CONFIDENCE BUILDERS

Evaluating Seasonal Climate Forecasts from User Perspectives

BY HOLLY C. HARTMANN, THOMAS C. PAGANO, S. SOROOSHIAN, AND R. BALES

Whether climate outlooks are "good" or "bad" depends on your perspective, including which regions, seasons, lead times, and aspects of forecast quality are relevant to your specific decision making situation.

For years seasonal climate forecasts had the potential to improve resource management but instead played only a marginal role in real-world decision making.[1] A widespread perception that the quality of the forecasts was poor presented an especially persistent dilemma for climate forecasters (Changnon 1990).

In 1997–98, however, use of the forecasts turned a corner. Water and emergency managers started paying serious attention to the forecasts during the 1997/98 El Niño event, because similar events in the past had produced exceptional flooding and drought (Office of Global Programs 1999; S. A. Changnon 2000; Pagano et al. 1999). As the subsequent strong La Niña produced two consecutive dry winters in the southern states, water managers continued to pay attention to the forecasts. Wildland fire managers and cattle ranchers also began expressing more interest in using the seasonal forecasts.

Given these experiences, one would expect resource managers to pay even more attention to seasonal climate forecasts as conflicts become more frequent over limited water supplies, evolving timber and grazing policies, and burgeoning energy demands. But skepticism about using seasonal forecasts remains.

For example, interactions with decision makers about their use of climate information during the 1997/98 El Niño event revealed well-entrenched resistance to use of seasonal forecasts (Pagano et al. 2002). One agency staff member told us: "In the many years that I've lived here [Arizona], I've learned that unless you're from the East Coast, you know you can't predict what's going to happen out here when it comes to mother nature." Another time, after making a presentation about climate variability and forecasts to a group of farmers and ranchers, discussion was becoming lively about how they and their competitors could possibly make use of the forecasts. But the meeting leader cut the discussion short, saying

[1] For a thorough discussion of the nature of seasonal forecasts and the distinction between these predictions of aggregate conditions and weather forecasts, which forecast specific events, see Hartmann et al. (1999). For more on the history of these forecasts and their use, see Namias (1968), Nicholls (1980), Changnon and Vonnhame (1986), Changnon (1992), Sonka et al. (1992), Pulwarty and Redmond (1997), Callahan et al. (1999), Pagano et al. (2002), and Pulwarty and Melis (2001).

AFFILIATIONS: HARTMANN, PAGANO, SOROOSHIAN, AND BALES—Department of Hydrology and Water Resources, The University of Arizona, Tucson, Arizona
CORRESPONDING AUTHOR: Holly C. Hartmann, Department of Hydrology and Water Resources, The University of Arizona, Tucson, AZ 85721
E-mail: hollyh@hwr.arizona.edu

In final form 19 February 2002

that climate forecast accuracy had a long way to go before the products would have any relevance to people actually making decisions.

Our conversations with ranchers, water managers, and wildland fire managers in recent years have sent us a clear warning. Uncertainty about the accuracy of the climate forecasts precludes users from making more effective use of them (Conley et al. 1999; Institute for the Study of Planet Earth 2000, hereafter ISPE 2000; Pagano et al. 2000, 2001, 2002).

Scientists can work to overcome that uncertainty. In this article we show how information can be conveyed better if we, as scientists, look at the forecasts from a user's perspective. The result is more useful forecasts.

The value of assessing forecast accuracy from the user's perspective is readily apparent. Thorough understanding of forecast performance helps decision makers determine when and how much to rely on the forecasts as well as how to respond to expected climate anomalies. Evaluations of seasonal forecasts also can help planners reduce their vulnerability to climate, because they can plan more informed, and thus more effective, preparation.

Even though assessments reveal the uncertainty of seasonal forecasting, uncertain forecasts can help proactive planning. Seasonal climate outlooks by the National Oceanic and Atmospheric Administration's (NOAA's) Climate Prediction Center (CPC) indeed have shown greater predictive skill than forecasts that project climatological probabilities (based on historical data)—but not in every case. Decision makers need to know when forecasts are not reliable enough for their purposes (Sarewitz et al. 2000). In fact, consistent communication of forecast uncertainty can increase forecast credibility (O'Grady and Shabman 1990). Without this credibility, the costly consequences of bad outcomes from the use of a single forecast can devastate user confidence in subsequent forecasts (Glantz 1982).

Forecast performance evaluations have periodically appeared in the scientific literature.[2] However, results generally do not apply directly to resource management decisions. The studies reflect perspectives of climate researchers and forecasters, not users (the few exceptions to this include Mjelde et al. 1993; Lehman 1987). Frequently updated evaluations, easily accessible to users, do accompany online climate forecasts by the CPC (www.cpc.ncep.noaa.gov/products/predictions/90day/). But they consist of skill scores computed for each lead time for the entire conterminous United States without regional breakdowns. Other skill assessments available online deal with individual forecasting techniques but cannot help users easily evaluate the official outlooks, which incorporate subjective judgments that vary from forecast to forecast.

PLANNERS' PERSPECTIVES. Basically, forecast users have two complementary questions about forecast accuracy. First, what is the probability that the climate forecasts will warn the user of climate extremes? Second, given a specific forecast predicting an increased likelihood of some event, what is the probability that the event will actually occur? Administrators of very large river systems or business managers who must keep track of competition in other regions may pose these questions for large regions. However, many forecast users are interested in evaluations specific to a relatively small base of operations. The questions also are framed with respect to a specific time frame or decision. Wildland fire managers, for instance, plan their summer season resource allocations early in the spring, so they are interested in forecasts issued in the January–March time frame. Evaluations that take into account forecast skill in other months may not reflect the concerns of these users.

To serve resource managers effectively, scientists need better-targeted evaluations of forecasts. In our research we have attempted to measure and communicate forecast performance in ways that are meaningful to potential users. Our framework for evaluating forecasts considers the climate conditions, seasons, and lead times relevant to decision makers. We address the diversity of users' meteorological savvy both with the varying levels of sophistication of the evaluations and with our method of presenting the evaluations: not all decision makers have ready access to a scientist trained to interpret forecasts and evaluations. The multiple levels of sophistication, with varying trade-offs between information and understanding, offer opportunities for users to develop deeper insights about climate forecasts and their credibility and the implications of these issues for decisions.

We found that the CPC seasonal climate outlooks clearly perform better for some users than for others. We focused on three groups of prospective climate forecast users in the southwest United States: water managers, wildland fire managers, and cattle ranchers. The different ways in which climate forecasts may

[2] For examples of forecast evaluations, see Nicholls (1980), Bettge et al. (1981), Priesendorfer and Mobley (1984), Barnett and Priesendorfer (1987), Lehman (1987), O'Lenic (1990), Livezey (1990), Murphy and Huang (1991), Mjelde et al. (1993), and Wilks (2000).

fail to be useful for one or another group are more subtle than simply the level of overall accuracy one normally sees in forecast evaluations. We are prompted to adapt what Tolstoy wrote in *Anna Karenina*: while perfect forecasts are all alike, every imperfect forecast is imperfect in its own way. Which aspect of forecast quality is most important to a user depends on what decisions they must make. Forecast quality is ultimately in the perception of the beholder, not just in the evaluation of the forecaster.

MULTIPLE SCORES FOR MULTIPLE USERS.

The CPC has produced climate forecasts in basically the current format since December 1994. While the complete forecast package consists of several elements (Hartmann et al. 1999, 2002), two examples are presented in Fig. 1: maps of predictions of probability anomalies of surface air temperature and precipitation. The contours indicate how likely it is that average air temperature or total precipitation during the forecast period will fall within the upper, mid-, or lower third (tercile) of conditions that occurred during 1961–90 (1971–2000 for forecasts issued since May 2001). A climatological probability (i.e., an outlook with zero probability of an anomaly) simply indicates that there is an equal chance (33.3%) that conditions will fall within any of the historical terciles. Overall, the forecasts consist of a 1-month outlook, issued with a 2-week lead time, and a series of thirteen 3-month outlooks, with lead times from one-half to twelve-and-one-half months. The CPC outlooks are produced for the 102 regions of the U.S. shown in Fig. 2. These regions are based on resampling the 344 climate divisions specified by NOAA's National Climate Data Center (NCDC). An entire suite is issued anew near the middle of each month (see online at www.cpc.ncep.noaa/gov/products/predictions).

Right from the appearance of these forecasts, one can detect potential difficulties in interpretation. In addition, one need look no further than Figs. 3–5 for evidence of how differently seasonal forecasts have performed in different areas, even within the Southwest. Figure 3 indicates that forecasts that depart from

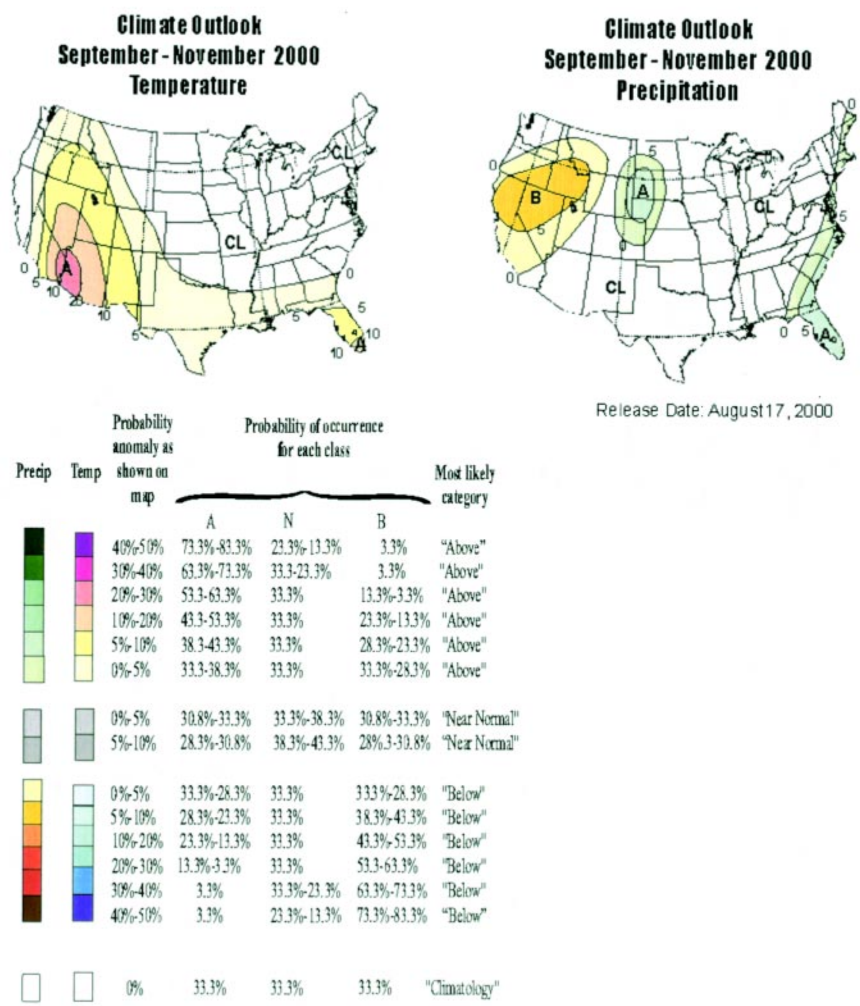

FIG. 1. Example official seasonal climate outlook and legend produced by the NOAA CPC. Outlook shown was issued Aug 2000 and covers Sep–Nov 2000. Maps show seasonal mean surface air temperature and seasonal total precipitation probability anomalies. For example, for the temperature outlook, the contour interval that includes Great Salt Lake, UT, shows a 5%–10% probability anomaly for warm temperatures. Adjustment of base probabilities (33.3% for each tercile) results in a 38.3%–43.3%, 33.3%, and 28.3%–23.3% probability, respectively, that seasonal temperatures will fall within the warm, near-normal, or cool tercile categories defined by regional conditions during 1961–90.

FORECASTS THAT COMMUNICATE

Qualitative aspects of climate forecast products can be as important as any quantitative attribute in affecting how users interpret, apply, and ultimately judge them (Nicholls 1999). Resource managers possess a range of abilities to obtain, interpret, and use climate forecasts (Pagano et al. 2002). Some resource managers employ meteorologists or experts, others hire consultants or rely on federal experts (D. Changnon 2000), and some interpret forecasts despite having no special training. Our discussions with users indicate that even managers with technical backgrounds consistently misinterpret CPC outlooks (Pagano et al. 2001). Forecasters should consider user input (Stern and Easterling 1999) to create outlooks that foster easy, accurate, and reliable interpretation. We raise here four presentation issues that deserve attention.

First, although CPC outlook maps illustrate temperature and precipitation probability anomalies, they are often interpreted as quantities. Map contours often form a "bull's-eye" (e.g., the area over southwest Arizona in Fig. 1), identifying regions for which stronger statements are being made about the likelihood of indicated conditions. However, these contours are often interpreted as meaning the center region is expected to have more extreme conditions than the surrounding areas. Although lower in spatial resolution, early versions of the experimental climate forecasts made by the International Research Institute for Climate Prediction (Fig. SB1) did not use contours, offering more reliable interpretation.

Second, anomalies are relative only to a limited historical period. In using 1961–90 as the reference climatology, recent conditions were not referenced (the problem will be temporarily reduced when new 1971–2000 climatological distributions are implemented). But recent conditions are what is fresh in decision makers' minds. For example, Changnon et al. (1988) found that agribusiness decision makers in the Midwest focus on conditions over the prior 3–6 yr. Decision makers may relate to climate forecasts better if they are compared to recent conditions. CPC's experimental probability of exceedance graphs (Barnston et al. 2000) show recent conditions (10 yr for temperature, 15 yr for precipitation), although the graphs pose their own interpretive difficulties.

Of course, the opposite problem is also true. Sometimes the forecasts unduly ignore the past. For some parts of the United States, periods separated by nearly a century have more in common than do intervening periods (e.g., Quinn 1981). With improving understanding of decadal- and centennial-scale climate regimes, it seems ill advised to ignore the longer historical record. The phenomenal population growth in the southwest United States means that many residents have lived there only a few years. Without historical perspective, newcomers will have little appreciation of the potential range of climatic variability.

Third, the climatology designation is misleading in the CPC outlooks, and sometimes it covers the entire conterminous states. This does not indicate that conditions are likely to be normal or even that forecasters actually believe each of the three conditions is equally likely. Rather, CPC states climatology when forecast techniques lack sufficient skill to differentiate or when individual forecast techniques

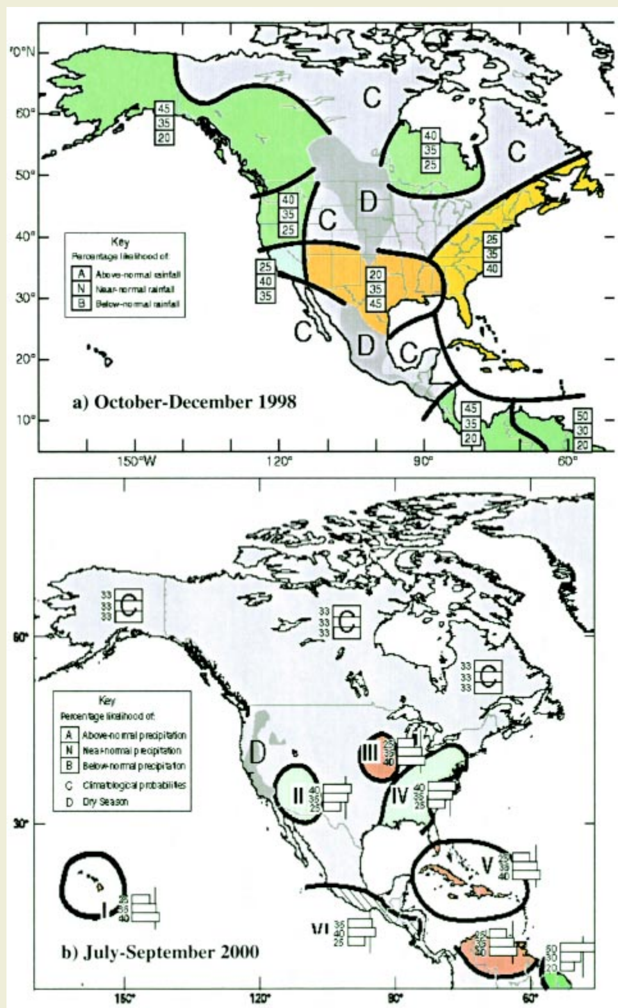

FIG. SB1. Seasonal climate outlooks produced by the International Research Institute for Climate Prediction: (a) older format, forecast produced Oct 1998 and covering Oct–Dec 1998, and (b) newer format, forecast produced Jun 2000 and covering Jul–Sep 2000.

produce conflicting guidance that cannot be resolved through forecaster expertise. Comparison with observations shows that climatology forecasts are not unbiased predictors of uniformly distributed conditions as some users have thought. Considering all regions and lead times, 52% and 21% of the climatology forecasts for temperature corresponded to observations in the warmest and coolest terciles, respectively, while 42% and 28% of climatology forecasts for precipitation corresponded to observations in the wettest and driest terciles.

We recommend that the climatology designation be replaced with a more explicit statement of no forecast confidence, like the phrase "indeterminate," which was used in NWS seasonal climate outlooks prior to mid-1982 (Bettge et al. 1981). Given the implications for decision making, to say "complete forecast uncertainty" or "no forecast confidence" would be more appropriate than to give an equal-probability forecast. Knowing this, users might decide to use other forecast approaches that have regional merit. For example, Lamb and Changnon (1981) found a 5-yr normal to work best overall in predicting seasonal precipitation and temperatures in Illinois; although when predictions were in error, the errors were larger than when based on 10- or 15-yr periods.

Finally, seasonal average temperatures have been consistently within the upper terciles of both the climatological period (1961–90) and the complete period of record. Thus an extended persistence forecast (e.g., 5-yr optimal climate normal or the 10- or 15-yr optimal climate normals considered by CPC in generating their forecasts) may be more appropriate as a "naïve" forecast than would a 30-yr climatology. Also, given such clear recent trends, seasonal temperature forecast skill may be good enough that forecasters should now aim for a target smaller than terciles. The International Research Institute for Climate Prediction is experimenting with this, assigning forecast probabilities to the extreme 15% distribution tails if they think sufficient information is available.

climatology have been made frequently from southeastern Arizona. Further, these nonclimatology forecasts generally turned out to be consistent with observations; that is, observations fell within the tercile category specified to have an enhanced probability of occurrence. Note that forecasts for a wet season can be correct, even though only a single month within the forecast period is actually wet (e.g., October–December 1997). On the other hand, (e.g., northeast Utah, Fig. 4) nonclimatology forecasts were relatively uncommon for regions in the upper Colorado River basin. Seasonal temperature forecasts for northwest Arizona and southern Nevada (Fig. 5) have been consistent with the extended and extremely warm observations of 1995–99. Tercile boundaries based on the region's entire historic record prior to the evaluation period (1895–1994; not shown) are similar to those of 1961–90, indicating how unusual recent temperatures have been.

Resource managers must be made aware of these regional differences, but further targeting of forecast evaluations is necessary: targeting by time, not just place. The water management agencies that govern delivery, reservoir levels, and flood control and response in the southwest United States are a case in point. Their responsibilities for making resource allocation decisions are as diverse as the watersheds they manage. We held extensive discussions with a broad range of water management professionals (Pagano et al. 2001; Carter et al. 2000) and identified periods especially important to decision makers concerned with seasonal water supplies originating primarily as mountain snowfall (Table 1). Snow accumulation in winter affects water supplies throughout the subsequent spring and summer. In the Southwest, the winter snow–summer flow relationship is particularly strong because in late spring and early summer there is little precipitation. Summer rains typically have little impact on useable water supplies in larger basins, but their transient local effects can be

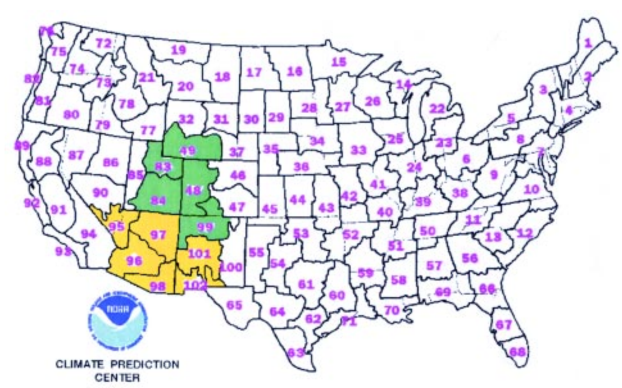

FIG. 2. Regions used in evaluating CPC seasonal climate outlooks. The upper and lower Colorado River basins are, respectively, indicated by the green and yellow highlighted regions.

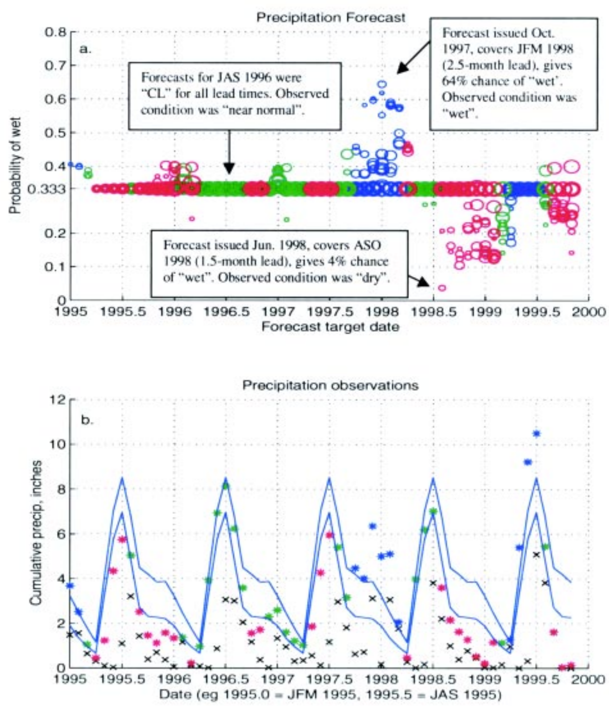

FIG. 3. Graphical comparison of CPC seasonal precipitation outlooks and observed seasonal total precipitation for southeastern Arizona (region 98 in Fig. 2). (a) Forecast probability of precipitation falling in the wet tercile; circle size indicates forecast lead time (0.5 to 12.5 months, smallest to largest). Color indicates observed category: wettest tercile (blue), middle tercile (green), driest tercile (red). (b) Observed precipitation; blue lines are 1961–90 precipitation tercile boundaries, asterisks are 3-month observed total precipitation using same color scheme as (a), and "Xs" are single-month observed precipitation.

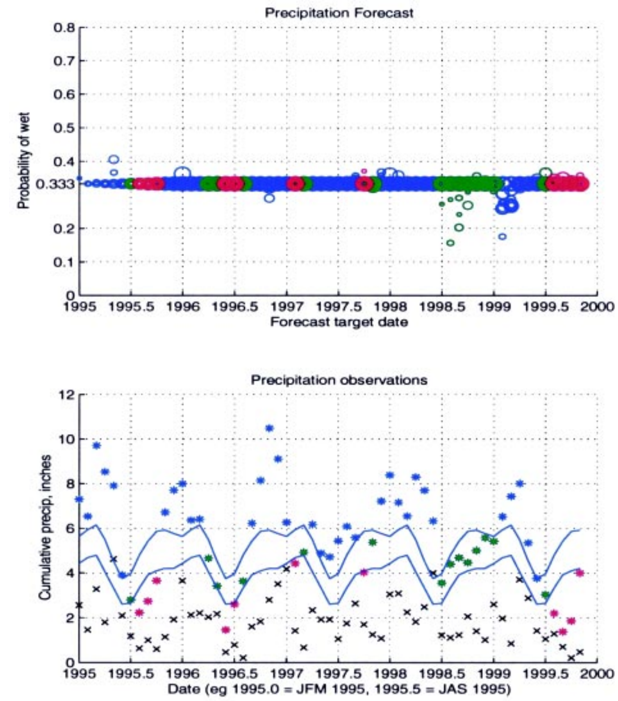

FIG. 4. Graphical comparison of CPC seasonal precipitation outlooks and observed seasonal total precipitation for northeastern Utah (region 83 in Fig. 2). Legend same as in Fig. 3.

important in irrigation districts where water managers augment rainfall with contracted water deliveries or groundwater pumping.

Typically, in October, several water supply management agencies meet with National Weather Service Colorado Basin River Forecast Center (RFC) personnel to review potential winter and early spring conditions that will ultimately affect seasonal water supplies. For basins in the southwest United States, the CPC climate outlooks relevant to this fall planning include forecasts issued in August, September, and October, covering December–May. Only those outlooks exclusively covering the decision period are considered here (i.e., December–February, January–March, February–April, March–May), to eliminate the influence of conditions in other months; this set of outlooks consists of 12 forecasts/ yr.

That is not all water managers in the Colorado Basin must consider. A second set of decisions involves the water supply outlooks that the RFC typically issues semimonthly, from January to May (Hartmann et al. 1999, 2002). The temporal coverage of the outlooks depends on basin seasonal flow characteristics. For the upper Colorado River basin (regions 48, 49, 83, 84, 99 in Fig. 2), water supply forecasts are issued from January through May and generally cover April–September, the period reflecting prolonged melting of extensive high-elevation snowpacks. The relevant climate forecasts are those issued December–April and covering January–September (25 forecasts/ yr).

A third set of decisions involves water supply outlooks for the lower Colorado River basin (regions 95–98, 101, 102). Here, the season of highest flows is much shorter, reflecting lower elevations, warmer temperatures, and a shallower snowpack. The RFC outlooks for this area involve climate forecasts issued during December–May and covering January–May. Only the climate forecasts issued December–February (6 forecasts/ yr) exclusively involve this period. Finally, a fourth decision period pertains to water managers who rely on summer delivery of seasonal water supplies to augment rainfall. Their perspective on climate forecasts includes forecasts issued January–May, covering June–September (10 forecasts/ yr).

DATA AND METHODS

We analyzed 3-month outlooks issued from December 1994 to October 1999 in a CPC dataset of digitized historical seasonal temperature and precipitation forecast maps. The digitization provides spatially weighted probability anomalies for each of 102 regions within the conterminous United States (Fig. 2). The regions are agglomerations of the 344 climate divisions used by the NOAA/NCDC. To calculate forecast probabilities, we added map anomalies to the climatological probability (33.3%). The observation dataset, covering January 1995–November 1999, consists of monthly average temperature and monthly total precipitation for each of the 102 regions based on spatially weighted averages of NCDC climate division data. Tercile boundaries are based on Gaussian and gamma distributions fitted to the 1961–90 observations for temperature and precipitation, respectively, with zero precipitation treated as censored data.

There are two caveats worth remembering about a quantitative evaluation of forecasts such as these. First, climate outlooks concern only average temperatures and total precipitation over an entire forecast period, not within it. They say nothing about daily, weekly, or even monthly extremes within a 3-month forecast period, or whether precipitation will occur as many small, or a few large, events. Yet in semiarid regions, seasonal precipitation can be the result of a single event. For example, the 1997/98 El Niño autumn climate was extremely wet in southwest Arizona due to the storm track of Hurricane Nora (Pagano et al. 1999). Second, limited sample sizes compromise even the most mathematically rigorous analyses. Spatial and temporal autocorrelations reduce effective sample sizes further, and forecast technology changes faster than sufficient data can accumulate.

Further, acceptable trade-offs between contextual specificity (e.g., region, lead time, season) and sufficient sampling differ by users' risk tolerance and are not reliably determined without their input.

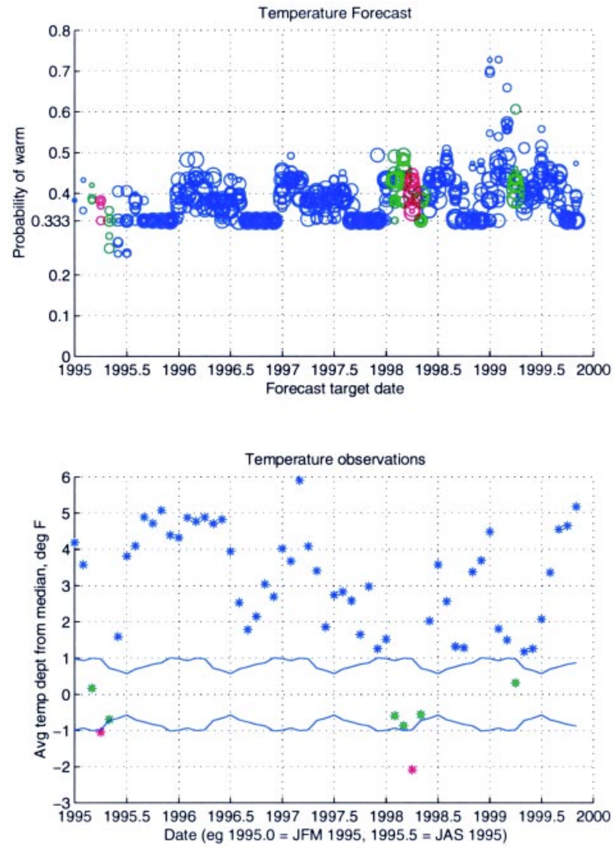

FIG. 5. Graphical comparison of CPC seasonal temperature outlooks and observed seasonal average temperatures for northwestern Arizona and southern Nevada (region 95 in Fig. 2). (a) Forecast probability of temperatures falling in the warm tercile; circle size indicates forecast lead time (0.5 to 12.5 months, smallest to largest). Color indicates observed category: warmest tercile (blue), middle tercile (green), coolest tercile (red). (b) Observed temperatures; blue lines are 1961–90 temperature tercile boundaries, asterisks are 3-month observed average temperature anomalies using same color scheme as (a).

CRITERIA FOR FORECAST EVALUATIONS. With these temporal and spatial considerations in mind, we selected several complementary forecast evaluations to show the range of information about climate outlooks that water resource managers in the southwest United States could use.

Myriad criteria exist for evaluating forecast quality (see Wilks 1995 for a detailed description of these criteria). It is helpful to select appropriate evaluation criteria that compare the CPC outlooks available at the time a decision is made to the information that decision makers would use if they had no CPC climate outlooks. "Climatology" forecasts (equal probabilities assigned to each tercile) are representative of what decision makers would use without forecasts. The relative improvement provided by the CPC outlooks, compared to forecasting climatology, is given as a skill score:

$$\text{Skill Score} = \frac{(\text{score}_{\text{forecasts}} - \text{score}_{\text{climatology}})}{(\text{score}_{\text{perfect}} - \text{score}_{\text{climatology}})} \times 100, \quad (1)$$

where $\text{score}_{\text{forecasts}}$, $\text{score}_{\text{climatology}}$, and $\text{score}_{\text{perfect}}$ are values (for the actual forecasts, climatology, and perfect

Table 1. Selected scenarios representing water management, ranching, and wildland fire management decision-making situations in the southwest United States.

Decision-making situation	When forecasts issued (months)	Season of interest (months)
Water management scenario		
Fall	Aug–Oct	Dec–May
Winter, upper Colorado	Dec–Apr	Jan–Sep
Winter, lower Colorado	Dec–Feb	Jan–May
Spring	Dec–May	Jun–Sep
Cattle ranching scenario		
Summer	Apr–May	Jul–Sep
Winter	Oct–Nov	Dec–Mar
Fire management scenario		
Spring	Jan–Mar	Apr–Jul

forecasts, respectively) determined by appropriate criteria.

CATEGORICAL MEASURES. In selecting the criteria for evaluating climate outlooks, one must bear in mind that these are not the same as typical weather forecasts. Weather forecasts project continuous variables like temperatures or the likelihood of discrete events like the occurrence of precipitation; the CPC climate forecasts, by contrast, indicate probability anomalies that are continuous variables, but organized into discrete intervals for discrete categories (e.g., "wet," "near normal," and "dry"). A plethora of forecast performance criteria use contingency tables that reflect whether forecasts "hit" or "miss" a discrete condition observed at verification time. These criteria include the Heidke skill scores that CPC posts online. To use these performance criteria with CPC forecasts, the condition (e.g., wet) receiving an enhanced likelihood is converted into a categorical forecast, having an implied 100% probability.

In this study we use instead the probability of detection (POD) and false alarm rate (FAR) scores, which address user concerns about specific climate conditions in simple terms. For a given condition, the probability of detection is the number of forecasts that ultimately prove correct relative to the total number of times the condition actually occurs. If the forecast correctly calls for wet conditions twice, yet over that time wet conditions occur 4 times, then the probability of detection is 50%, or 0.50. For a given condition, the false alarm rate is the number of forecasts that ultimately prove wrong, relative to the total number of times that forecast has been made. The false alarm rate is 0.50 if the forecasts call for wet conditions 4 times but are incorrect twice. The probability of detection considers all forecasts, while the false alarm rate considers only forecasts for anomalous conditions (i.e., nonclimatology forecasts).

Figure 6 shows probabilities of detection and false alarm rates for the forecasts relevant to the winter decision period for lower Colorado River managers. For example, Fig. 6d indicates that for southeast Arizona, when wet was forecast, then dry or normal occurred less often than expected by chance. CPC seasonal precipitation outlooks show skill for both wet and dry conditions for southeast Arizona; the specific forecasts used in the region's computations are illustrated in Fig. 7. However, with the exception of the false alarm rate for the wettest tercile, performance of the CPC seasonal precipitation outlooks for this decision period is poor for much of the nation outside the Southwest.

Forecast performance criteria based on "hitting" or "missing" associated observations (like probabilities of detection) offer users conceptually easy entry into discussions about forecast quality. Converting the probabilistic forecasts into categorical forecasts also highlights for users that climatology forecasts do not say any one category is more likely than another. Categorical measures are relatively simple to compute and communicate and can be related to specific user concerns. However, they unfairly penalize the CPC forecasts by neglecting differences between weak and strong confidence statements. Today's CPC outlooks are complicated by their probabilistic nature; the forecasts are never totally "wrong" because they never project a 100% chance of any condition. Each of the three terciles always shows at least a 3.3% probability, as indicated in the legend of Fig. 1.

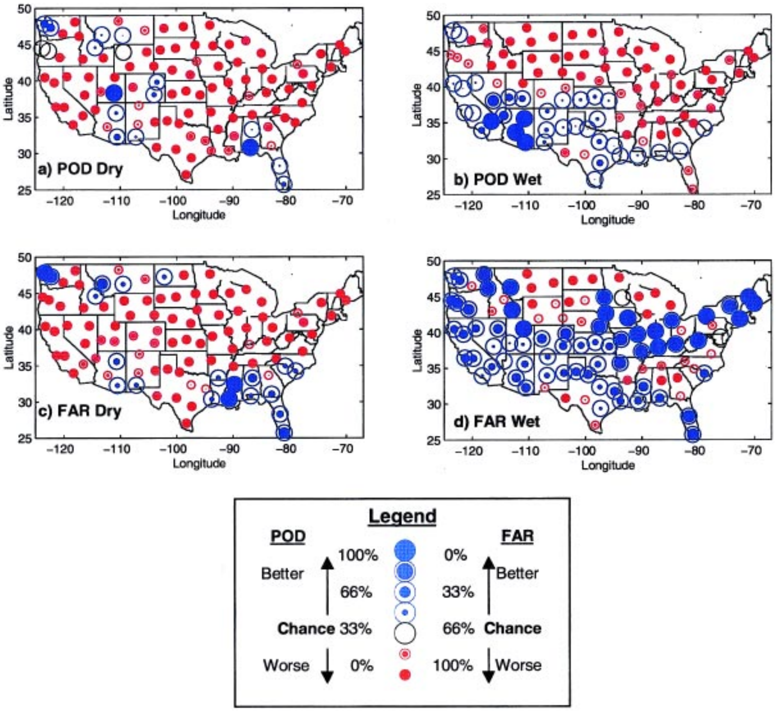

Fig. 6. Probability of detection (POD) and false alarm rate (FAR) for seasonal precipitation outlooks issued during Dec–Feb and covering Jan–May (water management scenario—winter, lower Colorado). POD corresponding to (a) driest and (b) wettest tercile and FAR for (c) driest and (d) wettest tercile. Blue (red) circles indicate climate outlooks are better (worse) than chance (33.3%) forecasts; black circle indicates absence of nonclimatology forecasts. Circle size indicates percent difference relative to potential shown by outer circle.

BRIER SCORES. The intuitive advantages of probability of detection and false alarm rate can be extended beyond categorical forecasts to probabilistic forecasts by using the Brier score (Brier 1950; Wilks 1995). This score is computed like the mean-square error, with the error for a single forecast–observation pair being the squared difference between the forecast probability for a tercile and the observed "probability" for that category (i.e., 0 if it did not occur and 1 if it did). Thus, if an event does not occur, forecasts indicating low probability for the event are not heavily penalized compared to those indicating high probability. Each tercile receives its own Brier score.

To evaluate CPC forecasts with the Brier score, only those forecasts that shift probabilities from tercile to tercile (nonclimatology forecasts) are included, since a climatology forecast actually signifies the absence of a forecast; thus the Brier score is comparable to the categorical false alarm rate. Figure 8 illustrates forecast skill based on the Brier score for the same lower Colorado River basin decision period as in Figs. 6 and 7. For the southwest United States, the seasonal climate outlooks are an improvement over using climatological probabilities for both wet and dry conditions, as expected from the evaluation with probabilities of detection and false alarm rates. Similarity between these scores and the Brier score results reflects that wet forecasts (generally limited to the winter of 1997/98) were made with relatively high probability and were consistent with subsequent observations (e.g., in Fig. 7). Forecasts for the driest tercile, however, are shown more favorably using the Brier score because some forecasts were made with low probability and dry conditions did not occur. Comparison of Brier scores and false alarm rates makes clear the potential for users to be disappointed in forecast performance. Ultimately, users attribute lower credibility to the outlooks than is deserved, when probabilities are converted to categorical forecasts.

Like probabilities of detection and false alarm rates, Brier scores allow users to focus on specific climate conditions (e.g., wet or dry). This is a drawback for some users, however, because the Brier score neglects the distribution of forecast probabilities outside the climate tercile of interest. Clearly, if a forecasts shifts most probability to the upper tercile, the user would

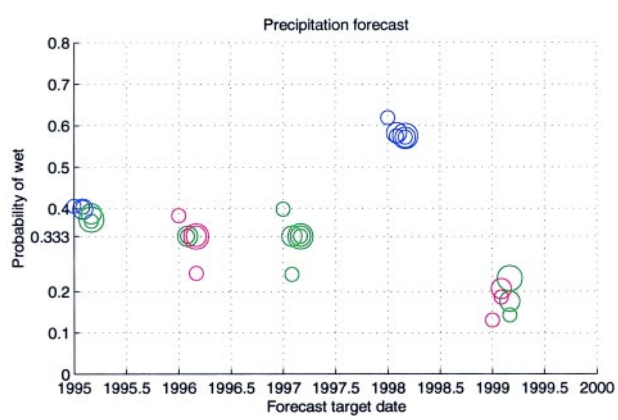

Fig. 7. Graphical comparison of CPC seasonal precipitation outlooks and observed seasonal total precipitation for southeastern Arizona (region 98 in Fig. 2) corresponding to outlooks issued during Dec–Feb and covering Jan–May (water management scenario—winter, lower Colorado). Legend same as in Fig. 3a; see Fig. 3b for corresponding observations.

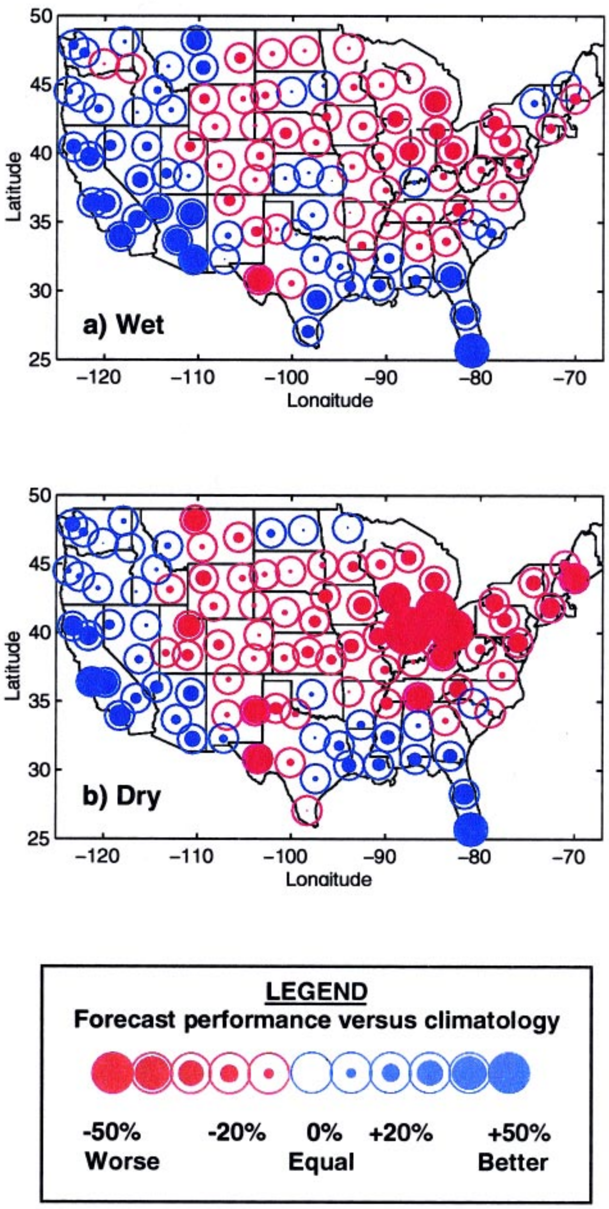

Fig. 8. Brier skill scores for seasonal precipitation outlooks issued during Dec–Feb and covering Jan–May (water management scenario—winter, lower Colorado). Only nonclimatological forecasts are considered: (a) wettest tercile, (b) driest tercile. Blue (red) circles indicate climate outlooks are better (worse) than climatological probabilities (33.3% each tercile); circle size indicates percent difference relative to 50% change shown by outer circle.

expect to see most of the rest of the probability distribution in the central tercile rather than in the lower tercile.

RANKED PROBABILITY SCORES. The ranked probability score allows consideration of multiple observation categories and cumulative forecast probabilities, which makes it more appropriate for users interested in the full range of conditions. Ranked probability scores, like Brier scores, are computed like mean-square error, but for each forecast–observation pair they compare cumulative forecast probabilities and multiple observation categories. Forecasts receive increasingly worse scores for assigning probabilities to categories increasingly distant from that observed (Epstein 1969; Wilks 1995). For example, if a season's observed total precipitation falls in the wet tercile, a forecast with probabilities of 28%, 33%, and 39% assigned to the dry, near-normal, and wet terciles, respectively, would not score as well as a forecast with respective probabilities of 3%, 24%, and 73%. Probabilities for all terciles are used to compute a single ranked probability score. As in the Brier score, climatology forecasts are considered nonforecasts and are excluded.

Figure 9 shows ranked probability scores computed for each of three water management scenarios—fall, winter (in the lower Colorado River basin), and summer. Seasonal precipitation outlooks covering the winter, made with the shortest lead times, show ranked probability score skill for most of the Pacific coast, Southwest, and Gulf coast regions (Fig. 9b). Even with longer lead times (up to 6.5 months in the fall), the CPC forecasts show a skillful ranked probability score for the Southwest and Gulf coast regions (Fig. 9a). However, few regions show any such skill for summer (Fig. 9c); some regions have never had nonclimatology forecasts issued for this period (including parts of the southwest United States).

Ranked probability scores logically build upon Brier score evaluations, because they are computed similarly. The ability to separately evaluate forecasts for specific climate conditions is lost, however. This means that ranked probability scores are most useful for situations where forecast consequences are similar among all climate conditions or where a decision maker can afford to play the odds across many events or regions. For example, energy or commodity traders would benefit from use of ranked probability scores. But an orchard irrigator holding junior water rights, who may be pleasantly surprised by wet conditions but devastated by unforeseen dry conditions, would benefit more from using Brier scores for dry conditions.

CONDITIONAL DISTRIBUTION DIAGRAMS. Conditional distributions of forecasts and observations provide the most comprehensive evaluations available to users. In this article we use two such dia-

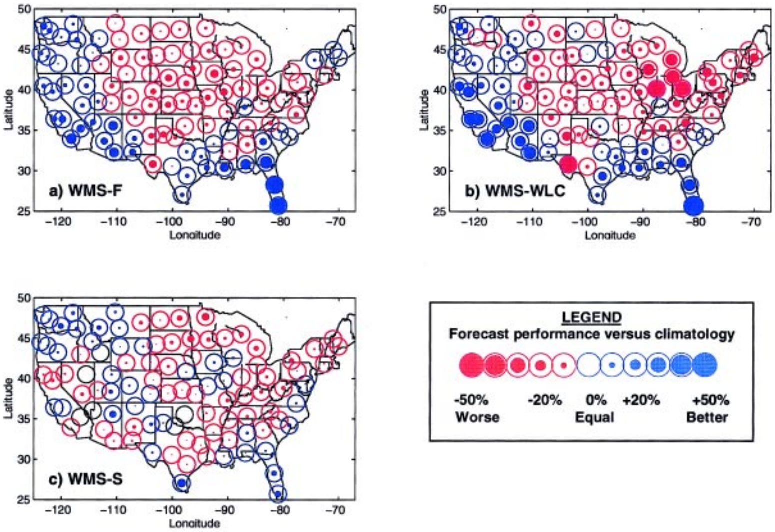

FIG. 9. Skill scores for ranked probability scores for seasonal precipitation outlooks for water management in the Southwest. Panels correspond to water management scenarios: (a) fall; (b) winter, lower Colorado; and (c) summer. Rest of legend same as in Fig. 8.

grams for evaluating forecasts. The first type of conditional distribution in this paper is discrimination diagrams that plot forecast probabilities by observed condition. In the terminology of Murphy (1993), these distributions identify the ability of the forecast system to "discriminate" among climate events. For example, the distribution diagram for wet tercile conditions shows how frequently (when such conditions occurred) the forecasts specified each possible probability of wet tercile conditions. Similarly, the dry tercile diagram charts how often the forecasts showed each probability when conditions turned out to be in the dry tercile (see Murphy and Winkler 1987, 1992 for a more complete background on conditional distributions). Markedly different probability distributions should be associated with different climate conditions.

The second type of conditional distribution is the reliability diagram. This shows, given each forecast probability interval, how frequently observations actually ended up in one or another tercile. One can thus see, for instance, whether conditions turned out to be dry 30% of the time that forecasts indicated 30% chance of dry tercile conditions. In the terminology of Murphy (1993), these distributions identify "reliability" by showing how well forecast probabilities correspond with their associated relative frequencies of "correct" observations. These diagrams address questions about the confidence that might be attributed to a forecast in hand, analogous to the false alarm rate.

Murphy and Huang (1991) and Wilks (2000) have already done comprehensive evaluations for CPC climate outlooks, but the two conditional distributions selected here for evaluation are those that can be related to simpler criteria and user concerns. Figure 10 shows discrimination diagrams for the winter management scenarios in the upper and lower Colorado River ba-

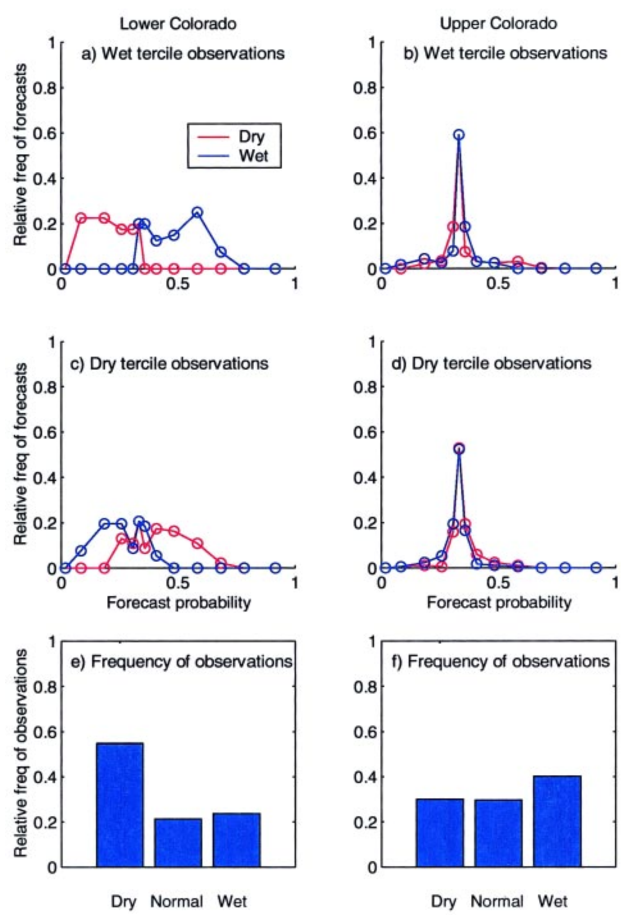

FIG. 10. (a)–(d) Discrimination diagrams and (e), (f) observation histograms for seasonal precipitation outlooks relevant to water supply management: (a), (c), (e) lower Colorado River basin, issued Dec–Feb and covering Jan–May (water management scenario—winter, lower Colorado); and (b), (d), (f) upper Colorado River basin, issued Dec–Apr and covering Jan–Sep (water management scenario—winter, upper Colorado). Discrimination diagrams conditioned on observations for (a), (b) wet and (c), (d) dry terciles. Blue (red) circles show relative frequency of forecasts issued with indicated probability for wet (dry) conditions.

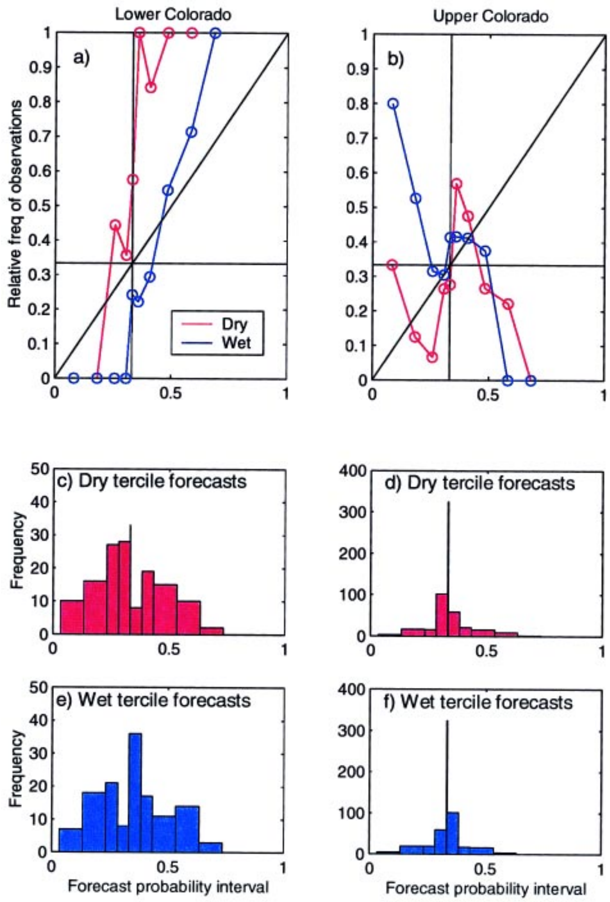

FIG. 11. (a), (b) Reliability diagrams and (c)–(f) forecast probability histograms for seasonal precipitation outlooks relevant to water supply management: (a), (c), (e) lower Colorado River basin, issued Dec–Feb and covering Jan–May (water management scenario—winter, lower Colorado); and (b), (d), (f) upper Colorado River basin, issued Dec–Apr and covering Jan–Sep (water management scenario—winter, upper Colorado). In reliability diagrams, blue (red) circles show fraction of wet (dry) observations occurring when forecast with the indicated probability. Black lines refer to climatological forecast probability (vertical) and observation frequency (horizontal), and perfect forecast reliability (1:1 line). Histograms conditioned on forecasts for (c), (d) dry and (e), (f) wet terciles, with bin widths based on forecast probability intervals (see Fig. 1).

sins, along with frequency histograms of the observation categories. In Fig. 10, perfect forecasts would show, for wet observations, that all forecasts had specified probabilities of 100% and 0% for wet and dry conditions to occur, respectively. While far from perfect, with the exception of climatology forecasts, lower Colorado River basin forecasts show complete discrimination for wet observations (i.e., no forecasts with enhanced likelihood of dry conditions have been followed by wet observations). Dry observations show some overlap in forecast probabilities specified for wet and dry conditions, but not nearly so much as for the upper Colorado River basin. There, the distributions of forecast probabilities for wet and dry conditions vary little between wet and dry observations, and are dominated by climatology statements.

Figure 11 shows reliability diagrams for the winter scenario in the upper and lower Colorado River basins along with frequency histograms of the probability intervals used to forecast wet and dry conditions. Perfect specification of forecast confidence would result in perfect alignment of forecast probabilities and observational frequencies (i.e., all points in Fig. 11 falling along a 1:1 line). For the lower basin, for forecasts of both wet and dry conditions, the frequency of wet and dry observations generally increases with the forecast probability, although forecast confidence has been understated, especially for forecasts of dry conditions. In contrast, upper basin forecasts show poor reliability, especially for the most extreme forecast probabilities.

Conditional distributions offer the advantage of identifying specific situations whereby forecasts perform particularly well or poorly. But they are sensitive to small sample sizes and unfortunately the CPC dataset included only six years of data at the time of this study. Obtaining sufficient sample sizes for each prospective situation is problematic, requiring pragmatic grouping of forecast–observation pairs across multiple regions, lead times, seasons, and distribution categories. The diagrams are also probably too complex to interpret for all but the large water management agencies and other groups staffed with specialists.

LESSONS FOR USERS. From the perspective of water managers in the southwest United States, winter precipitation outlooks made during fall and winter are better than climatology forecasts according to all criteria. Winter and spring forecasts of summer precipitation lack skill, and in many areas have provided no guidance at all, indicating only climatology. The implications of forecast performance are different for water managers of the upper versus lower Colorado River basins. The poor skill in forecasting the climate of the upper basin, where high-elevation snowpacks are the main source of water, reinforces the vital importance of high quality estimates of snowpack. In the lower basin, snowpack is less extensive and less reliable; flow forecasts depend on precipitation and thus are less predictable (Shafer and Huddleston 1985). The evident climate forecast skill in the lower basin during winter and spring offers potential for improving streamflow predictions.

Compared to the upper Colorado River basin, not only does the lower basin benefit from greater storage capacity (Harding et al. 1995), but from greater climate predictability as well.

These techniques are by no means limited in usefulness to water management. A glance at cattle ranching and wildland fire management in the southwest United States shows similar subtleties in the use of climate forecast evaluations (Table 1).

In the case of ranchers, vulnerability to climate stems, in part, from dependence on rain-fed rangelands, although specific vulnerabilities also depend on the location and type of operations (e.g., cow-calf or steer operations). Ranchers' interests in climate forecasts primarily involve the predictability of grass production. They may increase herd size to exploit a good grass season, or purchase additional feed and remove stock to prevent degradation of rangelands during drought.

Ranches are highly diverse, making it difficult to generalize about which climate forecasts are most important. The specific season of interest depends on whether rangelands are most productive during winter (generally low-elevation ranges) or summer (high-elevation ranges). We picked two representative scenarios to analyze, one for grasslands that peak in summer, encompassing forecasts made in April and May for July–September (2 forecasts/ yr), and a second for grasslands that peak in winter, encompassing forecasts made in October and November for December–March conditions (8 forecasts/ yr).

Figure 12 assesses seasonal temperature and precipitation outlooks using ranked probability scores from the perspective of cattle ranchers. Figure 12a shows that even with the relatively short lead times of ranchers, CPC outlooks have provided no information about summer precipitation, although summer temperature forecasts do show some skill for portions of the southwest United States (Fig. 12b). CPC outlooks for winter conditions show skill for both precipitation and temperature (Figs. 12c and 12d, respectively), with the Southwest showing generally the best combined performance.

Within the southwest United States, climate forecasts perform better from the perspective of cattle ranchers making use of winter range, compared to those using summer range. The greater climate predictability provides competitive advantages, if the ranchers have financial and operational flexibility to exploit them. Anything ranchers can do to shift operational risks from summer to winter would improve their competitive situation.

Not surprisingly, wildland fire managers also pay close attention to the growth of vegetation. The specific seasons of interest to fire management agencies depend on the climate, elevation, and land cover type within their area of jurisdiction (ISPE 2000). For instance, low-elevation grasslands can be at high risk throughout a dry winter, while high-elevation forests generally have later fire seasons.

Wildland fires in the southwest United States typically result from a reliably arid spring and early summer followed by lightning storms that occur prior to the summer monsoon (Swetnam and Betancourt 1990). Increasingly, prescribed burn programs are used to prevent catastrophic wildfires by preemptively removing fuels under ideal field conditions. Fire risk is also highly conditioned on prior climatic

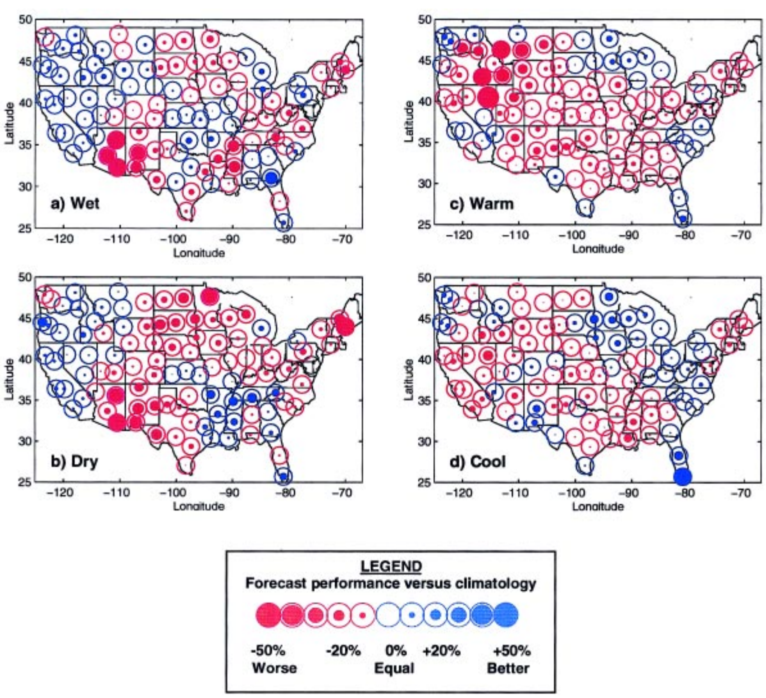

Fig. 12. Skill scores for ranked probability scores for seasonal precipitation and temperature outlooks for cattle ranching in the Southwest. Panels correspond to cattle ranching scenario—summer: (a) precipitation and (c) temperature, and winter: (b) precipitation and (d) temperature. Rest of legend same as in Fig. 8.

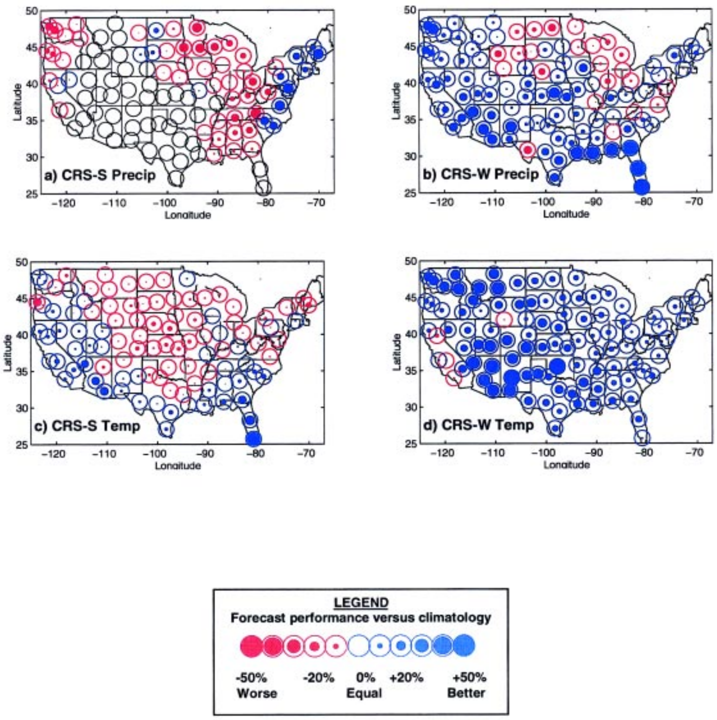

Fig. 13. Skill scores for Brier score for seasonal precipitation and temperature outlooks issued during Jan–Mar and covering Apr–Jul (fire management scenario—spring). Only nonclimatological forecasts are considered. Panels show precipitation: (a) wettest and (b) driest; and temperature: (c) warmest and (d) coolest terciles. Legend same as Fig. 8.

conditions, with the second dry winter after a wet winter, and a dry summer following a dry winter, being critical in the Southwest (Swetnam and Betancourt 1990).

Regional fire managers from throughout the country submit their requests for seasonal resource needs (e.g., fire crews, equipment) to national headquarters in March, regardless of their actual fire season (ISPE 2000). To analyze forecast reliability for this spring decision period we produced scores for forecasts issued January–March and covering April–July (6 forecasts/yr).

Figure 13 assesses seasonal temperature and precipitation outlooks in this period using Brier scores. For a few areas within the Southwest, there is some skill with regard to cool temperatures that may be useful for planning seasonal prescribed burn opportunities. However, precipitation and temperature forecasts covering April–July are generally poor, even for the relatively short lead times considered.

The poor performance of climate outlooks from the perspective of wildland fire managers suggests real potential for misuse if the forecasts are incorporated into decisions without extreme care. We explored whether requiring fire managers to submit budget, equipment, and personnel requests in March limits the potential usefulness of the climate outlooks. This analysis of forecasts issued at other times of the year involved combining the 102 regions of Fig. 2 into 10 regions (not shown) representing current wildland fire management divisions (www.nifc.gov/fireinfo/geomap.html) as well as combining versions of these divisions to produce more homogenous climatic zones. We looked at forecasts for each season made with one-half-month lead times. Only those regions that have important winter fire risks (e.g., southern California) could possibly benefit from seasonal climate outlook skill. However, for other regions, to the extent that fire risks are affected by winter conditions, forecast skill at even longer lead times may prove useful, but not under the current resource request schedule.

A STARTING POINT FOR IMPROVING EVALUATIONS.

Frequently updated forecast evaluations, using multiple criteria, should be available to potential users of seasonal climate outlooks. The evaluation framework presented here provides several criteria that accommodate variations in users' interpretive abilities. The framework offers trade-offs between different levels of informativeness and understandability, and enables users to increase the sophistication of their understanding about climate forecasts, their credibility, and implications of using them for decision making.

With the evaluations presented here, decision makers and forecasters can begin to determine essential forecast attributes, requisite performance thresholds, and relationships among the quality of forecasts and their usefulness in decision making, and ultimately their economic value. Graphing recent forecasts and observations together enables intuitive identification of multiple performance attributes. Products such as Figs. 3–5 represent a starting point for determining, through interaction with users, effective formats. Forecast characteristics that stand out graphically include the seasonality of forecast confidence, tendencies to use only climatology (i.e., to not provide a forecast), and consistency in the direction of predicted anomalies as forecast lead times get shorter. The extremity or normalcy of recent conditions can also be clearly identified and placed in historical context. Finally, specific forecasts can be directly compared to their associated observations. Although it is more appropriate to judge probabilistic forecasts in the aggregate rather than individually, a time series of all forecasts should help users visualize aggregate performance.

The demonstrated forecast skill, or lack thereof, provides a basis of experience for exploring the implications of forecast performance. Forecast quality can have implications for prioritizing scientific efforts, realizing competitive advantages, adjusting management processes, and changing climate forecasting efforts. With the evaluation framework presented here, we believe that resource managers will more readily realize the potential of climate forecasts from the CPC.

ACKNOWLEDGMENTS. We gratefully acknowledge the cooperation of many Arizona water management agencies, cattle ranchers, and wildland fire management agencies; forecasters at the NOAA CPC; and our CLIMAS colleagues. Support was provided by NASA/EOS Grant OSSA-A/88, NOAA OGP Grant NA 86GP0061, NASA Grant NAG5-8503, and NSF-STC Grant EAR-9876800.

REFERENCES

Barnett, T. P., and R. W. Priesendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for North American surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.,* **115,** 1825–1850.

Barnston, A. G., Y. He, and D. A. Unger, 2000: A forecast product that maximizes utility for state-of-the-art seasonal prediction. *Bull. Amer. Meteor. Soc.,* **81,** 1271–1279.

Bettge, T. W., D. P. Baumhefner, and R. M. Cherwin, 1981: On the verification of seasonal climate forecasts. *Bull. Amer. Meteor. Soc.,* **62,** 1654–1665.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.,* **78,** 1–3.

Callahan, B., E. Miles, and D. Fulharty, 1999: Policy implications of climate forecasting for water resources management in the Pacific Northwest. *Policy Sci.,* **32,** 269–293.

Carter, R. H., P. Tschakert, and B. Morehouse, 2000: Assessing the sensitivity of the Southwest's urban water sector to climate variability: Case studies in Arizona. CLIMAS CL1-00, Institute for the Study of Planet Earth, The University of Arizona, Tucson, AZ, 110 pp.

Changnon, D., 2000: Who used and benefited from the El Niño forecasts? *El Niño 1997–1998: The Climate Event of the Century,* S. A. Changnon, Ed., Oxford University Press, 109–135.

Changnon, S. A., 1990: The dilemma of climatic and hydrologic forecasting for the Great Lakes. *Proc. Great Lakes Water Level Forecast and Statistics Symp.,* Ann Arbor, MI, Great Lakes Commission, 13–25.

——, 1992: Contents of climate predictions desired by agricultural decision makers. *J. Appl. Meteor.,* **31,** 1488–1491.

——, 2000: *El Niño 1997–1998: The Climate Event of the Century.* Oxford University Press, 215 pp.

——, and D. R. Vonnhame, 1986: Use of climate predictions to decide a water management problem. *Water Resour. Bull.,* **22,** 649–652.

——, S. T. Sonka, and S. Hofing, 1988: Assessing climate information use in agribusiness. Part I: Actual and potential use and impediments to use. *J. Climate,* **1,** 757–765.

Conley, J., H. Eakin, T. E. Sheridan, and D. Hadley, 1999: CLIMAS ranching case study: Year 1. CLIMAS CL3-99, Institute for the Study of Planet Earth, The University of Arizona, Tucson, AZ, 27 pp.

Epstein, E. S., 1969: A scoring system for probabilities of ranked categories. *J. Appl. Meteor.,* **8,** 985–987.

Glantz, M. H., 1982: Consequences and responsibilities in drought forecasting—The case of Yakima, 1977. *Water Resour. Res.,* **18,** 3–13.

Harding, B. L., T. B. Sangoyomi, and E. A. Payton, 1995: Impacts of a severe sustained drought on Colorado River water resources. *Water Resour. Bull.,* **31,** 815–824.

Hartmann, H. C., R. Bales, and S. Sorooshian, 1999: Weather, climate, and hydrologic forecasting for the southwest U.S. CLIMAS CL2-99, Institute for the Study of Planet Earth, The University of Arizona, Tucson, AZ, 172 pp.

——, ——, and ——, 2002: Weather, climate, and hydrologic forecasting for the U.S. southwest: A survey. *Climate Res.,* in press.

Institute for the Study of Planet Earth, 2000: *The Implications of La Niña and El Niño for Fire Management.* Tucson, AZ, CLIMAS, ISPE, and the Laboratory for Tree Ring Research, The University of Arizona, 45 pp.

Lamb, P. J., and S. A. Changnon, 1981: On the best temperature and precipitation normals—The Illinois situation. *J. Appl. Meteor.,* **20,** 1383–1390.

Lehman, R. L., 1987: A model for decision making based on NWS monthly temperature outlooks. *J. Climate Appl. Meteor.,* **26,** 263–274.

Livezey, R. E., 1990: Variability of skill of long-range forecasts and implications for their use and value. *Bull. Amer. Meteor. Soc.,* **71,** 300–309.

Mjelde, J. W., D. S. Peel, S. T. Sonka, and P. L. Lamb, 1993: Characteristics of climate forecast quality: Implications for economic value to midwestern corn producers. *J. Climate,* **6,** 2175–2187.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting,* **8,** 281–293.

——, and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.,* **115,** 1330–1338.

——, and J. Huang, 1991: On the quality of CAC's probabilistic 30-day and 90-day forecasts. *Proc. 16th Annual Climate Diagnostics and Prediction Workshop,* Springfield, VA, National Weather Service, 390–399. [NTIS PB92-167378.]

——, and R. L. Winkler, 1992: Diagnostic verification of probability forecasts. *Int. J. Forecasting,* **7,** 435–455.

Namias, J., 1968: Long-range weather forecasting—History, current status and outlook. *Bull. Amer. Meteor. Soc.,* **49,** 438–470.

Nicholls, N., 1980: Long-range weather forecasting: Value, status, and prospects. *Rev. Geophys. Space Phys.,* **18,** 771–778.

——, 1999: Cognitive illusions, heuristics, and climate prediction. *Bull. Amer. Meteor. Soc.,* **80,** 1385–1398.

Office of Global Programs, 1999: An experiment in the application of climate forecasts: NOAA–OGP activities related to the 1997–98 El Niño event. Office of Global Programs, National Oceanic and Atmospheric Administration, Silver Spring, MD, 134 pp.

O'Grady, K., and L. Shabman, 1990: Communicating the probability of Great Lakes water levels and storms. *Proc. Great Lakes Water Level Forecast and Statistics Symp.,* Windsor, Ontario, Canada, Great Lakes Commission, 197–204.

O'Lenic, E. A., 1990: Short-term climate prediction in the Great Lakes region. *Proc. Great Lakes Water Level Forecast and Statistics Symp.,* Windsor, Ontario, Canada, Great Lakes Commission, 27–40.

Pagano, T. C., H. C. Hartmann, S. Sorooshian, and R. Bales, 1999: Advances in seasonal forecasting for water management in Arizona: A case study of the 1997–98 El Niño. Dept. Hydrol. Water Resources, Rep. HWR 99-040, The University of Arizona, Tucson, AZ, 221 pp.

——, ——, and ——, 2000: Climate forecasts: A new tool for emergency management in the southwest U.S. *Nat. Hazards Observ.,* **24,** 7–8.

——, ——, and ——, 2001: Using climate forecasts for water management: Arizona and the 1997–98 El Niño. *J. Amer. Water Resour. Assoc.,* **37,** 1139–1152.

——, ——, and ——, 2002: Use of climate forecasts for water management in Arizona: A case study of the 1997–98 El Niño. *Climate Res.,* in press.

Priesendorfer, R. W., and C. D. Mobley, 1984: Climate forecast verification, United States mainland, 1974–83. *Mon. Wea. Rev.,* **112,** 809–825.

Pulwarty, R. S., and K. T. Redmond, 1997: Climate and salmon restoration in the Columbia River basin: The role and usability of seasonal forecasts. *Bull. Amer. Meteor. Soc.,* **78,** 381–397.

——, and T. S. Melis, 2001: Climate extremes and adaptive management on the Colorado River: Lessons from the 1997–98 ENSO event. *J. Environ. Manage.,* **63,** 307–324.

Quinn, F. H., 1981: Secular changes in annual and seasonal Great Lakes precipitation, 1854–1979, and their implications for Great Lakes water resources studies. *Water Resour. Res.,* **17,** 1619–1624.

Sarewitz, D., R. A. Pielke Jr., and R. Byerly Jr., 2000: *Prediction: Science, Decision Making, and the Future of Nature,* Island Press, 405 pp.

Shafer, B. A., and J. M. Huddleston, 1985: Analysis of seasonal volume streamflow forecast errors in the western United States. *A Critical Assessment of Forecasting in Western Water Resources Management,* J. Cassidy and D. Lettenmaier, Eds., American Water Resources Association, 117–126.

Sonka, S. T., S. A. Changnon, and S. L. Hofing, 1992: How agribusiness uses climate predictions: Implications for climate research and provision of predictions. *Bull. Amer. Meteor. Soc.,* **73,** 1999–2008.

Stern, P. C., and W. E. Easterling, Eds., 1999: *Making Climate Forecasts Matter.* National Academy Press, 175 pp.

Swetnam, T. W., and J. L. Betancourt, 1990: Fire–Southern Oscillation relations in the southwestern United States. *Science,* **249,** 1017–1020.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, 467 pp.

——, 2000: Diagnostic verification of the Climate Prediction Center long-lead outlooks, 1995–98. *J. Climate,* **13,** 2389–2403.