difficulty that people have in interpreting and applying climate forecasts in practice, and argued for the necessity of systematically examining communication issues about climate forecasts so that people can take advantage of their potential benefits (e.g., Nicholls 1999; Pfaff et al. 1999). We consider two possible sources of difficulty in understanding climate forecasts: (a) the probabilistic nature of climate forecasts and (b) the complex presentation formats of current forecast products.

First, in climate prediction, it is not possible to offer a deterministic forecast; instead, forecasts are currently given as probabilities of predicted precipitation and temperature for specific regions falling into the upper, middle, and lower one-third in the past years' database for the regions. There is an extensive literature on human understanding of probabilities and decision making under uncertainty (e.g., Kahneman et al. 1982), which shows that people do have difficulty understanding probabilities. For instance, people make probability judgments on the basis of heuristics, such as "representativeness" and "availability," which lead to serious biases (Tversky and Kahneman 1974). And people (not only laypersons but also professionals such as physicians) tend to make errors in statistical reasoning, particularly to neglect base rates in Bayesian inference problems (Eddy 1982; Tversky and Kahneman 1982). When the same problem is presented in terms of frequencies rather than probabilities, more people come to the correct answer (Gigerenzer and Hoffrage 1995; Hoffrage et al. 2000). Also, while the forecaster attempts to convey, in a probability format, the existence of uncertainty inherent in climate forecasts, people in general feel uncomfortable with data containing uncertainty, and are reluctant to use such forecasts in decision making (Changnon 1992; Changnon et al. 1995; Slocum et al. 2003).

Second, current forecast maps are often presented in quite complicated formats, and people need to refer to and integrate multiple pieces of information, such as probabilistic three-category forecasts, threshold values separating the three categories, and the mean or median precipitation in the past years (details described in the Materials section below). People generally have difficulty integrating separately presented information into a whole picture. For example, Lloyd and Bunch (2003) presented geographic information in different formats to participants, and asked them to answer true-or-false questions from memory about geographic orientations of places. Participants who viewed the base map in its entirety answered the questions more accurately than participants who learned partial information on the base map separately (divided into nine sections, distributed across three layers, or shown at three cartographic scales).

The literature of cognitive cartography offers relevant research findings about communication based on spatial representations, or maps. Map-based communication entails two human parties, the map maker and the map reader, and two nonhuman entities, the map and the represented space (Figure 1A). Many researchers (e.g., Eastman 1985; Koláčný 1969; Robinson and Petchenik 1975) have pointed out that this communication is not unidirectional with pre-encoded information being transmitted to, and decoded by, the recipient. Rather, the map reader plays an active role in interpreting the map in light of his or her prior knowledge and experience (Gilhooly et al. 1988; Williamson and McGuinness 1990) and ability (Sholl and Egeth 1982; Underwood 1981). Also, the effectiveness of different map formats depends on the map user's purpose or goal (Patton and Cammack 1996; Phillips et al. 1975). The map reader's knowledge, in turn, is modified as a result of interacting with the map (cf. schemata as explained by Neisser 1976).

As for the influence of map design on map understanding, perception of individual map symbols has been extensively studied with respect to graduated symbol maps (e.g., Ekman et al. 1961; Flannery 1971; Olson 1975) and choropleth maps (e.g., Brewer et al. 1997; Bunch and Lloyd 2000; Lloyd and Steinke 1977). Studies of map symbols are generally based on Bertin's semiotics, which discusses the appropriateness of various graphic variables, such as size, texture, color, orientation, and shape, for depicting quantitative and qualitative information on a map (see, e.g., MacEachren 1995, 270). On the other hand, some researchers stressed the necessity of examining the understanding of the total map, not individual map symbols, as the level at which "meaning" is conveyed in a specific context (e.g., Dent 1972; Guelke 1976).

Although the present study is in line with these cognitive-cartographic studies, there is an important distinction that characterizes our study. The materials used in this study are climate forecast maps that have been disseminated for use by a broad spectrum of decision makers. Also, the maps and the represented information are quite complicated. As discussed above, communication of climate forecasts has been identified as an important issue to be considered, but it has not been studied empirically with human participants, except for interviews and surveys. Thus, this study aims to examine how well people understand forecast maps in the current design, and to obtain insights that could be further investigated in a future controlled experiment of alternative designs.

The specific audience of map readers in this study are graduate students in training in Columbia University's Master's of Public Administration (MPA) in Environmental Science and Policy, seeking to enter public service careers in the environmental arena. Current and future policy makers are an interesting and suitable audience for this study. First, policy makers examine maps with the goal of extracting information that can be used in decision making. In this regard, it is not sufficient that they merely "understand" the map; they must also "believe" that the map is an accurate representation and trustworthy enough to incorporate into decision making.

Second, policy makers wish to generate desirable outcomes for future events, and so they seek maps (and other forms of information) that predict future conditions that might influence their actions. Thus, when policy makers are the audience for a map, an interesting feedback pathway develops in cartographic communication (Figure 1B): Based on information contained in the map, the policy maker shapes a policy; the implementers of the policy, in turn, may cause change within the represented space. As an example, we see the decision to convert natural wetlands in South Florida into agricultural land (Marshall et al. 2003). The decision to relocate winter crops from north to south was made because of the damaging winter freezes that occur to the north, but due to the large-scale drainage of wetland, the frequency and severity of agriculturally damaging freezes have been inadvertently increased in the south. Hence, map reading (temperature delineation of the region) did influence policy making, which in turn affected the represented space, albeit in a direction opposite to the desired outcome.

In the study reported below, we presented a suite of climate forecast maps to prospective policy makers, and asked them questions designed to assess, in the context relevant to agricultural and environmental decision making, their understanding and evaluation of the forecast information shown on the maps. Major objectives of this study were to see if people have difficulty with currently issued forecast maps, and if so, to identify what aspects of the forecast maps are difficult to understand.

Method

Participants

The participants were 47 students (20 men and 27 women) in Columbia University's MPA program in Environmental Science and Policy, from the School of International and Public Affairs. This intensive and multidisciplinary program brings together people with backgrounds as varied as public policy, economics, earth and biological sciences, engineering, social sciences, and humanities. Past experience at public service is part of the admission criteria for this program and thus a large proportion of students are recruited from the workforce rather than directly out of undergraduate degrees. Usually 15-20% of the students are international students, who come mostly from Asia, as well as Western and

Eastern Europe. Throughout the course of this one-year program, the students are trained as environmental professionals who have an understanding of the science of environmental issues and use this knowledge to develop better management strategies, analytical tools that incorporate science in the process of policy formulation, and communication skills that permit a more effective transfer of complex information. The study participants, who have now graduated, are working in Federal and State Environmental and Energy Regulatory Agencies, as Presidential Management Fellows (within NASA and EPA), in NGOs acting on biodiversity and natural resource management, and in environmental consulting groups, so there is little doubt that they are recognized as decision makers. The particular group studied here ranged in age from 21 to 39 (mean age = 26.8 years).

Materials

We presented a set of five maps to each participant (Figures 2, 3, and 4). These maps are actual forecast maps issued by the International Research Institute for Climate Prediction (IRI) in September 2002 and January 2003, and are part of a large suite of maps provided by this organization for the use of individuals and organizations who are to make various decisions that are vulnerable to seasonal and interannual climate variability (http://iri.columbia.edu/climate/forecast/net_asmt). Each map covers North America, Central America, and northernmost South America. On the maps, latitude and longitude are shown at constant intervals; the effects of nonequal-area projection are minimal in the tropics, where the maps have the most important users as the targeted audience. For details of the IRI seasonal climate forecasts, see Barnston et al. (2003) and Goddard et al. (2003).

Two of the maps (Figures 2 and 3A) convey a forecast of precipitation for a 3-month "season" beginning with the month following the issuance of the prediction. The forecast is derived from the output of a set of climate prediction models that are based on the recent state of the atmosphere and oceans. For some areas during some months, the model has little predictive power, and these areas remain white on the map. Such white areas indicate equal chances (33.3%) that the precipitation will be below, near, or above normal. For areas where the model does make a prediction, color is used to indicate the category having the highest likelihood among the three: below normal (yellow to brown), near normal (gray), or above normal (light green to purple). *Above normal*, in this context, means that the predicted precipitation falls within the wettest one-third of the years in a 30-year database of precipitation observations, *below normal* within the driest tercile, and *normal* within the center tercile. Within the below-normal and above-normal categories, color gradations are used to indicate the probability of the prediction. For example, a yellow color means that the model finds a 40% probability that the precipitation will fall within the driest tercile of years for those months at that location, and a dark brown color indicates a 70% probability.

One color per region represents only the modeled probability of precipitation falling within the single most likely tercile, and so small bar graphs are superimposed upon the map to indicate the probabilities of precipitation falling within each of the three terciles. For example, the dark green area blanketing part of California in Figure 2 has a forecast of a 50% chance of above-normal precipitation, a 30% chance of normal precipitation, and a 20% chance of below-normal precipitation. “Dry season” masking is applied when the location receives an average of less than 30 mm of precipitation over the 3-month period. When such a small amount of rainfall is involved, the boundary between the below-normal and normal categories becomes close to 0 mm and the two categories cannot be well distinguished.

Three of the maps (Figures 3B, 4A, and 4B) represent historical precipitation, based on 30 years of weather-station observations. Figure 3B shows the observed precipitation during October-November-December 2002. The observed precipitation is shown as *precipitation anomaly*, which is defined as the ratio of the observed precipitation at each location during that specific 3-month period to the median precipitation in the 30-year database. Figures 4A and 4B show the threshold values of precipitation in millimeters, that is, the boundaries between the above-normal and normal categories (above-normal threshold), and between the normal and below-normal categories (below-normal threshold), of precipitation for a given place at a given time of year. By referring to and combining the probability forecast maps and the threshold maps, a user can obtain the precipitation amounts corresponding to the location of interest where a probabilistic forecast is made.

Test Questions

Presented with the precipitation forecast maps described above, participants were asked three sets of questions designed to assess their understanding and interpretation of the data shown on the maps.

Question 1

The first map (Figure 2) is a probability forecast map of precipitation for February-March-April 2003. The map was made in January 2003. It shows in color the probability that individual locations would have more or less precipitation than normal. *Normal* is defined as being in the center one-third of the years in a 30-year database. *Above normal* is defined as being in the wettest one-third of the years, and *below normal* is defined as being in the driest one-third of the years, in the 30-year database. The horizontal bar graphs shown on the colored areas indicate the probability of forecasts falling into, from top to bottom, the *above normal*, the *normal*, and the *below normal* categories.

Suppose that you are given this forecast map in January 2003. Based on this map, how would you answer the following question: "Which area will receive a greater amount of total precipitation for this forecast period, Southern California or Washington State?"

Please explain in a few sentences how you came to the answer (or the difficulty you had in trying to answer the question).

The first question was designed to assess participants' understanding of a standard IRI precipitation forecast map (Figure 2). Following a brief explanation of the map, the question asked which would receive a greater amount of precipitation, Southern California or Washington State. A few participants indicated that they were having trouble pinpointing these localities on the map; to these participants, the investigators individually pointed out the exact locations (same with the questions below). To answer this question correctly, participants need to understand that the map only shows the likelihood that a specific region is likely to receive below, near, or above its own normal precipitation for that region and time of year. Because no information is given about the amount of precipitation in the normal years, the correct answer is "cannot tell." If they simply note that Southern California is classified as above normal, and Washington State as below normal, and (mis)interpret it in terms of amount, then they are erroneously led to choose Southern California as having a greater amount of precipitation. Agricultural and environmental policy makers are frequently called upon to make decisions in the face of incomplete information, and recognizing whether available information is sufficient to answer a question is an important skill for this population. Also, agricultural and environmental decisions often require comparing conditions at different places, as our participants were asked to do in this question.

Question 2

The second pair of maps compares a forecast of precipitation for October-November-December 2002 (Figure 3A), and the observed precipitation during the same time period (Figure 3B). The forecast map (Figure 3A) was made in September 2002 and shows probability in the same way as the map in Figure 2. The observed precipitation map (Figure 3B) shows *precipitation anomaly*, which is defined as the percentage of the long-term median precipitation at that place for those 3 months. In other words, *100%* means that observed rainfall was the same as the median rainfall at that place for those 3 months. In Figure 3B, the colors and contours both show precipitation anomaly.

Based on these two maps, please answer the following questions:

(a) Identify a region that may have suffered drought conditions during October-November-December 2002.

(b) Identify a region that may have suffered flood damage during October-November-December 2002.

(c) Suppose that past data show that Nevada normally receives 10 inches of precipitation. Then, how would you answer the following question: "Was it wetter or drier in Nevada for this period, compared to the normal year?"

Please explain in a few sentences how you came to the answer (or the difficulty you had in trying to answer the question).

(d) Considering all of North America in Figures 3A and 3B, how would you characterize the correspondence between the forecast and the observation? Circe one number from 1 to 5 below:

1. They tend to be opposite.
2. They are unrelated to each other.
3. They agree only slightly.
4. They agree somewhat.
5. They agree quite closely.

Please explain in a few sentences how you came to the answer (or the difficulty you had in trying to answer the question).

(e) Imagine that you are the Secretary of Agriculture of the US and that this pair of forecast/observed maps is representative of the last 5 years of forecasts. Would you recommend that these forecasts be used to make decisions about what crops to plant? Circle one number from 1 to 5 below concerning your action:

1. Strong recommendation for using the forecasts.
2. Weak recommendation for using the forecasts.
3. No recommendation in either way.
4. Weak recommendation for NOT using the forecasts.
5. Strong recommendation for NOT using the forecasts.

Please explain in a few sentences how you came to the answer (or the difficulty you had in trying to answer the question).

The second set of questions asked participants to interpret a map of observed precipitation (Figure 3B), and then to compare the map of observed precipitation and a map of forecast precipitation (Figure 3A) for the same time period. Questions 2a and 2b required participants to translate between the climatologist's term, *precipitation anomaly*, into the more societally relevant terms, *drought conditions* and *flood damage*. Question 2c was intended to assess participants' attention to units. A participant who simply compared the numeral "10" in the question with the numeral "25" within Nevada on the map, would likely answer that Nevada was wetter than usual, whereas the yellow and orange colors blanketing Nevada in Figure 3B show that it was in fact drier than normal. The "10 inches" in the question is irrelevant to obtaining the correct answer. In Questions 2d and 2e, we begin to address participants' degree of "belief" or confidence in the forecast tendencies shown on the maps, as opposed to their "understanding" of the maps. Question 2d asked participants how well they thought the forecast agreed with the subsequent observation. The maps selected for this question covered a time period for which the forecasters, in retrospect, considered the quality of the forecast to be within the typical range. Question 2e asked whether participants would recommend the forecast be used in agricultural decision making for actions with potential societal and economic impact.

Question 3

The third pair of maps shows the *above normal* threshold (Figure 4A) and the *below normal* threshold (Figure 4B). The *above normal* threshold is the border between the *normal* category and the *above normal* category; the *below normal* threshold is the border between the *normal* category and the *below normal* category. These thresholds are used in making the forecast maps such as Figure 2. These two maps (Figures 4A and 4B) cover the same 3-month period as for the map in Figure 2.

Suppose that you are given these two maps (Figures 4A and 4B), together with the map in Figure 2. Please complete the following statements about the probabilities of *above normal* and *below normal* precipitation at the two locations for this forecast period:

(a) (i) *Above normal:* The probability is _____% that Charleston, South Carolina (denoted by C in Figures 4A and 4B), will receive *more* than _____ mm of precipitation.

(ii) *Below normal:* The probability is _____% that Charleston, South Carolina, will receive *less* than _____ mm of precipitation.

Please explain in a few sentences how you came to the answer (or the difficulty you had in trying to answer the question).

(b) (i) *Above normal:* The probability is _____% that Phoenix, Arizona (denoted by P in Figures 4A and 4B), will receive *more* than _____ mm of precipitation.

(ii) *Below normal:* The probability is _____% that Phoenix, Arizona, will receive *less* than _____ mm of precipitation.

Please explain in a few sentences how you came to the answer (or the difficulty you had in trying to answer the question).

Question 3 was designed to assess participants' understanding of the way in which the maps convey the probability of different forecast outcomes. The climate prediction models generate a full probability distribution. To communicate this information to a nonspecialist audience, the map makers have simplified the model output into three bins, representing the probability of precipitation characterized as below-normal, normal, or above-normal levels (Figure 2), and provided separate maps showing threshold values that bound each of these terciles (Figures 4A and 4B). To answer Question 3 correctly, participants first need to use the map in Figure 2 to determine what is the predicted probability that the city in question will receive above-normal or below-normal precipitation during the period in question. Because both cities in question are categorized as above normal, the above-normal probability for each city can be determined directly from the map color (in the green range). For the below-normal probability, participants need to consult the superimposed bar graph. Then, participants need to refer to the maps in Figures 4A and 4B to convert the categories of below normal and above normal into millimeters of precipitation.

Procedure

The test questions were administered during a laboratory session in the course "Climate and Water," which is a required course during the first semester of the MPA program in Environmental Science and Policy. At this point in the semester, general concepts in hydrology and climate variability had been taught, but the techniques of climate modeling had not been specifically covered. Students understood that participation was voluntary and that they would not be graded on their performance (after given this explanation, all the students in class agreed to participate). Following a brief introduction by the professor of the class (P.L.), instructions and clarifications were given by investigators (T.I. and K.K.)

who had no instructional authority over the students. The makers of the forecasts and the maps (A.B. and C.R) were not present, to minimize the tendency that participants would produce map-favorable answers out of politeness.

Although IRI forecast maps are disseminated to most users online, participants in this study viewed the maps as color photocopies on 8.5x11 in. paper, one map per page. The photocopies faithfully reproduced the details and color discrimination of the on-screen maps. In Questions 2 and 3, which required combining information from multiple maps, the paper maps allowed the participants to look at the two maps side by side, an option that is not readily available to users of the online maps. We do not have data on what percentage of actual users of IRI forecast maps view them on screen versus as color hardcopy. Some might even use black-and-white printouts, clearly an undesirable option, but one that might be used in countries where computer hardware is scarce (for such a case, IRI provides text discussions that may reveal what the colors would be). Descriptions of the forecast maps provided for online users are basically the same as those given to our participants, with an additional description of prediction methods.

Participants did not talk to each other as they answered the questions, nor had they ever seen the maps prior to the test. Participants appeared to take the task quite seriously. No strict time limit was imposed, but most participants completed all the questions within 30 minutes. After all the participants had handed in their answers, the professor of the class led a discussion of the questions and answers.

Results

Question 1

Of the 47 participants, 24 participants (51%) correctly wrote some variation of “cannot tell,” but 23 participants (49%) wrongly answered “Southern California.” No one chose Washington State. There was not a significant difference in the percentage of correct responses between men (50%) and women (52%) (with an alpha level of .05, which is the same for all the statistical tests reported below).

When asked how they came to the answer, participants gave fairly detailed explanations. Examples of descriptions by participants who correctly said "cannot tell" are "I wouldn't know, since this is a probability map, without actual precipitation values," "We can't answer that without knowing what their precipitation usually is," and "Normal is a term relative to the climate of each region." Examples of wrong descriptions are "The color of Southern California shows the precipitation is above normal. The color of Washington State shows the precipitation is below normal," and "Southern California region was indicated by green, while Washington State's region was indicated as yellow. According to the legend, green is more 'wet' than blue [sic]." These wrong descriptions show that the probability of precipitation falling within the above-normal or below-normal categories is confused with the amount of precipitation.

Question 2

The majority of participants answered the first three questions correctly (94% for Questions 2a and 2b; 87% for Question 2c), indicating that they generally did not have difficulty reading from the observation map in Figure 3B whether a specific region received more or less precipitation than normal. For Questions 2a and 2b, there was not a significant difference in the percentage of correct responses between men (100%) and women (89%); for Question 2c, the difference was significant between men (100%) and women (78%), $z = 2.76$, $p < .01$. The three women who made a mistake on Questions 2a and 2b identified "wet" regions for the drought conditions and "dry" regions for the flood conditions. They seem to have misunderstood, for some reason, what precipitation anomaly means. The six women who made a mistake on Question 2c (one of whom answered Questions 2a and 2b wrongly too), either (a) directly compared the numeral 10 in the question with 25 shown on Nevada in Figure 3B, (b) did not figure out that the answer could be obtained from the precipitation anomaly map, or (c) failed to correctly locate Nevada on the map, as revealed by their written descriptions.

Questions 2d and 2e examined how participants evaluated the degree of agreement between the forecast and the subsequent observation (Figure 5A), and their willingness to use the forecast as a basis for taking some action about agriculture (Figure 5B). Most participants (39 out of 47) evaluated the degree of agreement between the forecast and observation as either *agree only slightly* or *agree somewhat*. Evaluation and action scores were significantly correlated ($r = -.42$, $p < .01$), indicating that participants perceiving stronger agreement between the forecast and observation are more likely to recommend the forecast be used in agricultural decision making. However, it is noteworthy that the variability in the course of action that the participants would take based on the forecast is greater than the variability in their evaluation of the quality of the forecast. Although the forecasters consider the forecast in Figure 3A to be within the typical range and the participants agreed that there was some degree of agreement between the forecast and observation, 43% (20 out of 47) of the participants would recommend (either strongly or weakly) *against* the use of such forecast information in making decisions about what crops to plant. There was not a significant difference between men and women in the distributions of the evaluation and action scores, in either the Mann-Whitney test or the Kolmogorov-Smirnov test.

Outside of the test session, we also asked 10 climate prediction and diagnostics scientists at IRI, the organization which makes and disseminates the forecasts, to evaluate the degree of agreement between the same pair of forecast and observation maps, on the same 5-point scale used in Question 2d (Figure 5C). Due to small sample size, the difference was only marginally significant (the Mann-Whitney test, $p < .10$), but these scientists rated the degree of agreement as higher (mean = 3.6) than did the participants (mean = 3.2). The forecaster (A.B.) chose *agree somewhat*. There might be a bias toward the better in their evaluation as these scientists are in some way or other involved in the forecasting. On the other hand, they know from accumulated experience how good a forecast usually can be for a specific season and region. The spatial correlation between the forecast and observation maps turns out to be .15 (footnote 1), and this is very close to the average of a precipitation forecast for North America for the Oct-Nov-Dec season. A possible explanation of the higher evaluation by the scientists is that they are so used to such a weak relationship that they consider the forecast and observation to be in agreement *somewhat* rather than *only slightly*.

Question 3

Most of the participants (about 90%) correctly gave the above-normal and below-normal thresholds in millimeters; that is, they did not have difficulty reading the maps which showed the thresholds separating the above-normal, normal, and below-normal categories (Figures 4A and 4B). In contrast, participants had trouble answering the probabilities of each city falling into the above-normal and below-normal categories, especially the latter (only about 20% answered correctly; see Figure 6A). For both Questions 3a and 3b, there was a significant difference in the percentage of correct responses among the four items: above-probability, above-mm, below-probability, and below-mm (the Cochran test, $p < .001$). Because above normal is the category having the highest probability for both cities during the period in question, the above-normal probability can be read directly from the colors blanketing the two cities in Figure 2 (light to dark green). The below-normal probability, on the other hand, needs to be extracted from the horizontal bar graphs. The results show that participants had trouble with the horizontal bar graphs (or some students might possibly have ignored the bar graphs), and perhaps uncertainty inherent in climate forecasts. There was not a significant difference between men and women in the percentage of correct responses for any item in Questions 3a and 3b.

In the participants' written descriptions, the most frequent misconception was the idea that the below-normal probability is 100 minus the above-normal probability. This shows that these participants ignored the existence of the normal category. Some participants said that because the cities in question are classified as above normal, there should be no number for the below-normal probability (or the below-normal probability should be 0). This suggests that these participants failed to grasp the probabilistic nature of the forecast and thought that there should be one prediction of either above normal, normal, or below normal. Some participants explicitly mentioned that the bar graphs were difficult to understand. Overall, participants indicated their trouble or confusion in answering these questions (especially the below-probability items).

To further examine the validity of the above discussion, we conducted a second study 8 months later, asking a subset of the participants to answer the same questions using cities for which below normal was the category having the highest probability (Seattle, Washington; and Detroit, Michigan). In the same way as in Question 3, participants were given a forecast map (Figure 2) and threshold maps (Figures 4A and 4B, in which the locations of Seattle and Detroit were marked), and asked questions about the above-normal and below-normal precipitation probability and amount. This time, the below-normal probability can be read directly from the colors blanketing the two cities; whereas the above-normal probability needs to be extracted from the horizontal bar graphs. Out of the 47 students who took the original test, 23 students (8 men and 15 women) took the second test. We did not find any particular difference between the original and second groups of students in terms of, for example, class standing. Overall, the performance was better on the second test compared to the first test. Most of the participants (more than 90%) correctly answered the above-mm and below-mm items; but they did not do as well on the probability items, especially the above-probability items (only about 60% correct; see Figure 6B). Note that the pattern of performance on the above-probability and below-probability items was reversed in the original study (which used above-normal cities) and the second study (which used below-normal cities). These results provide further support for the interpretation that the participants had difficulty understanding what the bar graphs depicted. For both cities, the difference in performance among the four items (above-probability, above-mm, below-probability, below-mm) was significant (the Cochran test, $p < .001$). For the above-probability items, there was a significant difference between men and women in the percentage of correct responses, favoring men over women (for Seattle, 88% vs. 40%, $z = 2.76$, $p < .01$; for Detroit, 88% vs. 47%, $z = 2.35$, $p < .05$).

Discussion and Conclusions

The results of this study show that the efficacy of the climate forecast maps, which have been in actual use by a broad spectrum of decision makers, as a communication tool was less than one would hope for, at least in the current design. Our participants, qualified and motivated students in a professional master's degree program aimed for prospective policy makers, failed to interpret some of the forecast maps as the map maker intended.

The participants understood observed precipitation maps easily, whereas they had more trouble understanding probability forecast maps. The majority of participants (about 90%) answered correctly on the items that can be answered from the observation maps (Questions 2a, 2b, 2c and the millimeter items of Question 3). On items that require using the probabilistic forecast maps (Question 1 and the probability items of Question 3), the percentage of correct responses was lower, ranging from 23% to 74%. The reversed pattern of participants' performance on the above-probability and below-probability items in Question 3 and in the follow-up study provides further evidence of the difficulty with the probability maps, especially the superimposed small bar graphs.

The participants also had difficulty referring to and integrating different aspects of the climate forecasts, which was required in Questions 1 and 3. In Question 1, participants need to differentiate between probabilistic forecasts and precipitation amount. In Question 3, participants need to clearly understand probabilistic three-category forecasts, thresholds in millimeters, and distributions of forecast outcomes over the three categories, by interrelating three different maps.

To help the user understand the forecast maps, IRI has created online tutorials (http://iri.columbia.edu/climate/forecast/tutorial, http://iri.columbia.edu/climate/forecast/tutorial2), and runs an extensive training program aimed at mid-career professionals from developing countries, who are potential users of climate forecasts (http://www.ccnmtl.columbia.edu/projects/climate/course_html). However, it seems inevitable that many users will neither read tutorials nor attend short courses or workshops, so there should be improvements in the design of the maps as well. Since the maps are disseminated online, one powerful option may be to present a simpler map to the user initially, and layer additional information behind the map that can be accessed interactively. For example, the initially presented forecast maps could use color to show the most likely categorical level of precipitation, with the detailed probability distribution accessed by a click on a point of interest on the map, rather than shown as a superimposed bar graph. In fact, IRI has recently developed a forecast product presented in such a format, and the National Weather Service's Climate Prediction Center also issues a similar forecast product (http://www.cpc.ncep.noaa.gov/products/forecasts). Another possible idea for improvement in design concerns the legend of the forecast maps (Figures 2 and 3A). The below-normal probabilities might be better shown in reverse order, so that the entire legend depicts less to more precipitation (i.e., drier to wetter conditions compared to normal for a specific region) from left to right.

On the other hand, it should be noted that although the forecast maps, in the current design, are so complex and pose some difficulty for the user, they are still a simplified representation of the map maker's conception of the represented space. The forecasters who constructed the maps have a vast amount of knowledge about the climate system and experience in climate forecasting. However, they found it impossible to cram all of their insights into the map, and have chosen a subset of the information that they think important and necessary to the user. In other words, specialists' understanding of complex earth systems exceeds the capacity of available communication tools to convey that understanding to professional policy makers and other nonspecialists. How to simplify representations of complex scientific data, without sacrificing richness and subtlety of the data and overwhelming the recipient, is an important issue to be further investigated (e.g., whether to show the detailed distribution of forecast outcomes or classify them into three categories).

As for the evaluation of the quality of the forecasts, more than half of the participants were not inclined to rely on these types of climate forecasts for making agricultural decisions. In particular, we found in the process of our tests that our participants felt uncomfortable with the existence of uncertainty in the forecasts. In contrast, the forecasters consider it imperative to provide information about uncertainty inherent in the forecasts. Thus, there may exist a deep-rooted belief that uncertainty in natural sciences should be more constrained than that related to social and affective issues for public decision making. Some of the observations we reported in this study point to the potential impact of limited understanding of some fundamental aspects of science on decision-making processes. The lack of familiarity with the real context of climate forecasts may thus induce erroneous interpretation, or "distrust" in their applicability to public decisions. These observations are significant since actual users of such maps may have less training than was provided for the MPA students in this study. At the same time, we note that policy makers would probably base their decisions on more than one forecast product in practice (e.g., forecast products from different organizations, or forecasts for multiple seasons by the same organization). How decision makers' attitudes toward forecasts change with exposure to several sets of forecast products may be an interesting question.

To summarize, the climate forecast maps, which have been disseminated and in actual use, fell short, in the current design, of realizing the potential benefits of the forecasts. Improvement in design, education of the user, or preferably both, are necessary, so that the forecast maps would be understood correctly and the potentially beneficial forecasts would be considered persuasive enough as input for decision making. The findings from this study about the difficulty with the currently used maps should provide insights into a future study to that end.

Acknowledgments

This research was partially supported by a seed grant from the International Research Institute for Climate Prediction. We are grateful to David Krantz for his valuable comments on an earlier draft of this paper, and to James Hansen for providing us with some of the references. We acknowledge the insightful comments by three anonymous reviewers, which helped us refine the paper. We also thank the prospective policy makers who participated in this study. Their participation made it possible to empirically examine the communication issue about the currently disseminated IRI climate forecast maps. Part of the results of this study were presented at the 115th annual meeting of the Geological Society of America in Seattle, Washington, and at the NOAA Climate Prediction Applications Science Workshop in Tallahassee, Florida.

References

Barnston, A. G., S. J. Mason, L. Goddard, D. G. DeWitt, and S. E. Zebiak. 2003. Multimodel ensembling in seasonal climate forecasting at IRI. *Bulletin of the American Meteorological Society* 84: 1783-1796.

Brewer, C. A., A. M. MacEachren, L. W. Pickle, and D. Herrmann. 1997. Mapping mortality: Evaluating color schemes for choropleth maps. *Annals of the Association of American Geographers* 87: 411-438.

Bunch, R., and R. Lloyd. 2000. The search for boundaries on maps: Color processing and map pattern effects. *Cartography and Geographic Information Science* 27: 15-29.

Cane, M. A., S. E. Zabiak, and S. C. Dolan. 1986. Experimental forecasts of El Niño. *Nature* 321: 827-832.

Changnon, S. A. 1992. Contents of climate predictions desired by agricultural decision makers. *Journal of Applied Meteorology* 31: 1488-1491.

Changnon, S. A., J. M. Changnon, and D. Changnon. 1995. Uses and applications of climate forecasts for power utilities. *Bulletin of the American Meteorological Society* 76: 711-720.

Dent, B. D. 1972. Visual organization and thematic map communication. *Annals of the Association of American Geographers* 62: 79-93.

Eastman, J. R. 1985. Cognitive models and cartographic design research. *Cartographic Journal* 22: 95-101.

Eddy, D. M. 1982. Probabilistic reasoning in clinical medicine: Problems and opportunities. In: D. Kahneman, P. Slovic, and A. Tversky (eds), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press. pp. 249-267.

Ekman, G., R. Lindman, and W. William-Olsson. 1961. A psychophysical study of cartographic symbols. *Perceptual and Motor Skills* 13: 355-368.

Flannery, J. J. 1971. The relative effectiveness of some common graduated point symbols in the presentation of quantitative point data. *Canadian Cartographer* 8: 96-109.

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

Gilhooly, K. J., M. Wood, P. R. Kinnear, and C. Green. 1988. Skill in map reading and memory for maps. *Quarterly Journal of Experimental Psychology* 40A: 87-107.

Goddard, L., A. G. Barnston, and S. J. Mason. 2003. Evaluation of the IRI's "net assessment" seasonal climate forecasts: 1997-2001. *Bulletin of the American Meteorological Society* 84: 1761-1781.

Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane. 2001. Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology* 21: 1111-1152.

Guelke, L. 1976. Cartographic communication and geographic understanding. *Canadian Cartographer* 13: 107-122.

Hammer, G. L., J. W. Hansen, J. G. Phillips, J. W. Mjelde, H. Hill, A. Love, and A. Potgieter. 2001. Advances in application of climate prediction in agriculture. *Agricultural Systems* 70: 515–553.

Hansen, J. W. 2002. Realizing the potential benefits of climate prediction to agriculture: Issues, approaches, challenges. *Agricultural Systems* 74: 309–330.

Hoffrage, U., S. Lindsey, R. Hertwig, and G. Gigerenzer. 2000. Communicating statistical information. *Science* 290: 2261-2262.

Jones, J. W., J. W. Hansen, F. S. Royce, and C. D. Messina. 2000. Potential benefits of climate forecasting to agriculture. *Agriculture, Ecosystems and Environment* 82: 169–184.

Kahneman, D., P. Slovic, and A. Tversky (eds). 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Koláčný, A. 1969. Cartographic information: A fundamental concept and term in modern cartography. *Cartographic Journal* 6: 47-49.

Lloyd, R., and R. L. Bunch. 2003. Technology and map-learning: Users, methods, and symbols. *Annals of the Association of American Geographers* 93: 828-850.

Lloyd, R., and T. Steinke. 1977. Visual and statistical comparison of choropleth maps. *Annals of the Association of American Geographers* 67: 429-436.

MacEachren, A. M. 1995. *How maps work: Representation, visualization, and design*. New York: Guilford Press.

Marshall, C. H., R. A. Pielke, Sr., and L. T. Steyaert. 2003. Crop freezes and land-use change in Florida. *Nature* 426: 29-30.

Monmonier, M. 2000. *Air apparent: How meteorologists learned to map, predict, and dramatize weather*. Chicago: University of Chicago Press.

Neisser, U. 1976. *Cognition and reality*. New York: Freeman.

Nicholls, N. 1999. Cognitive illusions, heuristics, and climate prediction. *Bulletin of the American Meteorological Society* 80: 1385-1397.

Olson, J. M. 1975. Experience and the improvement of cartographic communication. *Cartographic Journal* 12: 94-108.

Patton, D. K., and R. G. Cammack. 1996. An examination of the effects of task type and map complexity on sequenced and static choropleth maps. In: C. H. Wood and C. P. Keller (eds), *Cartographic design: Theoretical and practical perspectives*. Chichester, England: Wiley. pp. 237-252.

Pfaff, A., K. Broad, and M. Glantz. 1999. Who benefits from climate forecasts? *Nature* 397: 645-646.

Phillips, R. J., A. De Lucia, and N. Skelton. 1975. Some objective tests of the legibility of relief maps. *Cartographic Journal* 12: 39-46.

Robinson, A. H., and B. B. Petchenik. 1975. The map as a communication system. *Cartographic Journal* 12: 7-15.

Sholl, M. J., and H. E. Egeth. 1982. Cognitive correlates of map-reading ability. *Intelligence* 6: 215-230.

Slocum, T. A., D. C. Cliburn, J. J. Feddema, and J. R. Miller. 2003. Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance. *Cartography and Geographic Information Science* 30: 299-317.

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124-1131.

Tversky, A., and D. Kahneman. 1982. Evidential impact of base rates. In: D. Kahneman, P. Slovic, and A. Tversky (eds), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press. pp. 153-160.

Underwood, J. D. M. 1981. Skilled map interpretation and visual-spatial ability. *Journal of Geography* 80: 55-58.

Williamson, J., and C. McGuinness. 1990. The role of schemata in the comprehension of maps. In: K. J. Gilhooly, M. T. G. Keane, R. H. Logie, and G. Erdos (eds), *Lines of thinking: Reflections on the psychology of thought* (vol. 2). Chichester, England: Wiley. pp. 29-40.

Footnote

[1]To calculate a spatial correlation, the three-category forecast was converted into a single number by the formula (*below_normal_prob* x 1) + (*normal_prob* x 2) + (*above_normal_prob* x 3). For example, a forecast of 25%-35%-40% for the below-normal, normal, and above-normal categories would be 2.15. The observations would be 1, 2, or 3, depending on which category was observed. Then these two sets of numbers were correlated. This method was applied to obtain one possible objective measure of (visual) correspondence between the forecast and observation maps.

Figure Captions

Figure 1. (*A*) In cartographic communication, information flows from the represented space, to the map maker, to the map, and to the map reader (after Robinson and Petchenik 1975). This communication flow is affected by attributes of both the map maker and the map reader. (*B*) The policy maker, being the intended map reader, uses information from the map to formulate a policy, which when implemented, brings about a change in the represented space.

Figure 2. Map used in Question 1, showing by color the predicted probability of a given locality receiving below, near, or above its historical normal precipitation for Feb-Mar-Apr 2003. Small bar graphs indicate the probabilities of precipitation falling into the three categories.

Figure 3. Maps used in Question 2. (*A*) Forecast map, showing a probabilistic precipitation forecast for Oct-Nov-Dec 2002. (*B*) Observation map, showing the ratio of the observed precipitation during Oct-Nov-Dec 2002 to the median precipitation over the past 30 years.

Figure 4. Maps used in Question 3. (*A*) Above-normal threshold map, showing the threshold value of precipitation in mm separating the above-normal and normal categories. (*B*) Below-normal threshold map, showing the threshold value separating the normal and below-normal categories. These threshold values were used to make the forecast map in Figure 2.

Figure 5. (*A*) Distribution of evaluation scores (1 = *opposite*; 2 = *unrelated*; 3 = *agree only slightly*; 4 = *agree somewhat*; 5 = *agree quite closely*). (*B*) Distribution of action scores (1 = *strong recom.*; 2 = *weak recom.*; 3 = *no recom.*; 4 = *weak recom. against*; 5 = *strong recom. against*). (*C*) Distribution of 10 scientists' evaluation of the degree of agreement between the forecast and observation, on the same 5-point scale as in panel A.

Figure 6. (*A*) Percentage of correct responses to the threshold questions for Charleston and Phoenix. The labels "Prob." and "mm" on the horizontal axis indicate the probability

items and the millimeter items, respectively. (*B*) Percentage of correct responses to the threshold questions for Seattle and Detroit.

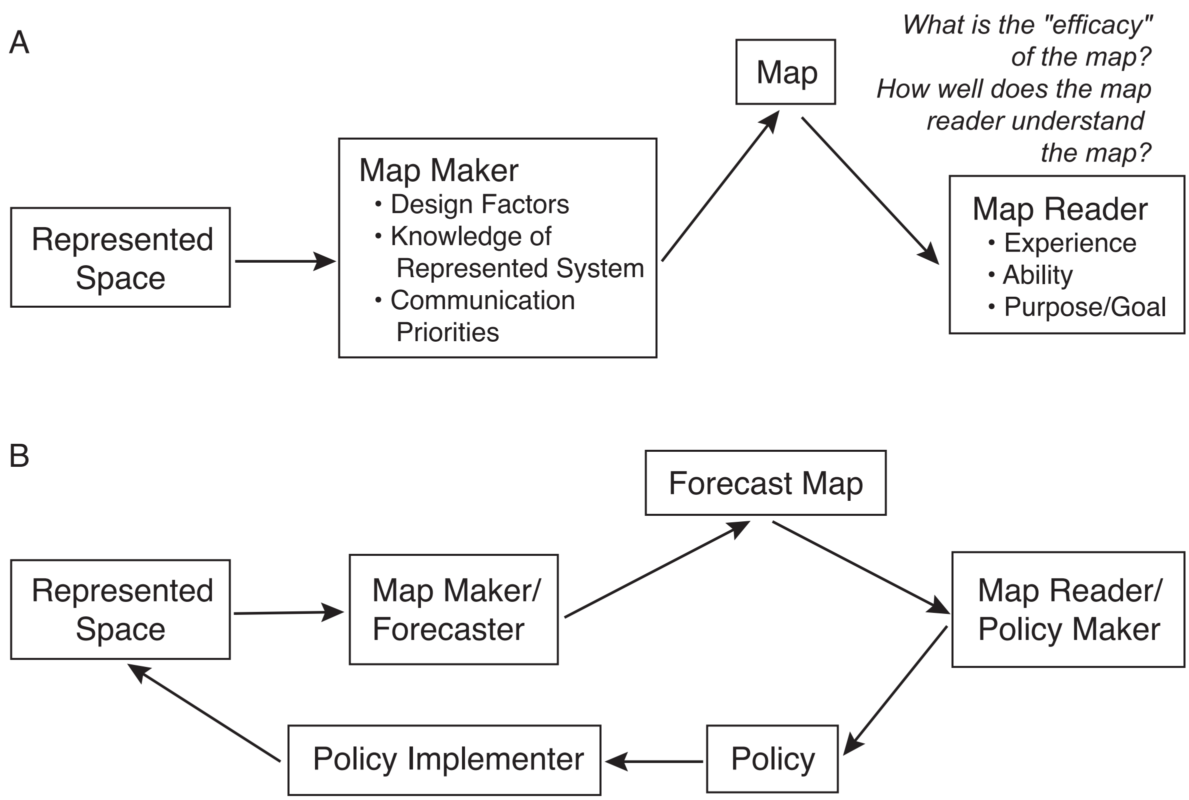
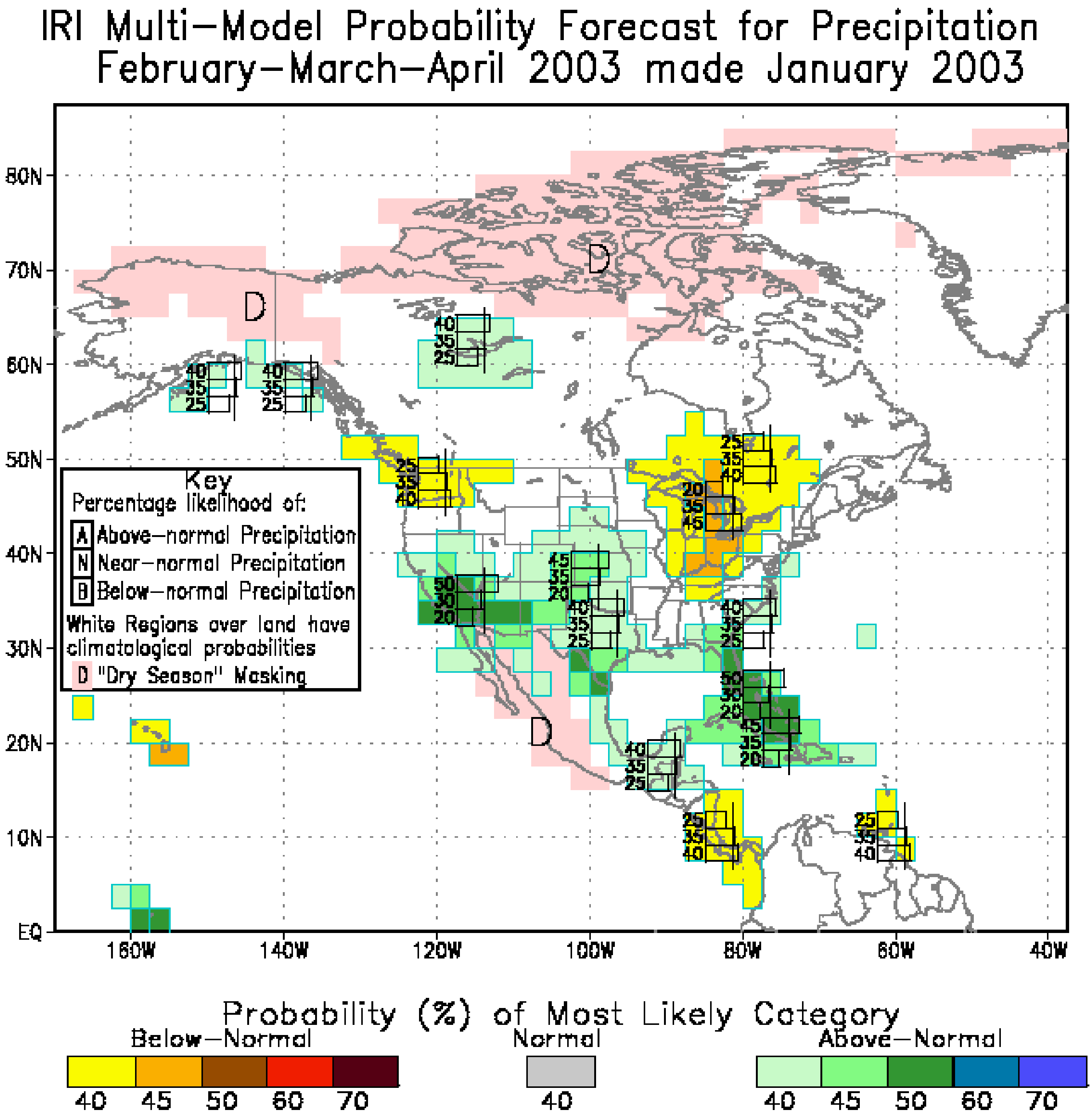
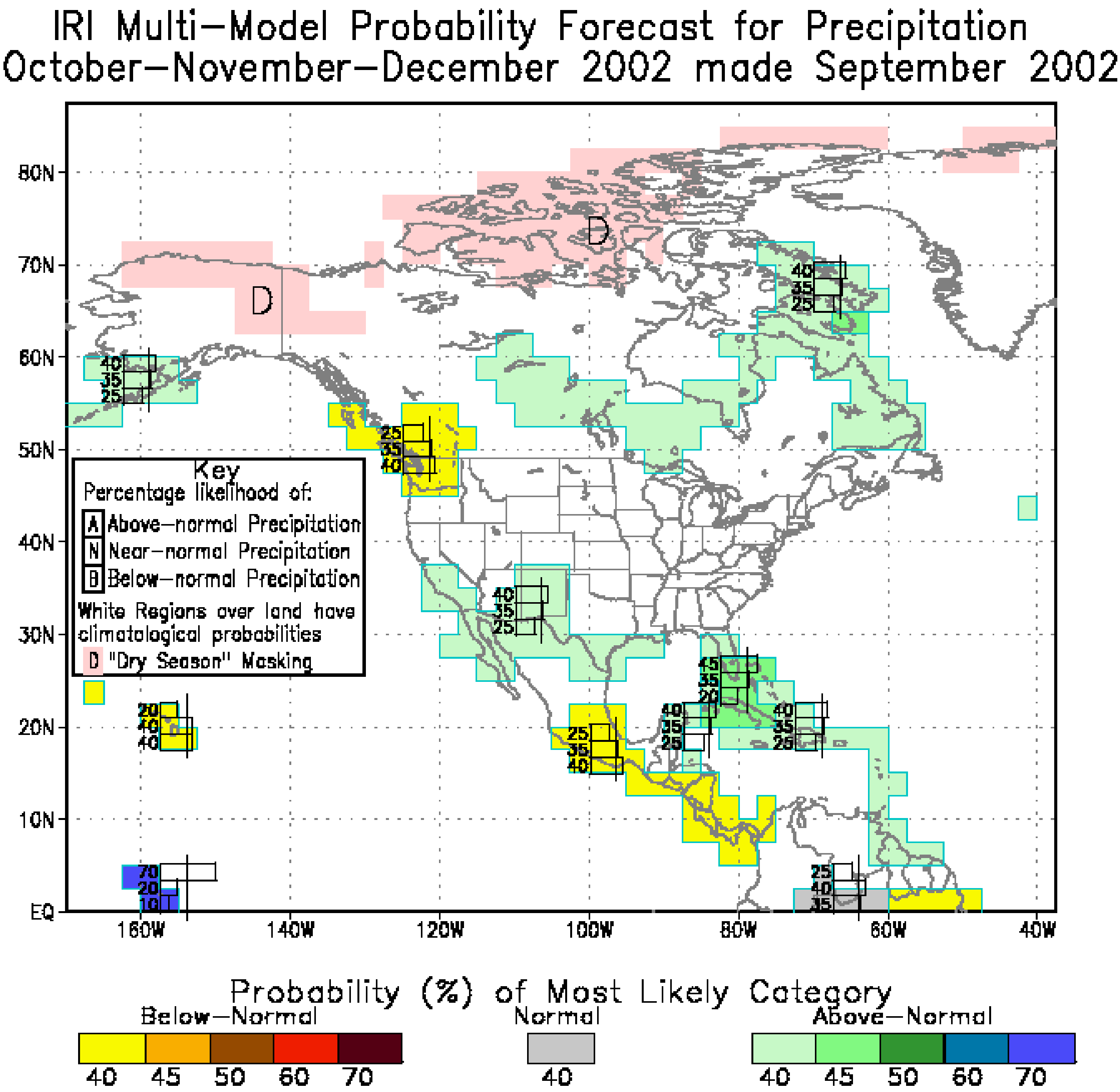

A
Represented Space
Map Maker
• Design Factors
• Knowledge of Represented System
• Communication Priorities
Map
What is the "efficacy" of the map? How well does the map reader understand the map?
Map Reader
• Experience
• Ability
• Purpose/Goal
B
Represented Space
Map Maker/ Forecaster
Forecast Map
Map Reader/ Policy Maker
Policy
Policy Implementer

Figure 2


IRI Multi-Model Probability Forecast for Precipitation
February-March-April 2003 made January 2003
Key
Percentage likelihood of:
A Above-normal Precipitation
N Near-normal Precipitation
B Below-normal Precipitation
White Regions over land have climatological probabilities
D "Dry Season" Masking
Probability (%) of Most Likely Category
Below-Normal
Normal
Above-Normal
40 45 50 60 70
40
40 45 50 60 70



IRI Multi-Model Probability Forecast for Precipitation
October-November-December 2002 made September 2002
Key
Percentage likelihood of:
A Above-normal Precipitation
N Near-normal Precipitation
B Below-normal Precipitation
White Regions over land have climatological probabilities
D "Dry Season" Masking
80N
70N
60N
50N
40N
30N
20N
10N
EQ
160W
140W
120W
100W
80W
60W
40W
Probability (%) of Most Likely Category
Below-Normal
40 45 50 60 70
Normal
40
Above-Normal
40 45 50 60 70

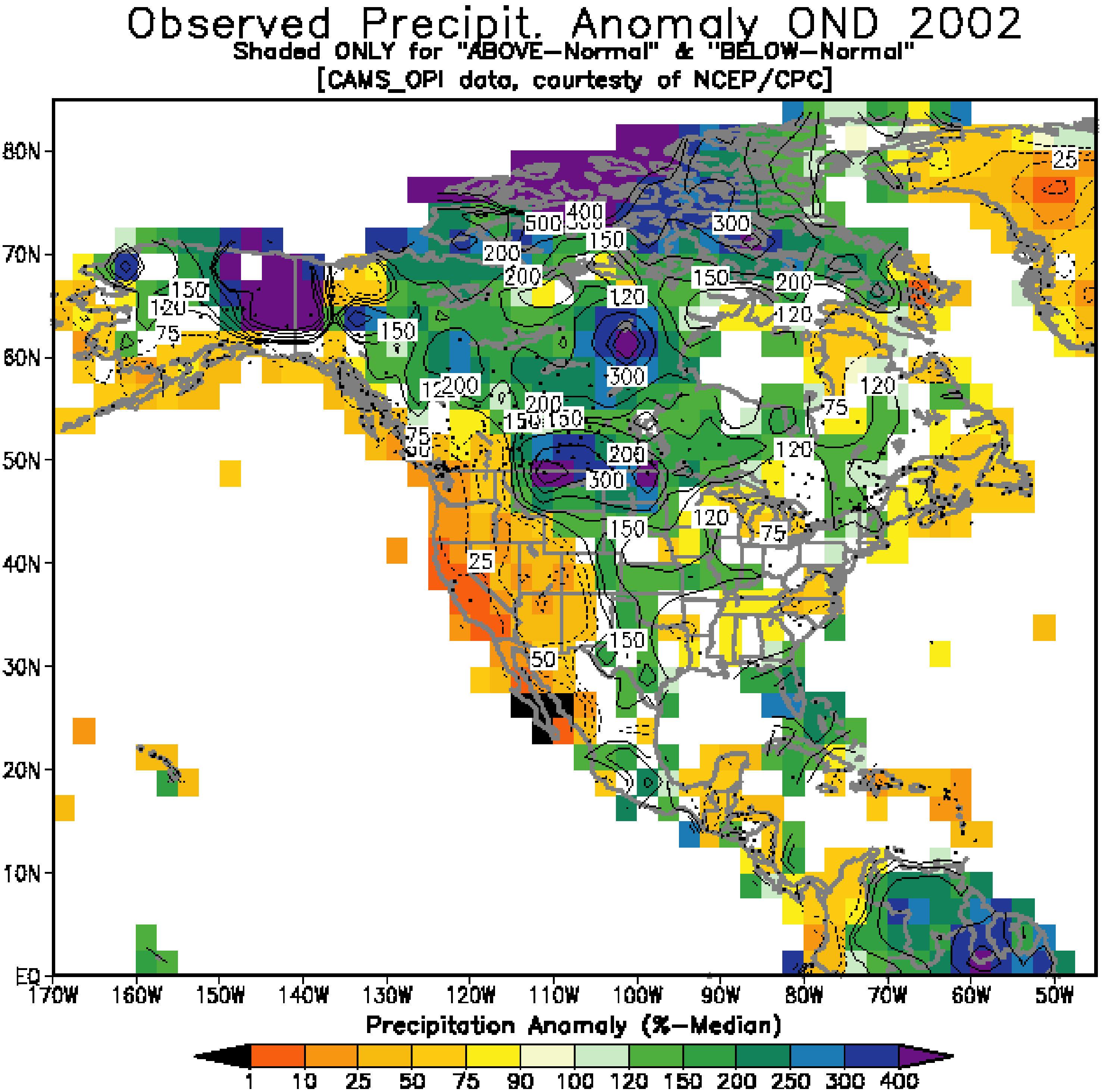
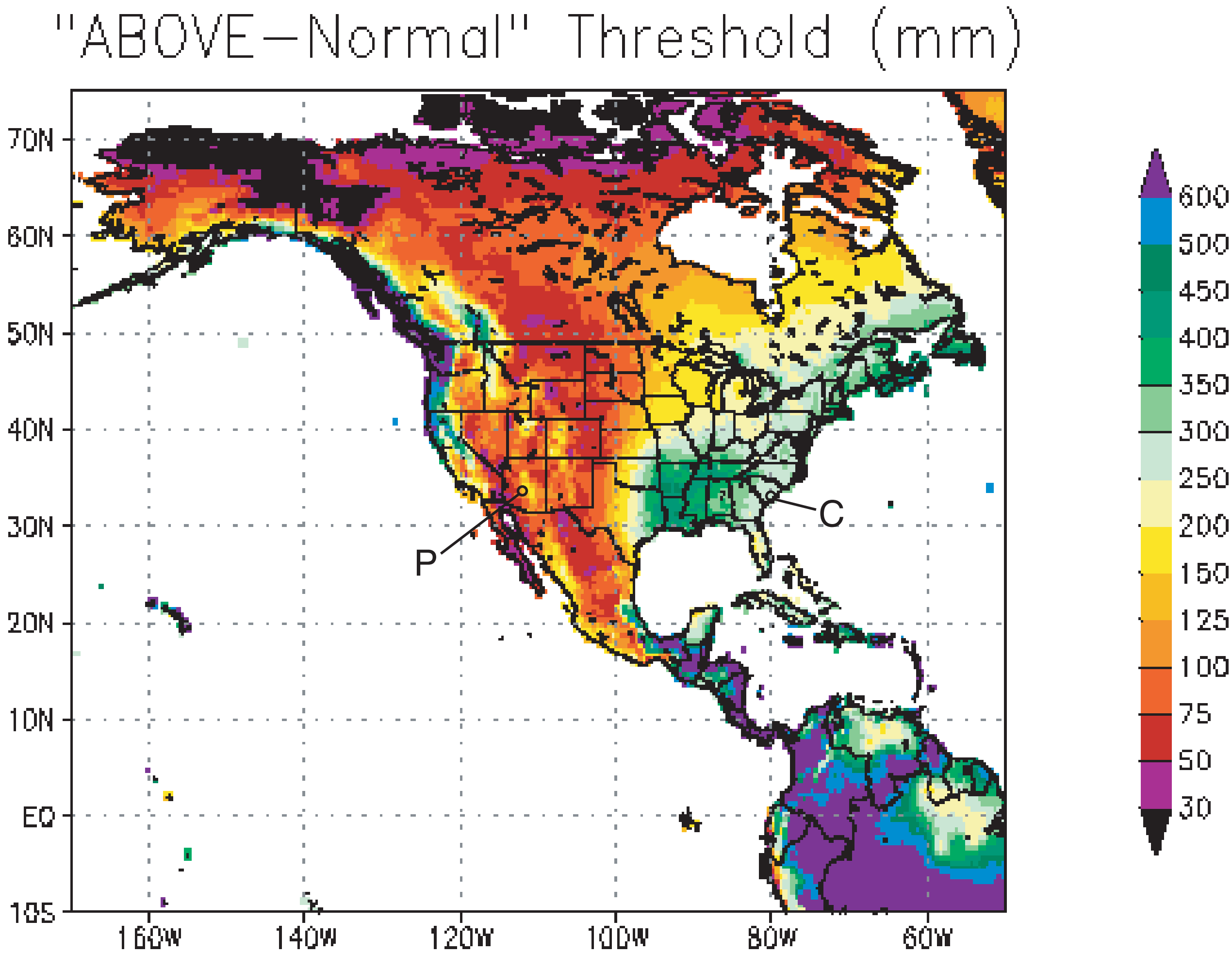

Observed Precipit. Anomaly OND 2002
Shaded ONLY for "ABOVE-Normal" & "BELOW-Normal"
[CAMS_OPI data, courtesty of NCEP/CPC]
80N
70N
60N
50N
40N
30N
20N
10N
EQ
170W
160W
150W
140W
130W
120W
110W
100W
90W
80W
70W
60W
50W
Precipitation Anomaly (%-Median)
1
10
25
50
75
90
100
120
150
200
250
300
400

Figure 4A


"ABOVE-Normal" Threshold (mm)
70N
60N
50N
40N
30N
20N
10N
EQ
10S
160W
140W
120W
100W
80W
60W
P
C
600
500
450
400
350
300
250
200
150
125
100
75
50
30

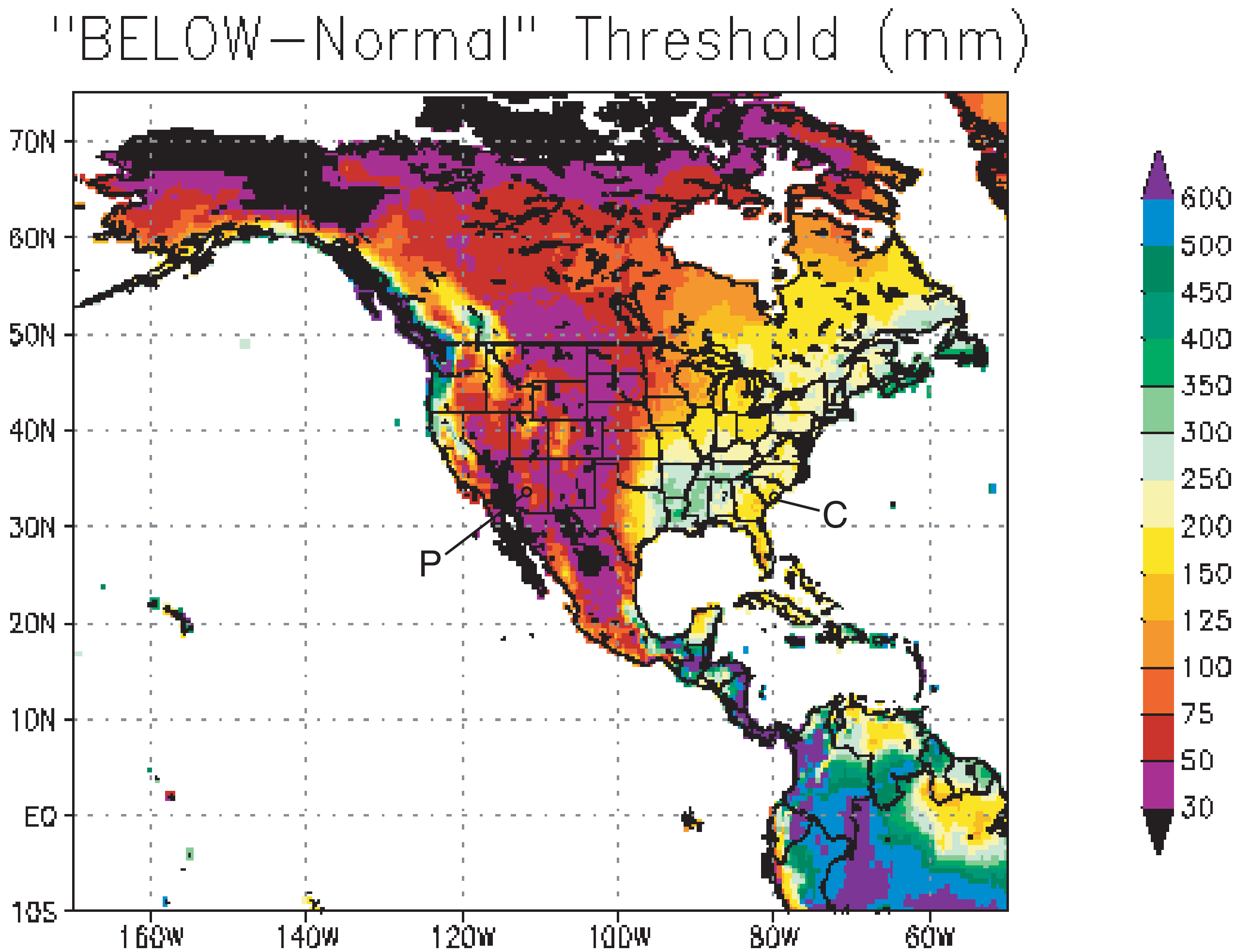

"BELOW-Normal" Threshold (mm)
70N
60N
50N
40N
30N
20N
10N
EQ
10S
160W
140W
120W
100W
80W
60W
P
C
600
500
450
400
350
300
250
200
150
125
100
75
50
30

Figure 5

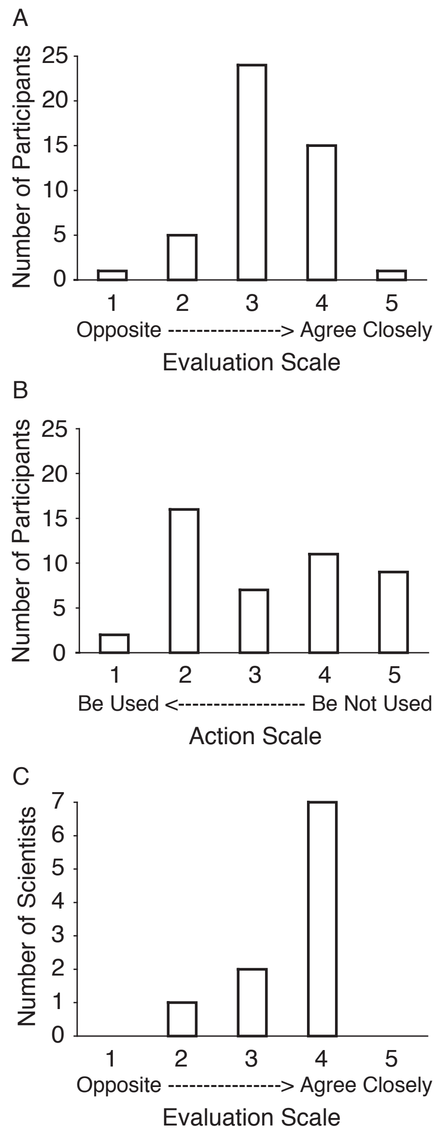
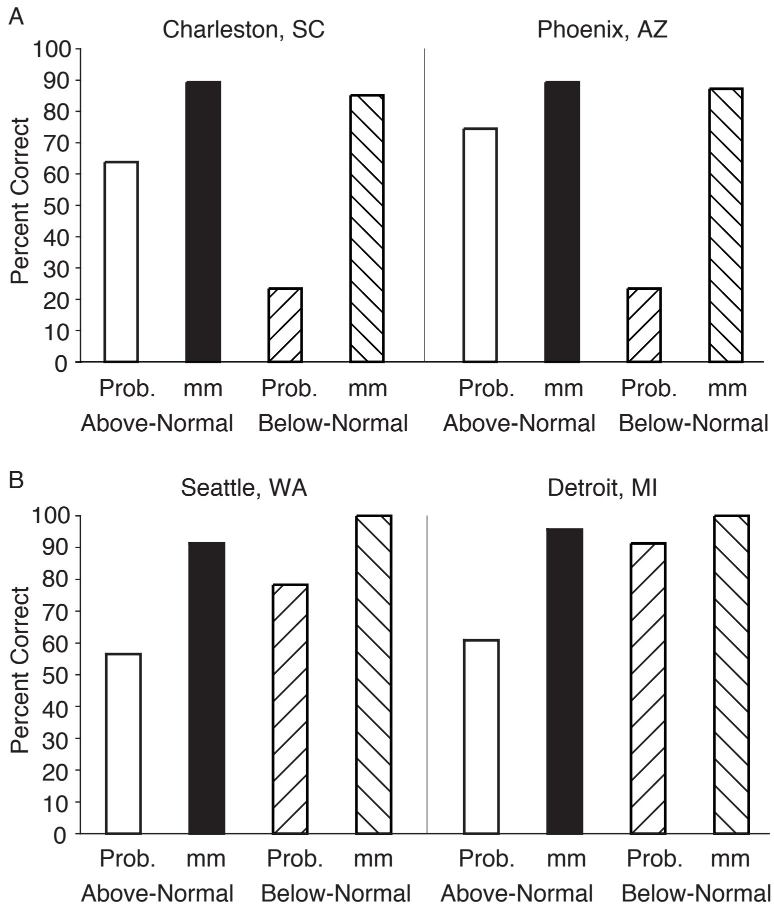

A
Number of Participants
25
20
15
10
5
0
1 2 3 4 5
Opposite ---------------> Agree Closely
Evaluation Scale
B
Number of Participants
25
20
15
10
5
0
1 2 3 4 5
Be Used <----------------- Be Not Used
Action Scale
C
Number of Scientists
7
6
5
4
3
2
1
0
1 2 3 4 5
Opposite ---------------> Agree Closely
Evaluation Scale

Figure 6


A
Charleston, SC
Phoenix, AZ
Percent Correct
100
90
80
70
60
50
40
30
20
10
0
Prob.
mm
Prob.
mm
Above-Normal
Below-Normal
Prob.
mm
Prob.
mm
Above-Normal
Below-Normal
B
Seattle, WA
Detroit, MI
Percent Correct
Prob.
mm
Prob.
mm
Above-Normal
Below-Normal
Prob.
mm
Prob.
mm
Above-Normal
Below-Normal