CHAPTER 2

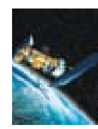
A Description and Evaluation of Hydrologic and Climate Forecast and Data Products that Support Decision-Making for Water Resource Managers

Convening Lead Author: Nathan Mantua, Climate Impacts Group, Univ. of Washington

Lead Authors: Michael Dettinger, U.S.G.S., Scripps Institution of Oceanography; Thomas C. Pagano, N.W.C.C., NRCS/USDA; Andrew Wood, 3TIER™, Inc./Dept. of Civil and Environmental Engineering, Univ. of Washington; Kelly Redmond, W.R.C.C., Desert Research Institute

Contributing Author: Pedro Restrepo, NOAA

KEY FINDINGS



There are a wide variety of climate and hydrologic data and forecast products currently available for use by decision makers in the water resources sector, ranging from seasonal outlooks for precipitation and surface air temperature to drought intensity, lake levels, river runoff and water supplies in small to very large river basins. However, the use of official seasonal-to-interannual (SI) climate and hydrologic forecasts generated by National Oceanic and Atmospheric Administration (NOAA) and other agencies remains limited in the water resources sector. Forecast skill, while recognized as just one of the barriers to the use of SI climate forecast information, remains a primary concern among forecast producers and users. Simply put, there is no incentive to use SI climate forecasts when they are believed to provide little additional skill to existing hydrologic and water resource forecast approaches. Not surprisingly, there is much interest in improving the skill of hydrologic and water resources forecasts. Such improvements can be realized by pursuing several research pathways, including:

- Improved monitoring and assimilation of real-time hydrologic observations in land surface hydrologic models that leads to improved estimates for initial hydrologic states in forecast models;
- Increased accuracy in SI climate forecasts; and,
- Improved bias corrections in existing forecast.

Because runoff and forecast conditions are projected to gradually and continually trend towards increasingly warmer temperatures as a consequence of human-caused climate change, the expected skill in regression-based hydrologic forecasts will always be limited by having only a brief reservoir of experience with each new degree of warming. Consequently, we must expect that regression-based forecast equations will tend to be increasingly and perennially out of date in a world with strong warming trends. This problem with the statistics of forecast skill in a changing world suggests that development and deployment of more physically-based, less statistically-based, forecast models should be a priority in the foreseeable future.

Another aspect of forecasts that serves to limit their use and utility is the challenge in interpreting forecast information. For example, from a forecast producer's perspective, confidence levels are explicitly and quantitatively conveyed by the range of possibilities described in probabilistic forecasts. From a forecast user's perspective, probabilistic forecasts are not always well understood or correctly interpreted. Although structured user testing is known to be an effective product development tool, it is rarely done. Evaluation should be an integral part of improving forecasting efforts, but that evaluation should be extended to factors that encompass use and utility of forecast information for stakeholders. In particular, very little research is done on effective seasonal forecast communication. Instead, users are commonly engaged only near the end of the product development process.

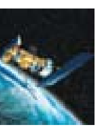
Other barriers to the use of SI climate forecasts in water resources management have been identified and those that relate to institutional issues and aspects of current forecast products are discussed in Chapters 3 and 4 of this Product.

Pathways for expanding the use and improving the utility of data and forecast products to support decision making in the water resources sector are currently being pursued at a variety of spatial and jurisdictional scales in the United States. These efforts include:

- An increased focus on developing forecast evaluation tools that provide users with opportunities to better understand forecast products in terms of their expected skill and applicability;
- Additional efforts to explicitly and quantitatively link SI climate forecast information with SI hydrologic and water supply forecasting efforts;
- An increased focus on developing new internet-based tools for accessing and customizing data and forecast products to support hydrologic forecasting and water resources decision making; and,
- Further improvements in the skill of hydrologic and water supply forecasts.

Many of these pathways are currently being pursued by the federal agencies charged with producing the official climate and hydrologic forecast and data products for the United States, but there is substantial room for increasing these activities.

An additional important finding is that recent improvements in the use and utility of data and forecast products related to water resources decision-making have come with an increased emphasis on these issues in research funding agencies through programs like the Global Energy and Water Cycle Experiment (GEWEX, a program initiated by the World Climate Research Programme) and NOAA's Regional Integrated Sciences and Assessment (RISA), Sectoral Applications Research Program (SARP), Transition of Research Applications to Climate Services (TRACS) and Climate Prediction Program for the Americas (CPPA) programs. Sustaining and accelerating future improvements in the use and utility of official data and forecast products in the water resources sector rests, in part, on sustaining and expanding federal support for programs focused on improving the skill in forecasts, increasing the access to data and forecast products, and supporting sustained interactions between forecast producers and consumers. One strategy is to support demonstration projects that result in the development of new tools and applications that can then be transferred to broader communities of forecast producers, including those in the private sector, and broader communities of forecast consumers.



2.1 INTRODUCTION

In the past, water resource managers relied heavily on observed hydrologic conditions such as snowpack and soil moisture to make seasonal-to-interannual (SI) water supply forecasts to support management decisions. Within the last decade, researchers have begun to link SI climate forecasts with hydrologic models (*e.g.*, Kim *et al.*, 2000; Kyriakidis *et al.*, 2001) or statistical distributions of hydrologic parameters (*e.g.*, Dettinger *et al.*, 1999; Sankarasubramanian and Lall, 2003) to improve hydrologic and water resources forecasts. Efforts to incorporate SI climate forecasts into water resources forecasts have been prompted, in part, by our growing understanding of the effects of global-scale climate phenomena, like El Niño-Southern Oscillation (ENSO), on U.S. climate, and the expectation that SI forecasts of hydrologically-significant climate variables like precipitation and temperature provide a basis for predictability that is not currently being exploited. To the extent that climate variables like temperature and precipitation can be forecasted seasons in advance, hydrologic and water-supply forecasts can also be made skillfully well before the end, or even beginning, of the water year[1].

More generally speaking, the use of climate data and SI forecast information in support of water resources decision making has been aided by efforts to develop programs focused on fostering sustained interactions between data and forecast producers and consumers in ways that support co-discovery of applications (*e.g.* see Miles *et al.*, 2006).

This Chapter focuses on a description and evaluation of hydrologic and climate forecast and data products that support decision making for water resource managers. Because the focus of this CCSP Product is on using SI forecasts and data for decision support in the water resources sector, we frame this Chapter around key forecast and data products that contribute towards improved hydrologic and water supply forecasts. As a result, this Product does not contain a comprehensive review and assessment of the entire national SI climate and hydrologic forecasting effort. In addition, the reader should note that, even today, hydrologic and water supply forecasting efforts in many places are still not inherently linked with the SI climate forecasting enterprise.

Surveys identify a variety of barriers to the use of climate forecasts (Pulwarty and Redmond, 1997; Callahan *et al.*, 1999; Hartmann *et al.*, 2002), but insufficient accuracy is always mentioned as a barrier. It is also well established that an accurate forecast is a necessary, but in and of itself, insufficient condition to make it useful or usable for decision making in management applications (Table 2.1). Chapters 3 and 4 provide extensive reviews, case studies, and analyses that provide insights into pathways for lowering or overcoming barriers to the use of SI climate

Table 2.1 Barriers to the use of climate forecasts and information for resource managers in the Columbia River Basin (Reproduced from Pulwarty and Redmond, 1997).

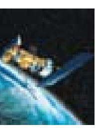
a. Forecasts not "accurate" enough.
b. Fluctuation of successive forecasts ("waffling").
c. The nature of what a forecast is, and what is being forecast (e.g., types of El Niño and La Niña impacts, non-ENSO events, what are "normal" conditions?).
d. Non-weather/climate factors are deemed to be more important (e.g., uncertainty in other arenas, such as freshwater and ocean ecology [for salmon productivity]).
e. Low importance is given to climate forecast information because its role is unclear or impacts are not perceived as important enough to commit resources.
f. Other constraints deny a flexible response to the information (e.g., meeting flood control or Endangered Species Act requirements).
g. Procedures for acquiring knowledge and making and implementing decisions which incorporate climate information, have not been clearly defined.
h. Events forecast may be too far in the future for a discrete action to be engaged.
i. Availability and use of locally-specific information may be more relevant to a particular decision.
j. "Value" may not have been demonstrated by a credible reliable organization or competitor.
k. Desired information not provided (e.g., number of warm days, regional detail).
l. There may be competing forecasts or other conflicting information.
m. Lack of "tracking" information; does the forecast appear to be verifying?
n. History of previous forecasts not available. Validation statistics of previous forecasts not available.

[1] The *water year*, or hydrologic year, is October 1st through September 30th. This reflects the natural cycle in many hydrologic parameters such as the seasonal cycle of evaporative demand, and of the snow accumulation, melt, and runoff periods in many parts of the United States.

forecasts in water resources decision making. It is almost impossible to discuss the perceived value of forecasts without also discussing issues related to forecast skill. Many different criteria have been used to evaluate forecast skill (see Wilks, 1995 for a comprehensive review). Some measures focus on aspects of deterministic skill (*e.g.*, correlations between predicted and observed seasonally averaged precipitation anomalies), while many others are based on categorical forecasts (*e.g.*, Heidke skill scores for categorical forecasts of "wet", "dry", or "normal" conditions). The most important measures of skill vary with different perspectives. For example, Hartmann *et al.* (2002) argue that forecast performance criteria based on "hitting" or "missing" associated observations offer users conceptually easy entry into discussions of forecast quality. In contrast, some research scientists and water supply forecasters may be more interested in correlations between the ensemble average of predictions and observed measures of water supply like seasonal runoff volume.

Forecast skill remains a primary concern among many forecast producers and users. Skill in hydrologic forecast systems derives from various sources, including the quality of the simulation models used in forecasting, the ability to estimate the initial hydrologic state of the system, and the ability to skillfully predict the statistics of future weather over the course of the forecast period. Despite the significant resources expended to improve SI climate forecasts over the past 15 years, few water-resource related agencies have been making quantitative use of climate forecast information in their water supply forecasting efforts (Pulwarty and Redmond 1997; Callahan *et al.*, 1999).

In Section 2.2 of this Chapter, we review hydrologic data and forecasts products. Section 2.3 provides a parallel discussion of the climate



BOX 2.1: Agency Support

Federal support for research supporting improved hydrologic forecasts and applications through the use of climate forecasts and data has received increasing emphasis since the mid-1990s. The World Climate Research Program's Global Energy and Water Cycle Experiment (GEWEX) was among the first attempts to integrate hydrology/land surface and atmosphere models in the context of trying to improve hydrologic and climate predictability.

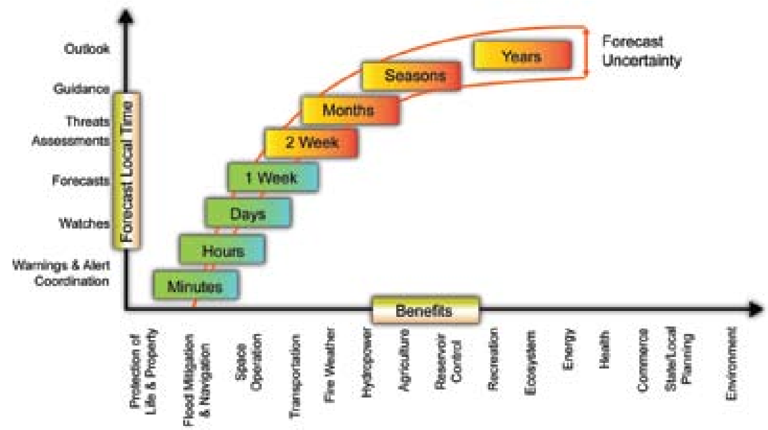
There have been two motivations behind this research: understanding scientific issues of land surface interactions with the climate system, and the development or enhancement of forecast applications, *e.g.*, for water, energy and hazard management. Early on, these efforts were dominated by the atmospheric (and related geophysical) sciences.

In the past, only a few U.S. programs have been very relevant to hydrologic prediction: the NOAA Climate Prediction Program for the Americas (CPPA), NOAA predecessors GEWEX Continental-scale International Project (GCIP), GEWEX Americas Prediction Project (GAPP) and the NASA Terrestrial Hydrology Program. The hydrologic prediction and water management focus of NOAA and NASA has slowly expanded over time. Presently, the NOAA Climate Dynamics and Experimental Prediction (CDEP), Transition of Research Applications to Climate Services (TRACS) and Sectoral Applications Research Program (SARP) programs, and the Water Management program within NASA, have put a strong emphasis on the development of both techniques and community linkages for migrating scientific advances in climate and hydrologic prediction into applications by agencies and end use sectors. The longer-standing NOAA Regional Integrated Sciences and Assessments (RISA) program has also contributed to improved use and understanding of climate data and forecast products in water resources forecasting and decision making. Likewise, the recently initiated postdoctoral fellowship program under the Predictability, Predictions, and Applications Interface (PPAI) panel of U.S. CLIVAR aims to grow the pool of scientists qualified to transfer advances in climate science and climate prediction into climate-related decision frameworks and decision tools.

Still, these programs are small in comparison with current federally funded science focused initiatives and are only just beginning to make inroads into the vast arena of effectively increasing the use and utility of climate and hydrologic data and forecast products.

data and forecast products that support hydrologic and water supply forecasting efforts in the United States. In Section 2.4, we provide a more detailed discussion of pathways for improving the skill and utility in hydrologic and climate forecasts and data products.

Section 2.5 contains a brief review of operational considerations and efforts to improve the utility of forecast and data products through efforts to improve the forecast evaluation and development process. These efforts include cases in which forecast providers and users have been engaged in sustained interactions to improve the use and utility of forecast and data products, and have led to many improvements and innovations in the data and forecast products generated by national centers. In recent years, a small number of water resource agencies have also developed end-to-end forecasting systems (*i.e.* forecasting systems that integrate observations and forecast models with decision-support tools) that utilize climate forecasts to directly inform hydrologic and water resources forecasts.



3TIER, Inc./Dept of Civil and Environmental Engineering, Univ. of Washington

Figure 2.1 The correspondence of climate and hydrologic forecast lead time to user sectors in which forecast benefits are realized (from National Weather Service Hydrology Research Laboratory). The focus of this Product is on climate and hydrologic forecasts with lead times greater than two weeks and up to approximately one year.

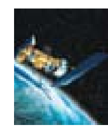
2.2 HYDROLOGIC AND WATER RESOURCES: MONITORING AND PREDICTION

The uses of hydrologic monitoring and prediction products, and specifically those that are relevant for water, hazard and energy management, vary depending on the forecast lead time (Figure 2.1). The shortest climate and hydrologic lead-time forecasts, from minutes to hours, are applied to such uses as warnings for floods and extreme weather, wind power scheduling, aviation, recreation, and wild fire response management. In contrast, at lead times of years to decades, predictions are used for strategic planning purposes rather than operational management of resources. At SI lead times, climate and hydrologic forecast applications span a wide range that includes the management of water, fisheries, hydropower and agricultural production, navigation and recreation. Table 2.2 lists aspects of forecast products at these time scales that are relevant to decision makers.

2.2.1 Prediction Approaches

The primary climate and hydrologic prediction approaches used by operational and research centers fall into four categories: statistical, dynamical, statistical-dynamical hybrid, and consensus. The first three approaches are objective in the sense that the inputs and methods are formalized, outputs are not modified on an *ad hoc* basis, and the resulting forecasts are potentially reproducible by an independent forecaster using the same inputs and methods. The fourth major category of approach, which might also be termed blended knowledge, requires subjective weighting of results from the other approaches. These types of approaches are discussed in Box 2.2.

Other aspects of dynamical prediction schemes related to model physical and computational structure are important in distinguishing one model or model version from another. These aspects are primary indicators of the sophistication of an evolving model, relative to other models, but are not of much interest to the forecast user community. Examples include the degree of coupling of model components, model vertical resolution, cloud microphysics package, nature of data assimilation approaches and of the data assimilated, and the ensemble generation scheme, among many other forecast system features.



Climate and hydrologic lead-time forecasts range from minutes to years.

2.2.2 Forecast Producers and Products

Federal, regional, state, and local agencies, as well as private sector companies, such as utilities, produce hydrologic forecasts. In contrast to climate forecasts, hydrologic forecast products more directly target end use sectors—*e.g.*, water, energy, natural resource or hazard management—and are often region-specific. Prediction methods and forecast products vary from region to region and are governed by many factors, but depend in no small measure on the hydroclimatology, institutional traditions and sectoral concerns in each region. A representative sampling of typical forecast producers and products is given in Appendix A.1. Forecasting activities at the federal, state, regional, and local scales are discussed in the following subsections.

2.2.2.1 Federal

The primary federal streamflow forecasting agencies at SI lead times are the NOAA, National Weather Service (NWS) and the U.S. Department of Agriculture (USDA) National Resource Conservation Service (NRCS) National Water and Climate Center (NWCC). The NWCC's four forecasters produce statistical forecasts of summer runoff volume in the western United States using multiple linear regression to estimate future streamflow from current observed snow water equivalent, accumulated water year precipitation, streamflow, and in some locations, using ENSO indicators such as the Niño3.4 index (Garen, 1992; Pagano and Garen, 2005). Snowmelt runoff is critical for a wide variety of uses (water supply, irrigation, navigation, recreation, hydropower, environmental flows) in the relatively dry summer season. The regression approach has been central to the NRCS since the mid-1930s, before which similar snow-survey based forecasting was conducted by a number of smaller groups. Forecasts are available to users both in the form of tabular summaries (Figure 2.2) that convey the central tendency of the forecasts and estimates of uncertainty, and maps showing the median forecast anomaly for each river basin area for which the forecasts are operational

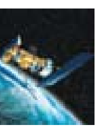


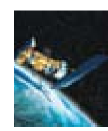
Table 2.2 Aspects of forecast products that are relevant to users.

Forecast Product Aspect	Description / Example
Forecast product variables	Precipitation, temperature, humidity, wind speed, and atmospheric pressure
Forecast product spatial resolution	Grid cell longitude by latitude, climate division
Domain	Watershed, river basin, regional, national, and global
Product time step (temporal resolution)	Hourly, sub-daily, daily, monthly, and seasonal
Range of product lead times	1 to 15 days, 1 to 13 months
Frequency of forecast product update	Every 12 hours, every month
Lag of forecast product update	The length of time from the forecast initialization time before forecast products are available: *e.g.*, two hours for a medium range forecast, one day for a monthly to seasonal forecast.
Existence of historical climatology	Many users require a historical climatology showing forecast model performance to use in bias-correction, downscaling, and/or verification.
Deterministic or probabilistic	Deterministic forecasts have a single prediction for each future lead time. Probabilistic forecasts frame predicted values within a range of uncertainty, and consist either of an ensemble of forecast sequences spanning all lead times, or of a distinct forecast distribution for each future lead time.
Availability of skill/accuracy information	Published or otherwise available information about the performance of forecasts is not always available, particularly for forecasts that are steadily evolving. In principle, the spread of probabilistic forecasts contains such information about the median of the forecast; but the skill characteristics pertaining to the spread of the forecast are not usually available.

BOX 2.2: Forecast Approaches

Dynamical: Computer models designed to represent the physical features of the oceans, atmosphere and land surface, at least to the extent possible given computational constraints, form the basis for dynamical predictions. These models have, at their core, a set of physical relationships describing the interactions of the Earth's energy and moisture states. Inputs to the models include estimates of the current moisture and energy conditions needed to initialize the state variables of the model (such as the moisture content of an atmospheric or soil layer), and of any physical characteristics (called parameters—one example is the elevation of the land surface) that must be known to implement the relationships in the model's physical core. In theory, the main advantage of dynamical models is that influence of any one model variable on another is guided by the laws of nature as we understand them. As a result, the model will correctly simulate the behavior of the earth system even under conditions that may not have occurred in the period during which the model is verified, calibrated and validated. The primary disadvantages of dynamical models, however, are that their high computational and data input demands require them to approximate characteristics of the Earth system in ways that may compromise their realism and therefore performance. For example, the finest computational grid resolution that can be practically achieved in most atmospheric models (on the order of 100 to 200 kilometers per cell) is still too coarse to support a realistic representation of orographic effects on surface temperature and precipitation. Dynamical hydrologic models can be implemented at much finer resolutions (down to ten meters per cell, for catchment-scale models) because they are typically applied to much smaller geographic domains than are atmospheric models. While there are many aspects that distinguish one model from another, only a subset of those (listed in Table 1.1) is appreciated by the forecast user, as opposed to the climate modeler, and is relevant in describing the dynamical forecast products.

Statistical: Statistical forecast models use mathematical models to relate observations of an earth system variable that is to be predicted to observations of one or more other variables (and/or of the same variable at a prior time) that serve as predictors. The variables may describe conditions at a point location (e.g., flow along one reach of a river) or over a large domain, such as sea surface temperatures along the equator. The mathematical models are commonly linear relationships between the predictors and the predictand, but also may be formulated as more complex non-linear systems.



Statistical models are often preferred for their computational ease relative to dynamical models. In many cases, statistical models can give equal or better performance to dynamical models due in part to the inability of dynamical models to represent fully the physics of the system (often as a result of scale or data limitations), and in part to the dependence of predictability in many systems on predominantly linear dynamics (Penland and Magorian, 1993; van den Dool, 2007). The oft-cited shortcomings of statistical models, on the other hand, include their lack of representation of physical causes and effects, which, in theory, compromise their ability to respond to unprecedented events in a fashion that is consistent with the physical constraints of the system. In addition, statistical models may require a longer observational record for "training" than dynamical models, which are helped by their physical structure.

Objective hybrids: Statistical and dynamical tools can be combined using objective approaches. A primary example is a weighted merging of the tools' separate predictions into a single prediction (termed an objective consolidation; van den Dool, 2007). A second example is a tool that has dynamical and statistical subcomponents, such as a climate prediction model that links a dynamical ocean submodel to a statistical atmospheric model. A distinguishing feature of these hybrid approaches is that an objective method exists for linking the statistical and dynamical schemes so as to produce a set of outputs that are regarded as "optimal" relative to the prediction goals. This objectivity is not preserved in the next consensus approach.

Blended Knowledge or Subjective consensus: Some forecast centers release operational predictions, in which expert judgment is subjectively applied to modify or combine outputs from prediction approaches of one or more of the first three types, thereby correcting for perceived errors in the objective approaches to form a prediction that has skill superior to what can be achieved by objective methods alone. The process by which the NOAA Climate Predication Center (CPC) and International Research Institute for Climate and Society (IRI) constructs their monthly and seasonal outlooks for example, includes subjective weighting of the guidance provided by different climate forecast tools. The weighting is often highly sensitive to recent evolution and current state of the tropical ENSO, but other factors, like decadal trends in precipitation and surface temperature, also have the potential to influence the final official climate forecasts.

Streamflow Forecasts as of June 1, 2008

Stream and Station	Forecast Period	Forecasts This Year Most Probable kaf	%avg	Reasonable Max %avg	Min &avg	30 Year '71–'00 Average Runoff kaf
Arkansas River Basin						
Arkansas River Granite at,CO	Apr-Sep	260	124	177	118	210
Salida at, CO	Apr-Sep	450	145	177	118	310
Canon City at, CO	Apr-Sep	540	136	172	111	397
Pueblo abv, CO	Apr-Sep	650	134	167	105	485
Grape Creek West-cliffe nr, CO	Apr-Sep	33.0	168	245	107	19.6
Cucharas River La Veta nr, CO	Apr-Sep	11.1	85	108	68	13.0
Purgatoire River-Trinidad at, CO	Apr-Sep	32.0	73	107	48	44
Huerfano River Redwing nr, CO	Apr-Sep	12.8	83	103	65	15.5
Chalk Creek Nathrop nr, CO	Apr-Sep	43.0	159	211	115	27
Vermejo River Dawson nr, NM	Mar-Jun	6.20	89	113	73	7.0
Eagle Nest Reservoir Reser-voir Inflow, NM	Mar-Jun	14.70	126	143	118	11.7
Cimarron River Cimarron nr, NM	Mar-Jun	18.60	117	138	106	15.9
Ponil Creek Cimarron nr, NM	Mar-Jun	6.10	91	109	81	6.7
Rayado Creek Sauble Ranch, NM	Mar-Jun	5.90	83	101	73	7.1

Figure 2.2 Example of NRCS tabular summer runoff (streamflow) volume forecast summary, showing median ("most probable") forecasts and probabilistic confidence intervals, as well as climatological flow averages. Flow units are thousand-acre-feet (KAF), a runoff volume for the forecast period. This table was downloaded from <http://www.wcc.nrcs.usda.gov/wsf/wsf.html>.

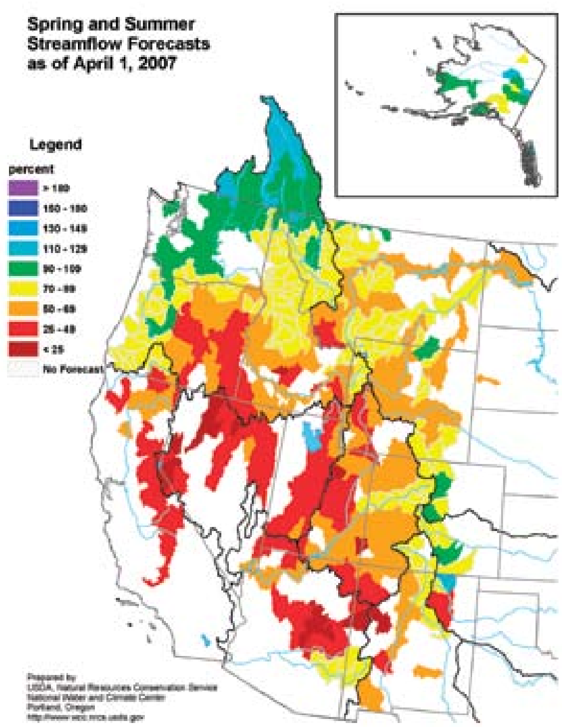


Figure 2.3 Example of NRCS spatial summer runoff (April-September streamflow) volume forecast summary, showing median runoff forecasts as an anomaly (percent of average).

(Figure 2.3). Until 2006, the NWCC's forecasts were released near the first of each month, for summer flow periods such as April through July or April through September. In 2006, the NWCC began to develop automated daily updates to these forecasts, and the daily product is likely to become more prevalent as development and testing matures. The NWCC has also just begun to explore the use of physically-based hydrologic models as a basis for forecasting.

NWCC water supply forecasts are coordinated subjectively with a parallel set of forecasts produced by the western U.S. NWS River Forecast Centers (RFCs), and with forecasts from Environment Canada's BC Hydro. The NRCS-NWS joint, official forecasts are of the subjective consensus type described earlier, so the final forecast products are subjective combinations of information from different sources, in this case, objective statistical tools (*i.e.*, regression models informed by observed snow water equivalent, accumulated water year precipitation, and streamflow) and model based forecast results from the RFCs.

The NWS surface water supply forecast program began in the 1940s in the Colorado Basin. It has since expanded to include seasonal forecasts (of volume runoff during the spring to summer snow melt period) for most of the snowmelt-dominated basins important to water management in the western United States. These forecasts rely on two primary tools: Statistical Water Supply (SWS), based on multiple-linear regression, and Ensemble Streamflow Prediction (ESP), a technique based on hydrologic modeling (Schaake, 1978; Day, 1985). Results from both approaches are augmented by forecaster experience and the coordination process with other forecasting entities. In contrast to the western RFCs, RFCs in the eastern United States are more centrally concerned with short to medium-range flood risk and drought-related water availability out to about a three month lead time. At some eastern RFC websites, the seasonal forecast is linked only to the CPC Drought Outlook rather than an RFC-generated product (Box 2.3).

The streamflow prediction services of the RFCs have a national presence, and, as such, are able to leverage a number of common technological elements, including models, databases and software for handling meteorological and hydrological data, and for making, assessing and disseminating forecasts (*i.e.*, website structure). Nonetheless, the RFCs themselves are regional entities with regional concerns.

The NWS's ESP approach warrants further discussion. In the mid 1970s, the NWS developed the hydrologic modeling, forecasting and analysis system—NWS River Forecast System (NWSRFS)—the core of which is the Sacramento soil moisture accounting scheme coupled to the Snow-17 temperature index snow model, for ESP-based prediction (Anderson, 1972, 1973; Burnash *et al.*, 1973). The ESP approach uses a deterministic simulation of the hydrologic state during a model spin-up (initialization) period, leading up to the forecast start date to estimate current hydrologic conditions, and then uses an ensemble of historical meteorological sequences as model inputs (*e.g.*, temperature and precipitation) to simulate hydrology in the future (or forecast period). Until several years ago, the RFC dissemination of ESP-based forecasts for streamflows at SI lead times was rare, and the statistical forecasts were the accepted standard. Now, as part of the NWS Advanced Hydrologic Prediction Service (AHPS) initiative, ESP forecasts are being aggressively implemented for basins across the United States (Figure 2.4) at lead times from hours to SI (McEnery *et al.*, 2005).

At the seasonal lead times, several western RFCs use graphical forecast products for the summer period streamflow forecasts that convey the probabilistic uncertainty of the forecasts. A unified web based suite of applications that became operational in 2008 provides forecast users with a number of avenues for exploring the RFC water supply forecasts. For example, Figure 2.5 shows (in clockwise order from top left) (a) a western United States depiction of the median water supply outlook for the RFC forecast basins, (b) a progression of forecasts (median and bounds) during the water year together with flow normals and observed flows; (c) monthly forecast distributions, with the option to display individual forecast ensemble members (*i.e.*, single past years) and also select ENSO-based categorical forecasts (ESP subsets); and (d) various skill measures, such as mean absolute error, for the forecasts based on hindcast performance. Access to raw ensemble member data is also provided from the same website.

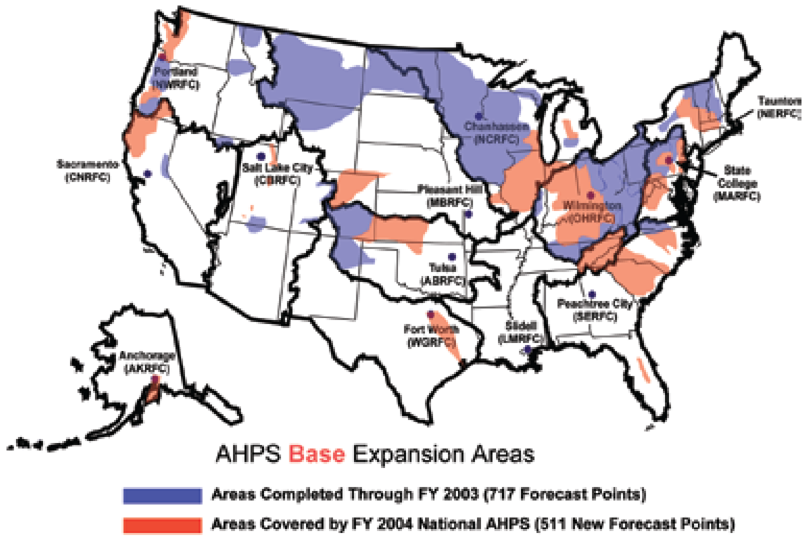


Figure 2.4 Areas covered by the NWS Advanced Hydrologic Prediction Service (AHPS) initiative (McEnery *et al.*, 2005).

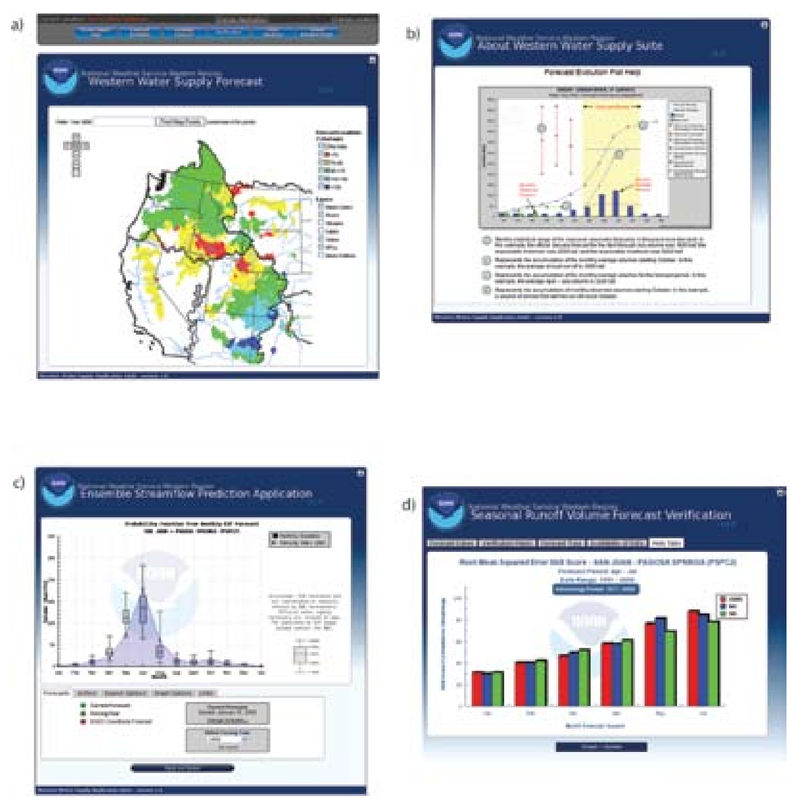
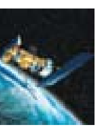


Figure 2.5 A graphical forecast product from the NWS River Forecast Centers, showing a forecast of summer (April through July) period streamflow on the Colorado River, Colorado to Arizona. These figures were obtained from <http://www.nwrfc.noaa.gov/westernwater>.

The existence in digitized form of the retrospective archive of seasonal forecasts is critical for the verification of forecast skill.



The provision of a service that assists hydrologic forecast users in either customizing a selection of ESP possibilities to reflect, perhaps, the users' interest in data from past years that they perceive as analogues to the current year, or the current ENSO state, is a notable advance from the use of "climatological" ESP (*i.e.*, using all traces from a historical period) in the prior ESP-related seasonal forecast products. Some western RFCs have also experimented with using the CPC seasonal climate outlooks as a basis for adjusting the precipitation and temperature inputs used in climatological ESP, but it was found that the CPC outlook anomalies were generally too small to produce a distinct forecast from the climatological ESP (Hartmann *et al.*, 2002). In some RFCs, NWS statistical water supply forecasts have also provided perspective (albeit more limited) on the effect of future climate assumptions on future runoff by including results from projecting 50, 75, 100, 125 and 150 percent of normal precipitation in the remaining water year. At times, the official NWS statistical forecasts have adopted such assumptions, *e.g.*, that the first month following the forecast date would contain other than 100 percent of expected precipitation, based on forecaster judgment and consideration of a range of factors, including ENSO state and CPC climate predictions.

Figure 2.6 shows the performance of summer streamflow volume forecasts from both the NWS and NRCS over a recent ten-year period; this example is also part of the suite of forecast products that the western RFC designed to improve the communication of forecast performance and provide verification information. Despite recent literature (Welles *et al.*, 2007) that has underscored a general scarcity of such information from hydrologic forecast providers, the NWS has recently codified verification approaches and developed verification tools, and is in the process of disbursing them throughout the RFC organization (NWS, 2006). The existence in digitized form of the retrospective archive of seasonal forecasts is critical for the verification of forecast skill. The ten-year record shown in Figure 2.6, which is longer than the record available (internally or to the public) for many public agency forecast variables, is of inadequate length for some types of statistical assessment, but is an undeniable advance in forecast communication relative to the services that were previously available. Future development priorities include a climate change scenario application, which would leverage climate change scenarios from IPCC or similar to produce inputs for future water supply planning exercises. In addition, forecast calibration procedures (*e.g.*, Seo *et al.*, 2006; Wood and Schaake, 2008) are being developed for the ensemble forecasts to remove forecast biases. The current NOAA/NWS web service Internet web address is: <http://www.nwrfc.noaa.gov/westernwater>

A contrast to these probabilistic forecasts is the deterministic five-week forecast of lake water level in Lake Lanier, GA, produced by the U.S. Army Corps of Engineers (USACE) based on probabilistic inflow forecasts from the NWS southeastern RFC. Given that the lake is a managed system and the forecast has

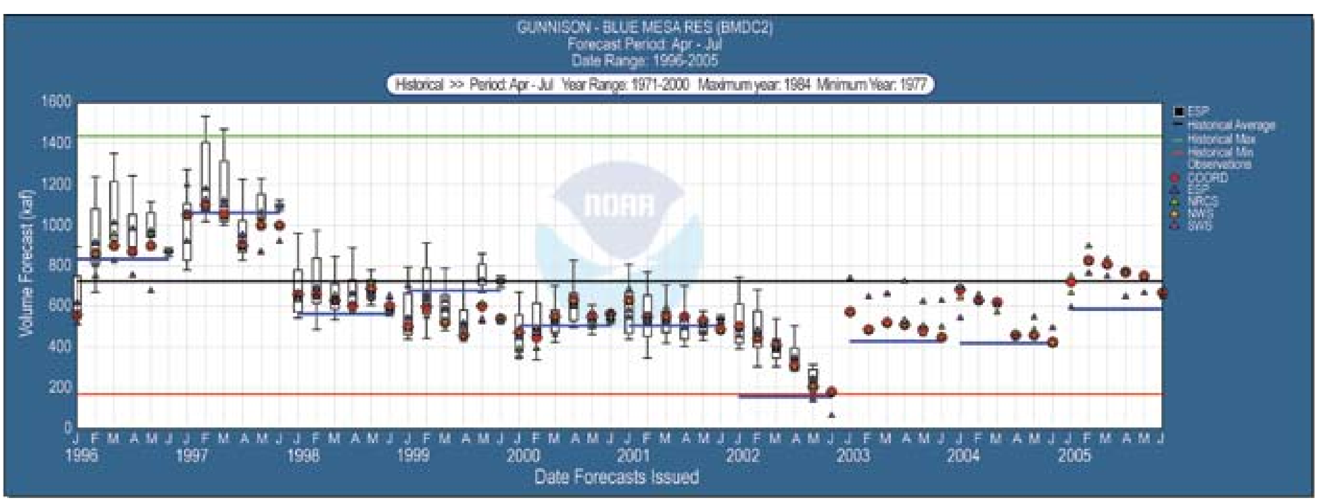


Figure 2.6 Comparing ESP and statistical forecasts from the NRCS and NWS for a recent 10-year period. The forecasts are for summer (April through July) period streamflow on the Gunnison River, Colorado.

a sub-seasonal lead time, the single-valued outlook may be justified by the planned management strategy. In such a case, the lake level is a constraint that requires transferring uncertainty in lake inflows to a different variable in the reservoir system, such as lake outflow. Alternatively, the deterministic depiction may result from an effort to simplify probabilistic information in the communication of the lake outlook to the public.

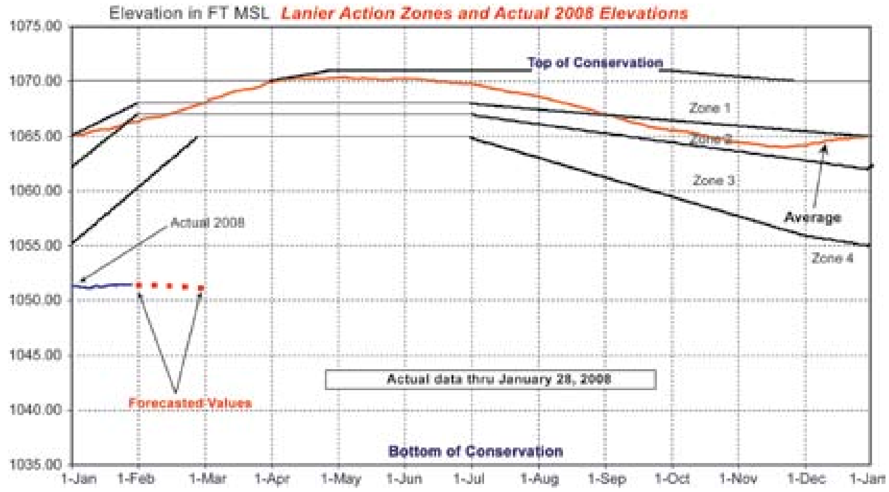


Figure 2.7 A deterministic five-week forecast of reservoir levels in Lake Lanier, Georgia, produced by USACE <http://water.sam.usace.army.mil/lanfc.htm>.

2.2.2.2 State and Regional

Regionally-focused agencies such as the U.S. Bureau of Reclamation (USBR), the Bonneville Power Administration (BPA), the Tennessee Valley Authority (TVA), and the Great Lakes Environmental Research Laboratory (GLERL) also produce forecasts targeting specific sectors within their priority areas. Figure 2.8 shows an example of an SI lead forecast of lake levels produced by GLERL. GLERL was among the first major public agencies to incorporate climate forecast information into operational forecasts using hydrologic and water management variables. Forecasters use coarse-scale climate forecast information to adjust climatological probability distribution functions (PDFs) of precipitation and temperature that are the basis for generating synthetic ensemble inputs to hydrologic and water management models, the outputs of which include lake level as shown in the figure. In this case, the climate forecast information is from the CPC seasonal outlooks (method described in Croley, 1996).

The Bonneville Power Administration (BPA), which helps manage and market power from the Columbia River reservoir system, is both a consumer and producer of hydrologic forecast products. The BPA generates their own ENSO-state conditioned ESP forecasts of reservoir system inflows as input to management decisions, a practice supported by research into the benefits of ENSO information for water management (Hamlet and Lettenmaier, 1999).

A number of state agencies responsible for releasing hydrologic and water resources forecasts also make use of climate forecasts in the process of producing their own hydrologic forecasts. The South Florida Water Management District (SFWMD) predicts lake (*e.g.*, Lake Okeechobee) and canal stages, and makes drought assessments, using a decision tree in which the CPC seasonal outlooks play a role. SFWMD follows GLERL's lead in using the Croley (1996) method for translating the CPC seasonal outlooks to variables of interest for their system.

2.2.2.3 Local

At an even smaller scale, some local agencies and private utilities may also produce forecasts or at least derive applications-targeted forecasts from the more general climate or hydrology forecasts generated at larger agencies or centers.

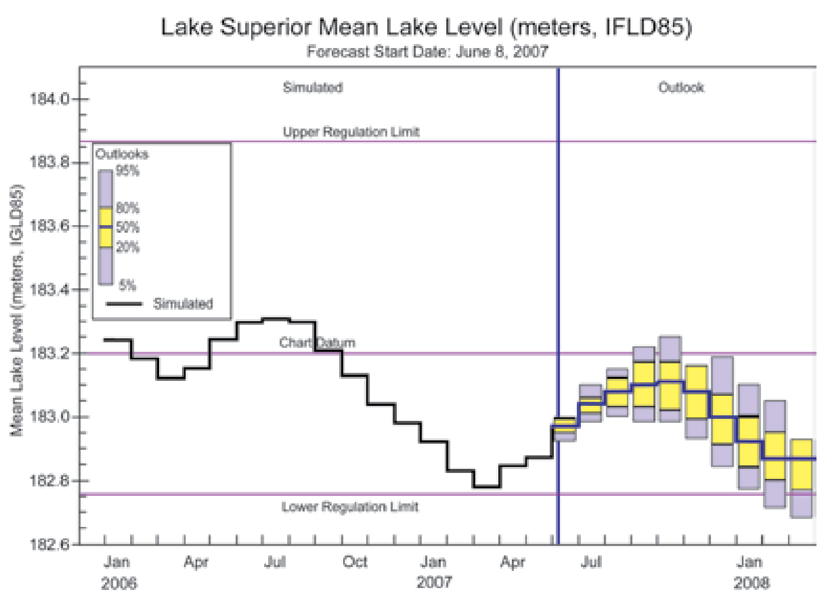


Figure 2.8 Probabilistic forecasts of future lake levels disseminated by GLERL. From: <http://www.glerl.noaa.gov/wr/ahps/curfcst/>.

Seattle Public Utilities (SPU; see Experiment 4, Section 4.2.1), for example, operates a number of reservoirs for use primarily in municipal water supply. SPU makes SI reservoir inflow forecasts using statistical methods based on observed conditions in their watersheds (*i.e.*, snow and accumulated precipitation), and on the current ENSO state, in addition to consulting the Northwest River Forecast Center (NWRFC) volume runoff forecasts. The SPU forecasts are made and used internally rather than disseminated to the public.

2.2.2.4 Research

Research institutions such as universities also produce hydrologic forecasts of a more experimental nature. A prime example is the Integrated Forecast and Reservoir Management (INFORM) project housed at the Hydrologic Research Center (HRC), which produces not only streamflow forecasts in the State of California, but also reservoir system forecasts. This project is discussed at greater length in Chapter 4 (Georgakakos *et al.*, 2005). Approximately five years ago, researchers at the University of Washington and Princeton University launched an effort to produce operational hydrologic and streamflow predictions using distributed land surface models that were developed by an interagency effort called the Land Data Assimilation System (LDAS) project (Mitchell *et al.*, 2004). In addition to generating SI streamflow forecasts in the western and eastern United States, the project also generates real-time forecasts for land surface variables such as runoff, soil moisture, and snow water equivalent (Wood and Lettenmaier, 2006; Luo and Wood, 2008), some of which are used in federal drought monitoring and prediction activities (Wood, 2008; Luo and Wood, 2007). Figure 2.9 shows an example (a runoff forecast) from this body of work that is based on the use of the Climate Forecast System (CFS) and CPC climate outlooks. Similar to the NWS ESP predictions, these hydrologic and streamflow forecasts are physically-based, dynamical and objective. The effort is supported primarily by NOAA, and like the INFORM project collaborates with public forecast agencies in developing research-level prediction products. The federal funding is provided with the intent of migrating operational forecasting advances that arise in the course of these efforts into the public agencies, a topic discussed briefly in Section 2.1.

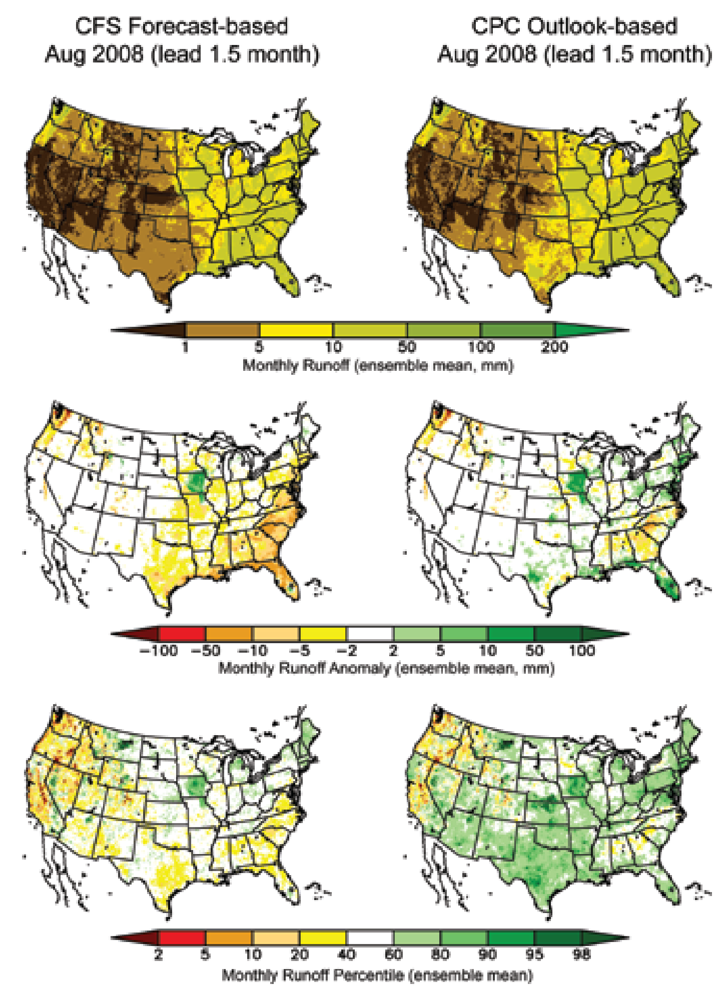


Figure 2.9 Ensemble mean forecasts of monthly runoff at lead 1.5 months created using an LDAS hydrologic model driven by CFS and CPS climate outlooks. The hydrologic prediction techniques were developed at the University of Washington and Princeton University as part of a real-time streamflow forecasting project sponsored by NOAA. Other variables, not shown, include soil moisture, snow water equivalent, and streamflow. This map is based on those available from <http://hydrology.princeton.edu/~luo/research/FORECAST/forecast.php>.

2.2.3 Skill in Seasonal-to-Interannual Hydrologic and Water Resource Forecasts

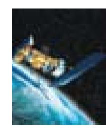
This Section focuses on the skill of hydrologic forecasts; Section 2.5 includes a discussion of forecast utility. Forecasts are statements about events expected to occur at specific times and places in the future. They can be either deterministic, single-valued predictions about specific outcomes, or probabilistic descriptions of likely outcomes that typically take the form of ensembles, distributions, or weighted scenarios.

The hydrologic and water resources forecasts made for water resources management reflect three components of predictability: the seasonality of the hydrologic cycle, the predictability associated with large-scale climate teleconnections, and the persistence of anomalies in hydrologic initial conditions. Evapotranspiration, runoff (*e.g.*, Pagano *et al.*, 2004) and ground-water recharge (*e.g.*, Earman *et al.*, 2006) all depend on soil moisture and (where relevant) snowpack conditions one or two seasons prior to the forecast windows, so that these moisture conditions, directly or indirectly, are key predictors to many hydrologic forecasts with lead times up to six months. Although hydrologic initial conditions impart only a few months of predictability to hydrologic systems, during their peak months of predictability, the skill that they contribute is often paramount. This is particularly true in the western United States, where much of the year's precipitation falls during the cool season, as snow, and then accumulates in relatively easily observed form, as snowpack, until it predictably melts and runs off in the warm season months later. Information about large-scale climatic influences, like the current and projected state of ENSO, are valued because some of the predictability that they confer on water resources has influence even before snow begins to accumulate or soil-recharging fall storms arrive. ENSO, in particular, is strongly synchronized with the annual cycle so that, in many instances, the first signs of an impending warm (El Niño) or cold (La Niña) ENSO event may be discerned toward the end of the summer before the fluctuation reaches its maturity and peak of influence on the United States climate in winter. This advance warning for important aspects of water year climate allows forecasters in some locations to incorporate the expected ENSO influences into hydrologic forecasts before or near the beginning of the water year (*e.g.*, Hamlet and Lettenmaier, 1999).

These large-scale climatic influences, however, rarely provide the high level of skill that can commonly be derived later in the water year from estimates of land surface moisture state, *i.e.*, from precipitation accumulated during the water year, snow water equivalent or soil moisture, as estimated indirectly from streamflow. Finally, the unpredictable, random component of variability remains to limit the skill of all real-world forecasts. The unpredictable component reflects a mix of uncertainties and errors in the observations used to initialize forecast models, errors in the models, and the chaotic complexities in forecast model dynamics and in the real world.

Many studies have shown that the single greatest source of forecast error is unknown precipitation after the forecast issue date. Schaake and Peck (1985) estimate that for the 1947 to 1984 forecasts for inflow to Lake Powell, almost 80 percent of the January 1st forecast error is due to unknown future precipitation; by April 1st, Schaake and Peck find that future precipitation still accounts for 50 percent of the forecast error. Forecasts for a specific area can perform poorly during years with abnormally high spring precipitation or they can perform poorly if the spring precipitation in that region is normally a significant component of the annual cycle. For example, in California, the bulk of the moisture falls from January to March and it rarely rains in spring (April to June), meaning that snowpack-based April 1st forecasts of spring-summer streamflow are generally very accurate. In comparison (see Figure 2.10), in eastern Wyoming and the Front Range of Colorado, April through June is the wettest time of year and, by April 1st, the forecaster can only guess at future precipitation events because of an inability to skillfully forecast springtime precipitation in this region one season in advance.

Forecasts made for water resources management reflect three components of predictability: the seasonality of the hydrologic cycle, the predictability associated with large-scale climate teleconnections, and the persistence of anomalies in hydrologic initial conditions.



Pagano *et al.* (2004) determined that the second greatest factor influencing forecasting skill is how much influence snowmelt has on the hydrology of the basin and how warm the basin is during the winter. For example, in basins high in the mountains of Colorado, the temperature remains below freezing for most of the winter. Streamflow is generally low through April until temperatures rise and the snow starts to melt. The stream then receives a major pulse of snowmelt over the course of several weeks. Spring precipitation may supplement the streamflow, but any snow that falls in January is likely to remain in the basin until April when the forecast target season starts. In comparison, in western Oregon, warm rain-producing storms can be interspersed with snow-producing winter storms. Most of the runoff occurs during the winter and it is possible for a large snowpack in February to be melted and washed away by March rains. For the forecaster, predicting April-to-July streamflow is difficult, particularly in anticipating the quantity of water that is going to "escape" before the target season begins. Additional forecast errors in snowmelt river basins can arise from the inability to accurately predict the sublimation of snow (sublimation occurs when ice or snow converts directly into atmospheric water vapor without first passing through the liquid state), a complex process that is influenced by cloudiness, sequences of meteorological conditions (wind, relative humidity as well as temperature) affecting crust, internal snow dynamics, and vegetation.

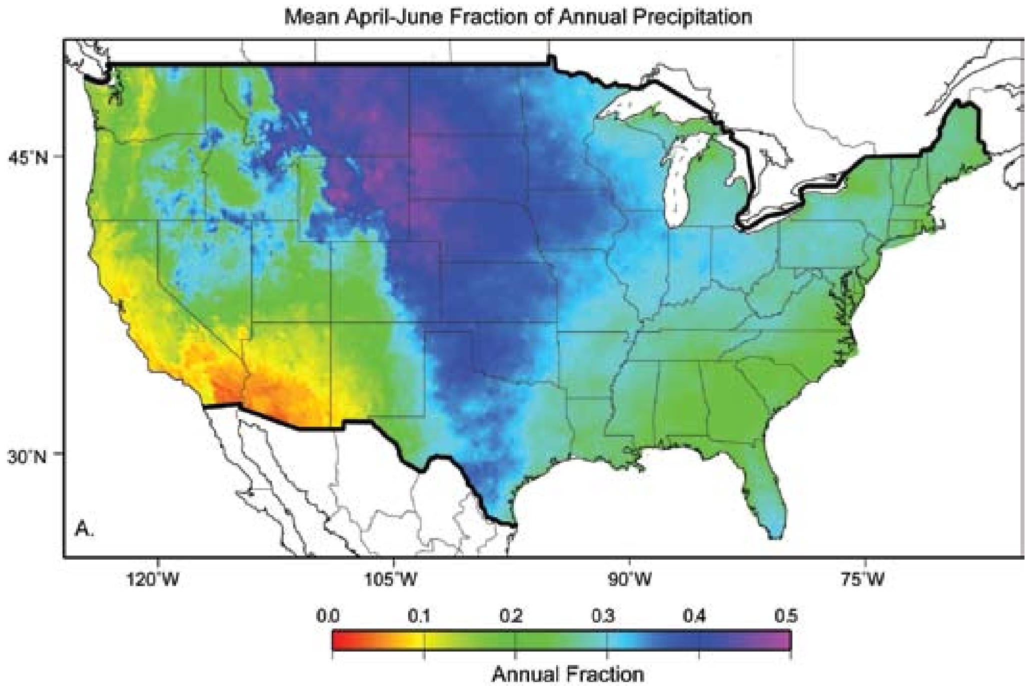


Figure 2.10 Mean percentages of annual precipitation that fell from April through June, 1971 to 2000 (based on 4-km PRISM climatologies). This figure was obtained from <http://www.prism.oregonstate.edu/>.

Some element of forecast accuracy depends on the variability of the river itself. It would be easy to incur a 100 percent forecast error on, for example, the San Francisco River in Arizona, whose observations vary between 17 percent to more than 750 percent of average. It would be much more difficult to incur such a high error on a river such as the Stehekin River in Washington, where the streamflow ranges only between 60 percent and 150 percent of average. A user may be interested in this aspect of accuracy (*e.g.*, percent of normal error), but most forecasters use skill scores (*e.g.*, correlation) that would normalize for this effect and make the results from these two basins more comparable. As noted by Hartmann *et al.* (2002), consumers of forecast information may be more interested in measures of forecast skill other than correlations.

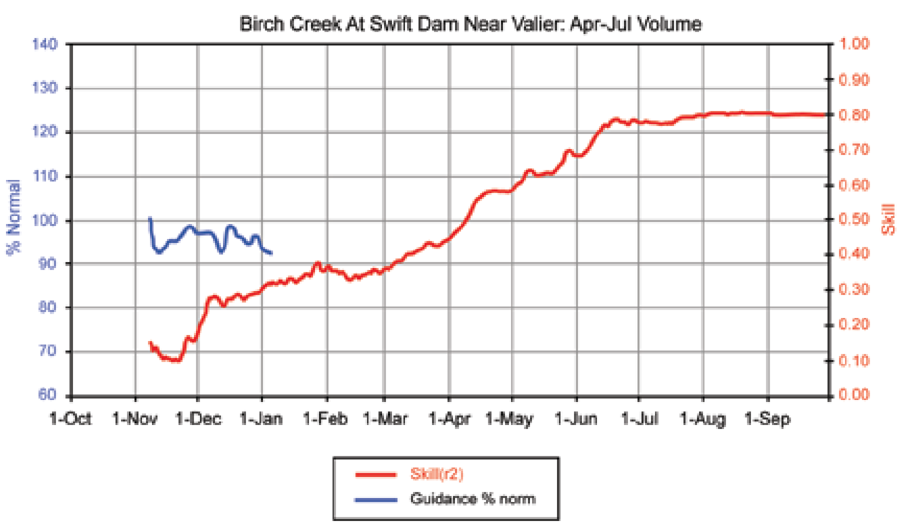
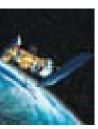


Figure 2.11 Recent operational National Water and Climate Center (NWCC) forecasts of April-July 2007 streamflow volume in Birch Creek at Swift Dam near Valier, Montana, showing daily median-forecast values of percentages of long-term average streamflow total for summer 2007 (blue) and the long-term estimates of correlation-based forecast skill corresponding to each day of the year. Figure obtained from the NWCC <http://www.wcc.nrcs.usda.gov/>.



2.2.3.1 Skill of Current Seasonal Hydrologic and Water-Supply Forecasts

As previously indicated, hydrologic and streamflow forecasts that extend to a nine-month lead time are made for western United States rivers, primarily during the winter and spring, whereas in other parts of the United States, where seasonality of precipitation is less pronounced, the forecasts link to CPC drought products, or are qualitative (the NWS Southeastern RFC, for instance, provides water supply related briefings from their website), or are in other regards less amenable to skill evaluation. For this reason, the following discussion of water supply forecast skill focuses mostly on western United States streamflow forecasting, and in particular water supply (*i.e.*, runoff volume) forecasts, for which most published material relating to SI forecasts exists.

In the western United States, the skill of operational forecasts generally improves progressively during the winter and spring months leading up to the period being forecasted, as increasing information about the year's land surface water budget are observable (*i.e.*, reflected in snowpack, soil moisture, streamflow and the like). An example of the long-term average seasonal evolution of NWCC operational forecast skill at a particular stream gage in Montana is shown in Figure 2.11. The flow rates that are judged to have a 50 percent chance of not being exceeded (*i.e.*, the 50th percentile or median) are shown by the blue curve for the early part of 2007. The red curve shows that, early in the water year, the April to July forecast has little skill, measured by the regression coefficient of determination (r^2, or correlation squared), with only about ten percent of historical variance captured by the forecast equations. By about April 1st, the forecast equations predict about 45 percent of the historical variance, and at the end of the season, the variance explained is about 80 percent. This measure of skill does not reach 100 percent because the observations available for use as predictors do not fully explain the observed hydrologic variation.

Comparisons of "hindcasts"—seasonal flow estimates generated by applying the operational forecast equations to a few decades (lengths of records differ from site to site) of historical input variables at each location with observed flows provide estimates of the expected skill of current operational forecasts. The actual skill of the forecast equations that are operationally used at as many as 226 western stream gages are illustrated in Figure 2.12, in which skill is measured by correlation of hindcast median with observed values.

The symbols in the various panels of Figure 2.12 become larger and bluer in hue as the hindcast dates approach the start of the April to July seasons being forecasted. They begin with largely unskillful beginnings each year in the January 1st forecast; by April 1st the forecasts are highly skillful by the correlation measures (predicting as much as 80 percent of the year-to-year fluctuations) for most of the California, Nevada, and Idaho rivers, and many stations in Utah and Colorado.

The general increases in skill and thus in numbers of stations with high (correlation) skill scores as the April 1st start of the forecast period approaches is shown in Figure 2.13.

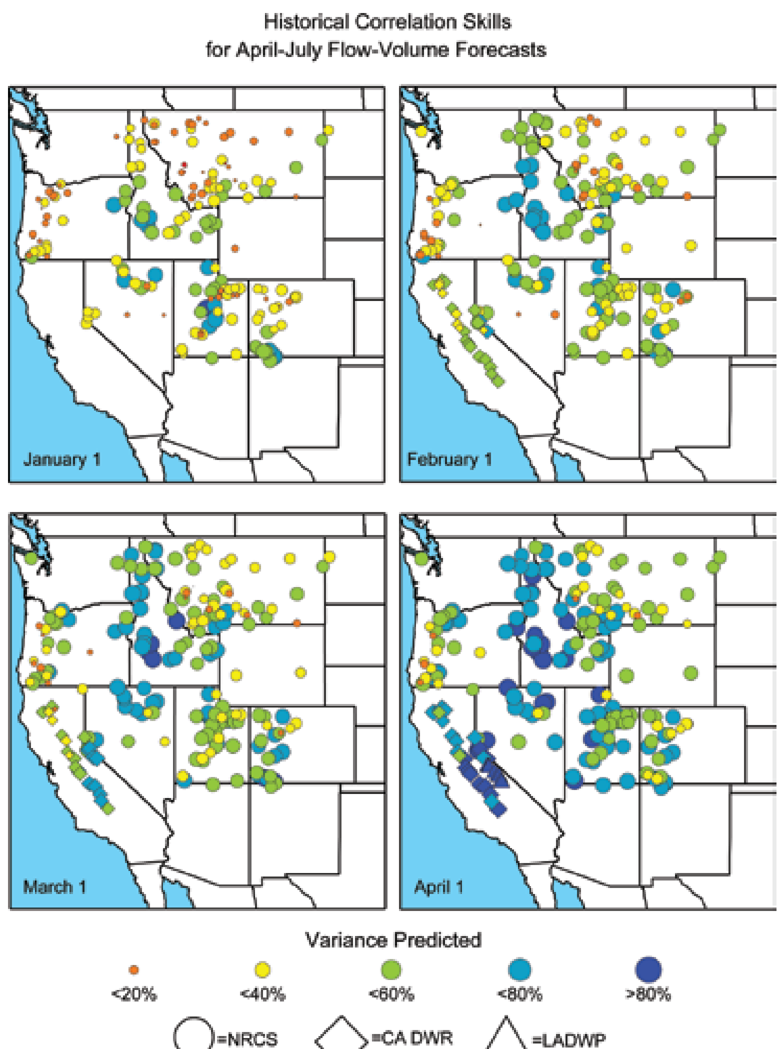
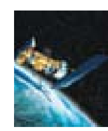


Figure 2.12 Skills of forecast equations used operationally by NRCS, California Department of Water Resources, and Los Angeles Department of Water and Power, for predicting April to July water supplies (streamflow volumes) on selected western rivers, as measured by correlations between observed and hindcasted flow totals over each station's period of forecast records. Figure provided by Tom Pagano, USDA NRCS.



A question not addressed in this Product relates to the probabilistic skill of the forecasts: How reliable are the confidence limits around the median forecasts that are provided by the published forecast quantiles (10th and 90th percentiles, for example)? In a reliable forecast, the frequencies with which the observations fall between various sets of confidence bounds matches the probability interval set by those bounds. That is, 80 percent of the time, the observed values fall between the 10th and 90th percentiles of the forecast. Among the few analyses that have been published focusing on the probabilistic performance of United States operational streamflow forecasts, Franz *et al.* (2003) evaluated Colorado River basin ESP forecasts using a number of probabilistic measures and found reliability deficiencies for many of the streamflow locations considered.

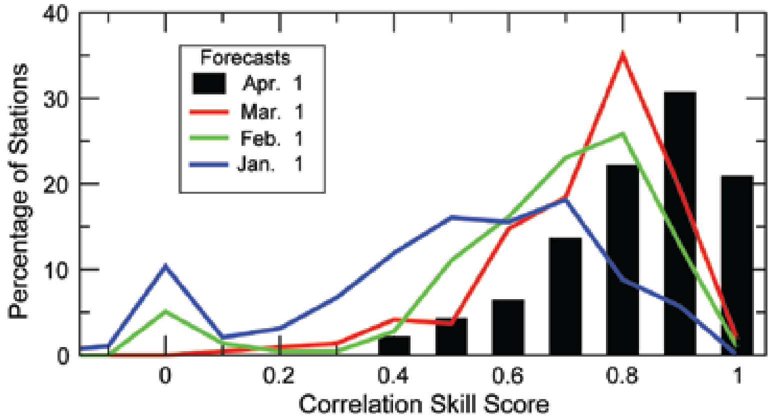


Figure 2.13 Percentages of stations with various correlation skill scores in the various panels (forecast dates) of Figure 2.12.

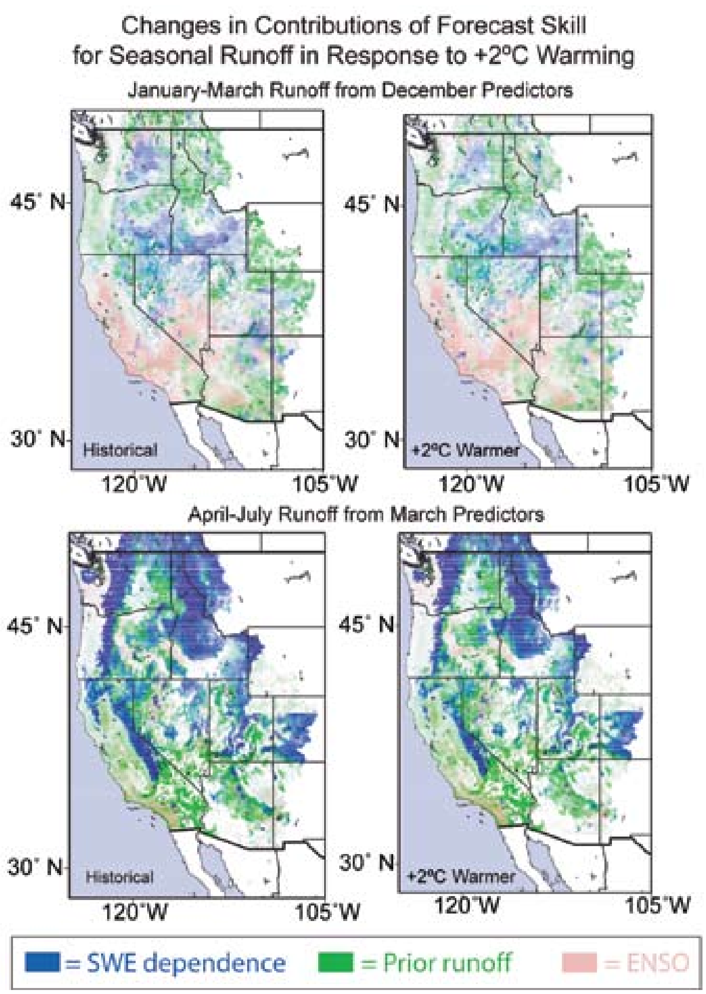
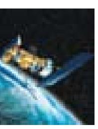


Figure 2.14 Potential contributions of antecedent snowpack conditions, runoff, and Niño 3.4 sea-surface temperatures to seasonal forecast skills in hydrologic simulations under historical, 1950 to 1999, meteorological conditions (left panels) and under those same conditions but with a 2°C uniform warming imposed (Dettinger, 2007).



2.2.3.2 The Implications of Decadal Variability and Long Term Change in Climate for Seasonal Hydrologic Prediction Skill

In the earlier discussion of sources of water-supply forecast skill, we highlighted the amounts and sources of skill provided by snow, soil moisture, and antecedent runoff influences. IPCC projections of global and regional warming, with its expected strong effects on western United States snowpack (Stewart *et al.*, 2004; Barnett *et al.*, 2008), raises the concern that prediction methods, such as regression, that depend on a consistent relationship between these predictors, and future runoff may not perform as expected if the current climate system is being altered in ways that then alters these hydro-climatic relationships. Decadal climate variability, particularly in precipitation (*e.g.*, Mantua *et al.*, 1997; McCabe and Dettinger, 1999), may also represent a challenge to such methods, although some researchers suggest that knowledge of decadal variability can be beneficial for streamflow forecasting (*e.g.*, Hamlet and Lettenmaier, 1999). One view (*e.g.*, Wood and Lettenmaier, 2006) is that hydrologic model-based forecasting may be more robust to the effects of climate change and variability due to the physical constraints of the land surface models, but this thesis has not been comprehensively explored.

The maps shown in Figure 2.14 are based on hydrologic simulations of a physically-based hydrologic model, called the Variable Infiltration Capacity (VIC) model (Liang *et al.*, 1994), in which historical temperatures are uniformly increased by 2ºC. These figures show that the losses of snowpack and the tendencies for more precipitation to fall as rain rather than snow in a warmer world reduce overall forecast skill, shrinking the areas where snowpack contributes strong predictability and also making antecedent runoff a less reliable predictor. Thus, many areas where warm-season runoff volumes are accurately predicted historically are likely to lose some forecast skill along with their snowpack. Overall, the average skill declines by about 2 percent (out of a historical average of 35 percent) for the January to March volumes and by about 4 percent (out of a historical average of 53 percent) for April to July. More importantly, though, are the declines in skill at grid cells where historical skills are greatest, nearly halving the occurrence of high-end (>0.8) January-to-March skills and reducing high-end April-to-July skills by about 15 percent (Figure 2.15).

This enhanced loss among the most skillful grid cells reflects the strong reliance of those grid cells on historical snowpacks for the greater part of their skill, snowpacks which decline under the imposed 2ºC warmer conditions. Overall, skills associated with antecedent runoff are more strongly reduced for the April-to-July runoff volumes, with reductions from an average contribution of 24 percent of variance predicted (by antecedent runoff) historically to 21 under the 2ºC warm conditions; for the January to March volumes, skill contributed by antecedent runoff only declines from 18.6 percent to 18.2 percent under the imposed warmer conditions. The relative declines in the contributions from snowpack and antecedent runoff make antecedent runoff (or, more directly, soil moisture, for which antecedent runoff is serving as a proxy here) a more important predictor to monitor in the future (for a more detailed discussion, see Section 2.4.2).

It is worth noting that the changes in skill contributions illustrated in Figure 2.14 are best-case scenarios. The skills shown are skills that would be provided by a complete recalibration of forecast equations to the new (imposed) warmer conditions, based on 50 years of runoff history. In reality, the runoff and forecast conditions are projected to gradually and continually trend towards increasingly warm conditions, and fitting new, appropriate forecast equations (and models) will always be limited by having only a brief reservoir of experience with each new degree of warming. Consequently, we must expect that regression-based forecast equations will tend to be increasingly and perennially out of date in a world with strong warming trends. This problem with the statistics of forecast skill in a changing world suggests development and deployment of more physically based, less statistically based forecast models should be a priority in the foreseeable future (Herrmann, 1999; Gleick, 2000; Milly *et al.*, 2008).

2.2.3.3 Skill of Climate Forecast-Driven Hydrologic Forecasts

The extent to which the ability to forecast U.S. precipitation and temperature seasons in advance can be translated into long-lead hydrologic forecasting has been evaluated by Wood *et al.* (2005). That evaluation compared hydrologic variables in the major river basins of the western conterminous United States as simulated by the VIC hydrologic model (Liang *et al.*, 1994), forced by two different sources of temperature and precipitation data: (1) observed historical meteorology (1979 to 1999); and (2) by hindcast climate-model-derived six-month-lead climate forecasts.

The Wood *et al.* (2005) assessment quantified and reinforced an important aspect of the hydro-

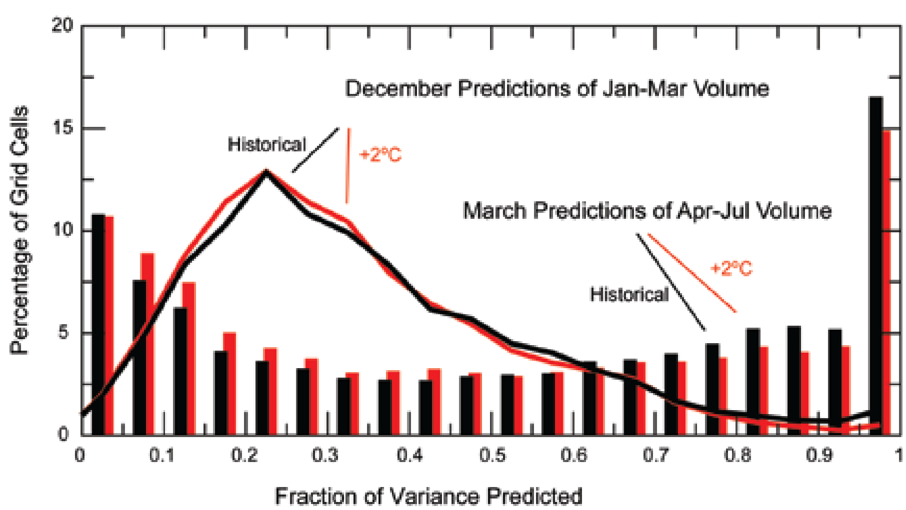


Figure 2.15 Distributions of overall fractions of variance predicted, in Figure 2.13, of January to March (curves) and April to July (histograms) runoff volumes under historical (black) and +2°C warmer conditions (Dettinger, 2007).

logic forecasting community's intuition about the current levels of hydrologic forecast skill using long-lead climate forecasts generated from various sources. The analysis first underscored the conclusions that, depending on the season, knowledge of initial hydrologic conditions conveys substantial forecast skill. A second finding was that the additional skill available from incorporating current (at the time) long-lead climate model forecasts into hydrologic prediction is limited when all years are considered, but can improve streamflow forecasts relative to climatological ESP forecasts in extreme ENSO years. If performance in all years is considered, the skill of current climate forecasts (particularly of precipitation) is inadequate to provide readily extracted hydrologic-forecast skill at monthly to seasonal lead times. This result is consistent with findings for North American climate predictability (Saha *et al.*, 2006). During El Niño years, however, the climate forecasts have adequate skill for temperatures, and mixed skill for precipitation, so that hydrologic forecasts for some seasons and some basins (especially California, the Pacific Northwest and the Great Basin) provide measurable improvements over the ESP alternative.

The authors of the Wood *et al.* (2005) assessment concluded that "climate model forecasts presently suffer from a general lack of skill, [but] there may be locations, times of year and conditions (*e.g.*, during El Niño or La Niña) for which they improve hydrologic forecasts relative to ESP". However, their conclusion was that improvements to hydrologic forecasts based on other forms of climate forecasts, *e.g.*, statistical or hybrid methods that are not completely reliant on a single climate model, may prove more useful in the near term in situations where alternative approaches yield better forecast skill than that which currently exists in climate models.

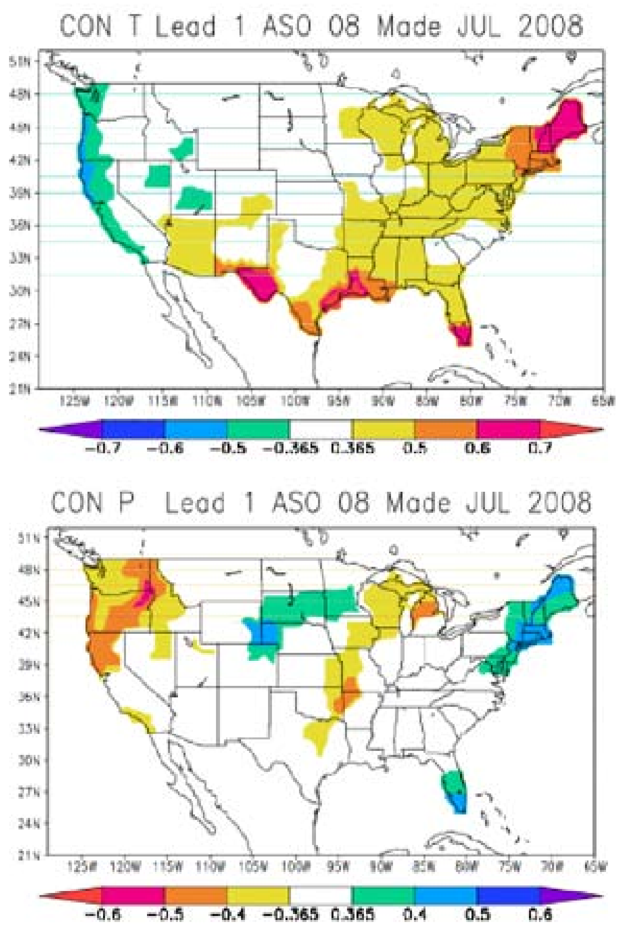


Figure 2.16 CPC objective consolidation forecast made in June 2007 (lead 1 month) for precipitation and temperature for the three month period Aug-Sep-Oct 2007. Figure obtained from <http://www.cpc.ncep.noaa.gov>.

2.3 CLIMATE DATA AND FORECAST PRODUCTS

2.3.1 A Sampling of Seasonal-to-Interannual Climate Forecast Products of Interest to Water Resource Managers

At SI lead times, a wide array of dynamical prediction products exist. A representative sample of SI climate forecast products is listed in Appendix A.1. The current dynamical prediction scheme used by NCEP, for example, is a system of models comprising individual models of the oceans, global atmosphere and continental land surfaces. These models were developed and originally run for operational forecast purposes in an uncoupled, sequential mode, an example of which is the so-called "Tier 2" framework in which the ocean model runs first, producing ocean surface boundary conditions that are prescribed as inputs for subsequent atmospheric model runs. Since 2004, a "Tier 1" scheme was introduced in which the models, together called the Coupled Forecast System (CFS) (Saha *et al.*, 2006), were fully coupled to allow dynamic exchanges of moisture and energy across the interfaces of the model components.

At NCEP, the dynamical tool, CFS, is complemented by a number of statistical forecast tools, three of which, Screening Multiple Linear Regression (SMLR), Optimal Climate Normals (OCN), and Canonical Correlation Analysis (CCA), are merged with the CFS to form an objective consolidation forecast product (Figure 2.16). While the consolidated forecast exceeds the skill of the individual tools, the official seasonal forecast from CPC involves a subjective merging of it with forecast and nowcast information sources from a number of different sources, all accessible to the public at CPC's monthly briefing. The briefing materials comprise 40 different inputs regarding the past, present and expected future state of the land, oceans and atmosphere from sources both internal and external to CPC. These materials are posted online at: <http://www.cpc.ncep.noaa.gov/products/predictions/90day/tools/briefing/>.

The resulting official forecast briefing has been the CPC's primary presentation of climate forecast information each month. Forecast products are accessible directly from CPC's root level home page in the form of maps of the probability anomalies for precipitation and temperature in three categories, or "terciles", representing below-normal, normal and above-normal values; a two-category scheme (above and below normal) is also available. This framework is used for the longer lead outlooks (Figure 2.17). The seasonal forecasts are also available in the form of maps of climate anomalies in degrees Celsius for temperature and inches for precipitation (Figure 2.18). The forecasts are released monthly, have a time-step of three months, and have a spatial unit of the climate division (Figure 2.19). For users desiring more information about the probabilistic forecast than is given in the map products, a "probability of exceedence" (POE) plot, with associated parametric information, is also available for each climate division (Figure 2.20). The POE plot shows the shift of the forecast probability distribution from the climatological distribution for each lead-time of the forecast.

In addition to NCEP, a few other centers, (*e.g.*, the International Research Institute for Climate and Society [IRI]) produce similar consensus forecasts and use a similar map-based, tercile-

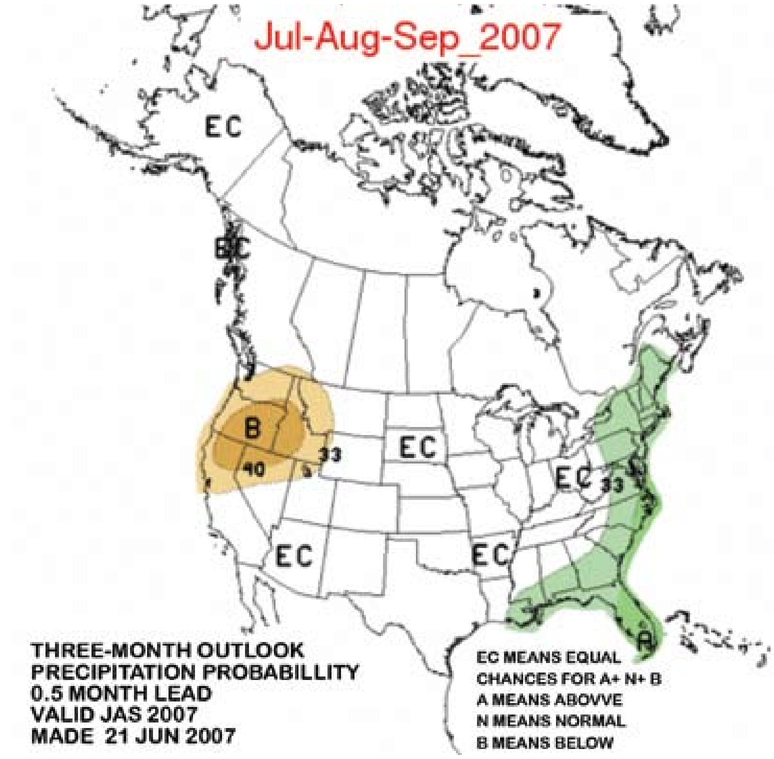


Figure 2.17 The National Center for Enivironmental Predictions CPC seasonal outlook for precipitation also shown as a tercile probability map. Tan/brown (green) shading indicates regions where the forecast indicates an increased probability for precipitation to be in the dry (wet) tercile, and the degree of shift is indicated by the contour labels. EC means the forecast predicts equal chances for precipitation to be in the A (above normal), B (below normal), or N (normal) terciles. Figure obtained from <http://www.cpc.ncep.noaa.gov/products/predictions/multi_season/13_seasonal_outlooks/color/page2.gif>.

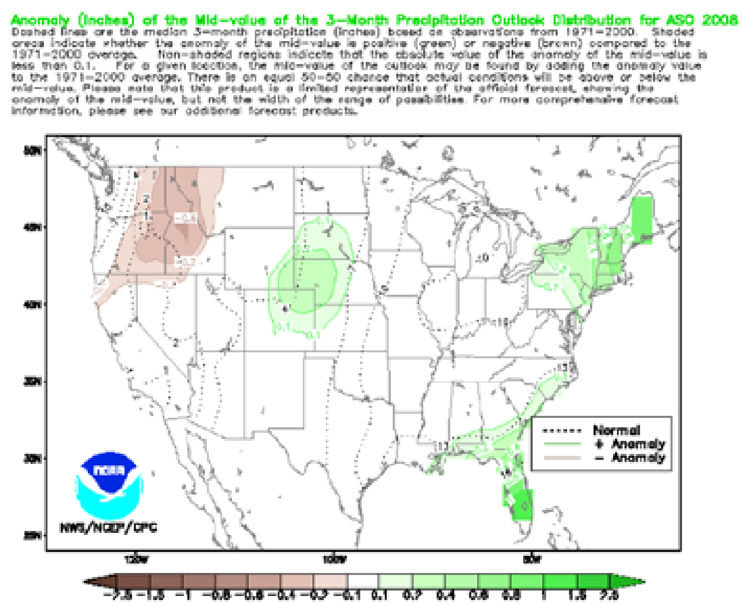
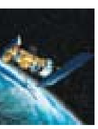


Figure 2.18 The National Center for Enivironmental Predictions CPC seasonal outlook for precipitation shown as inches above or below the total normal precipitation amounts for the 3-month target period (compare with the probability of exceedence forecast product shown in Figure 2.20). Figure obtained from <http://www.cpc.ncep.noaa.gov/products/predictions/long_range/poe_index.php?lead=3&var=p>.

focused framework for exhibiting their results. A larger number of centers run dynamical forecast tools, and the NOAA Climate Diagnostics Center, which produces monthly climate outlooks internally using statistical tools, also provides summaries of climate forecasts from a number of major sources, both in terms of probabilities or anomalies, for selected surface and atmospheric variables. Using dynamical models, the Experimental Climate Prediction Center (ECPC) at Scripps Institute provides monthly and seasonal time step forecasts of both climate and land surface variables at a national and global scale. Using these model outputs, ECPC also generates forecasts for derived variables that target wildfire management—*e.g.*, soil moisture and the Fireweather Index (see Chapter 4 for a more detailed description of Water Resource Issues in Fire-Prone U.S. Forests and the use of this index). The CPC has made similar efforts in the form of the Hazards Assessment, a short- to medium-range map summary of hazards related to extreme weather (such as flooding and wildfires), and the CPC Drought Outlook (Box 2.3), a subjective consensus product focusing on the evolution of large-scale droughts that is released once a month, conveying expectations for a three-month outlook period.

Seasonal-to-interannual forecast products are national to global in scale.



The foregoing is a brief survey of climate forecast products from major centers in the United States, and, as such, is far from a comprehensive presentation of the available sources. It does, however, provide examples from which the following observations about the general nature of climate prediction in the United Sates may be drawn. First, that operational SI climate forecasting is conducted at a relatively small number of federally-funded centers, and the resulting forecast products are national to global in scale. These products tend to have a coarse resolution in space and time, and are typically for basic earth system variables (*e.g.*, temperature, precipitation, atmospheric pressure) that are of general interest to many sectors. Forecasts are nearly always probabilistic, and the major products attempt to convey the inherent uncertainty via maps or data detailing forecast probabilities, although deterministic reductions (such as forecast variable anomalies) are also available.

2.3.2 Sources of Climate-Forecast Skill for North America

Much as with hydrologic forecasts, the skill of forecasts of climate variables (notably, temperature and precipitation) is not straightforward as it varies from region to region as well as with the forecast season and lead time; it is also limited by the chaotic and uncertain character of the climate system and derives from a variety of sources. While initial conditions are an important source for skill in SI hydrologic forecasts, the initial conditions of an atmospheric forecast are of little use after about 8 to 10 days as other forecast errors and/or disturbances rapidly grow, and therefore have no influence on SI climate forecast skill (Molteni *et al.*, 1996). SI forecasts are actually forecasts of those variations of the climate system that reflect predictable changes in boundary conditions, like seasurface temperatures (SSTs), or in external 'forcings,' disturbances in the radiative energy budget of the Earth's climate system. At time scales of decades-to-centuries, potential skill rests in predictions for slowly varying components of the climate system, like the atmospheric concentrations of carbon dioxide that influence the greenhouse effect, or slowly

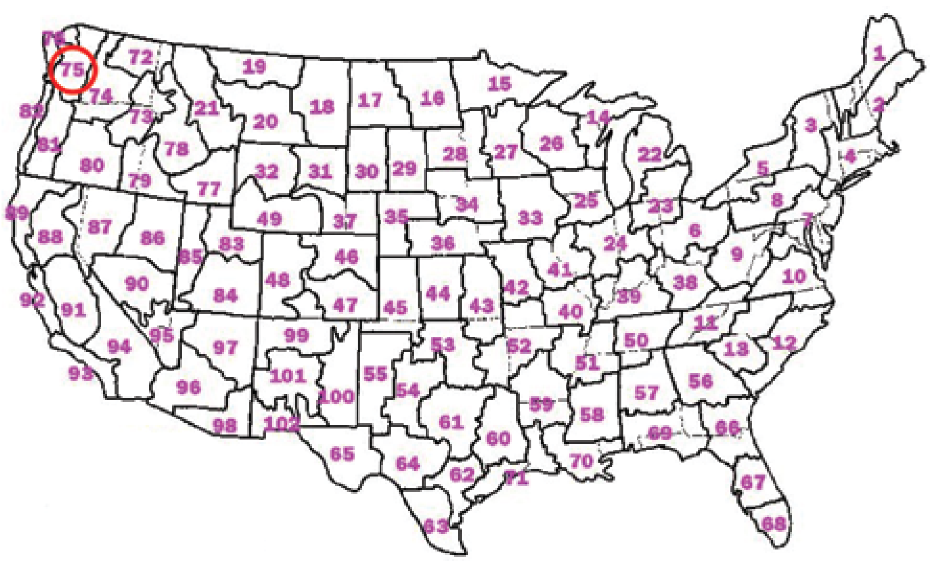


Figure 2.19 The CPC climate division spatial unit upon which the official seasonal forecasts are based. Figure obtained from <http://www.cpc.ncep.noaa.gov/products/predictions/long_range/poe_index.php?lead=3&var=p>.

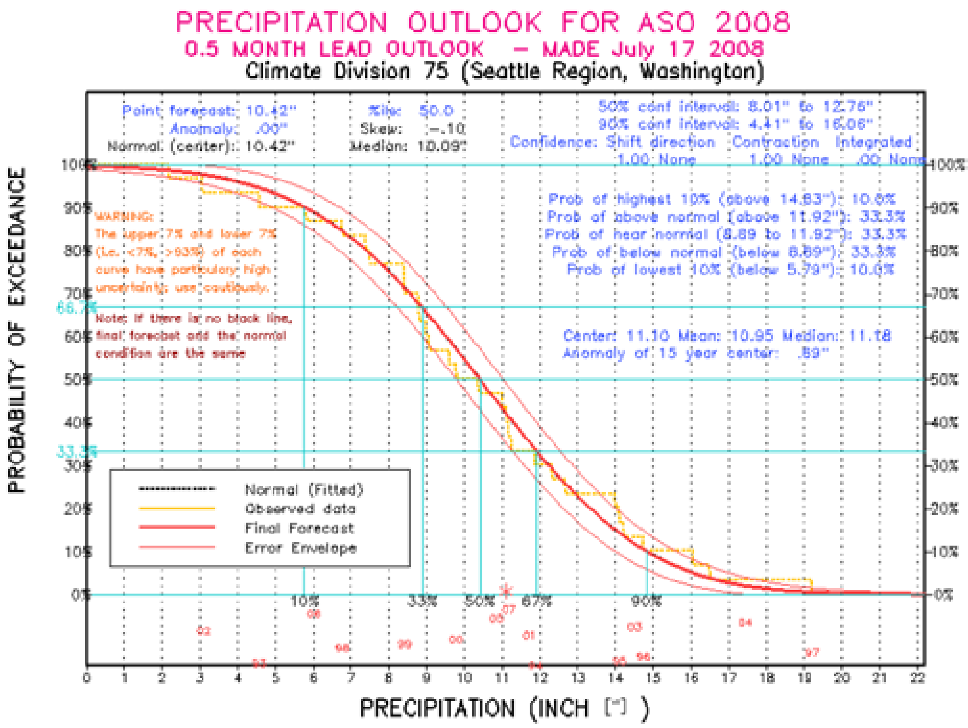
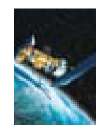


Figure 2.20 The NCEP CPC seasonal outlook for precipitation in the Seattle Region Climate Division (Division 75 in Figure 2.19) shown as the probability of exceedence for total precipitation for the three-month target period <http://www.cpc.ncep.noaa.gov/products/predictions/long_range/poe_graph_index.php?lead=3&climdiv=75&var=p.>.



evolving changes in ocean circulation that can alter SSTs and thereby change the boundary conditions for the atmosphere. Not all possible sources of SI climate-forecast skill have been identified or exploited, but contributors that have been proposed and pursued include a variety of large-scale air-sea connections (*e.g.*, Redmond and Koch, 1991; Cayan and Webb, 1992; Mantua *et al.*, 1997; Enfield *et al.*, 2001; Hoerling and Kumar, 2003), snow and sea-ice patterns (*e.g.*, Cohen and Entekhabi, 1999; Clark and Serreze, 2000; Lo and Clark, 2002; Liu *et al.*, 2004), and soil moisture and vegetation regimes (*e.g.*, Koster and Suarez, 1995, 2001; Ni-Meister *et al.*, 2005).

In operational practice, however, most of the forecast skill provided by current forecast systems (especially including climate models) derives from our ability to predict the evolution of ENSO events on time scales of 6 to 12 months, coupled with the teleconnections from the events in the tropical Pacific to many areas of the globe. Barnston *et al.* (1999), in their explanation of the advent of the first operational long-lead forecasts from the NOAA Climate Prediction Center, stated that "while some extratropical processes probably develop independently of the Tropics… much of the skill of the forecasts for the extratropics comes from anomalies of ENSO-related tropical sea surface temperatures". Except for the changes associated with diurnal cycles, seasonal cycles, and possibly the (30 to 60 day) Madden-Julian Oscillation of the tropical ocean-atmosphere system, "ENSO is the most predictable climate fluctuation on the planet" (McPhaden *et al.*, 2006). Diurnal cycles and seasonal cycles are predictable on time scales of hours-to-days and months-to-years, respectively, whereas ENSO mostly provides predictability on SI time scales. Figure 2.21a shows that temperatures over the tropical oceans and lands and extratropical oceans are more correlated from season to season than the extratropical continents. To the extent that they can anticipate the slow evolution of the tropical oceans, indicated by these correlations, SCFs in the extratropics that derive their skill from an ability to forecast conditions in the tropical oceans are provided a basis for prediction skill. To the extent that the multi-seasonal long-term potential predictability of the ENSO episodes (Figure 2.21b) can be drawn upon in certain regions at certain times of year,

Most of the skill provided by current forecast systems derives from our ability to predict the evolution of El Niño–Southern Oscillation events on time scales of 6 to 12 months.

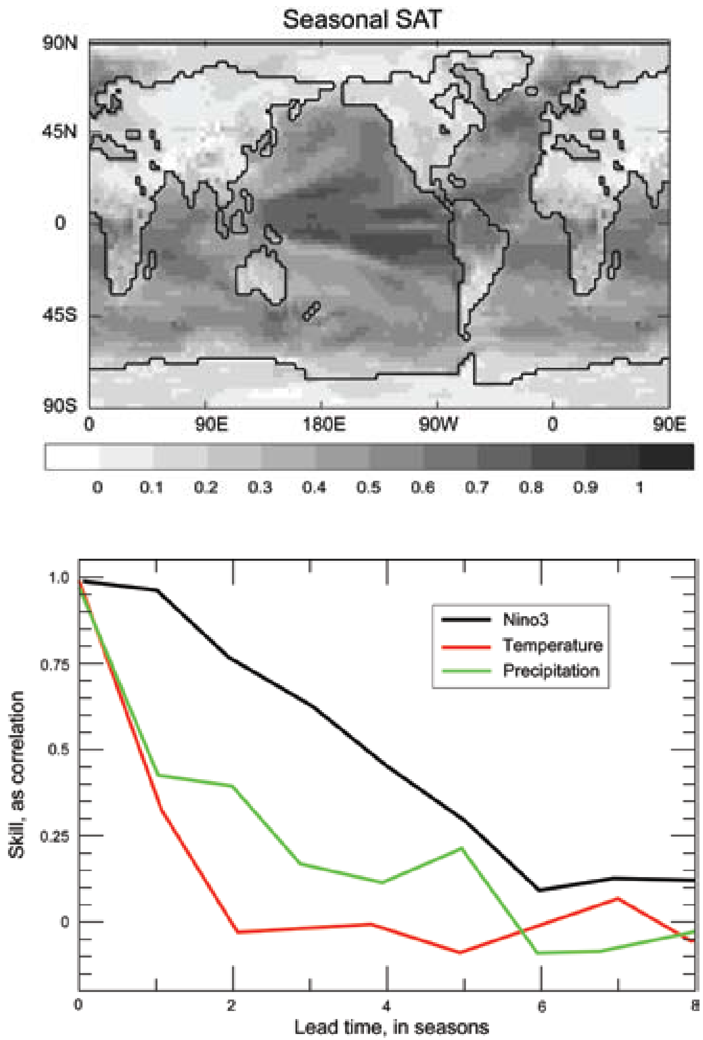
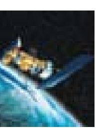


Figure 2.21 (a, top) Map of correlations between surface-air temperatures in each season and the following season in 600 years of historical climate simulation by the HadCM3 model (Collins 2002); (b, bottom) Potential predictability of a common ENSO index (Niño3 SST, the average of SSTs between 150°W and 90W, 5°S, and 5°N), average temperatures over the United States and Canada, and average precipitation over the United States and Canada, with skill measured by anomaly correlations and plotted against the forecast lead times; results extracted from Collins (2002), who estimated these skills from the reproducibility among multiple simulations of 30 years of climate by the HadCM3 coupled ocean-atmosphere model. Correlations below about 0.3 are not statistically significant at the 95 percent level.



the relatively meager predictabilities of North American temperatures and precipitation can be extended.

The scattered times between ENSO events drastically limits skillful prediction of events until, at least, the first faltering steps towards the initiation of an ENSO event have been observed. ENSO events, however, are frequently (but not always) phase-locked (synchronized) with aspects of the seasonal cycle (Neelin *et al.*, 2000), so that (a) forecasters know when to look most diligently for those "first faltering steps" and (b) the first signs of the initiation of an event are often witnessed 6 to 9 months prior to ENSO's largest expressions in the tropics and Northern Hemisphere (*e.g.*, Penland and Sardeshmukh, 1995). Thus, ENSO influences, however irregular and unpredictable they are on multiyear time scales, regularly provide the basis for SI climate forecasts over North America. ENSO events generally begin their evolution sometime in late (northern) spring or early summer, growing and maturing until they most often reach full strength (measured by either their SST expressions in the tropical Pacific or by their influences on the Northern Hemisphere) by about December – March (*e.g.*, Chen and van den Dool 1997). An ENSO event's evolution in the tropical ocean and atmosphere during the interim period is reproducible enough that relatively simple climate indices that track ENSO-related SST and atmospheric pressure patterns in the tropical Pacific provide predictability for North American precipitation patterns as much as two seasons in advance. Late summer values of the Southern Oscillation Index (SOI), for instance, are significantly correlated with a north-south see-saw pattern of wintertime precipitation variability in western North America (Redmond and Koch, 1991).

2.4 IMPROVING WATER RESOURCES FORECAST SKILL AND PRODUCTS

Although forecast skill is only one measure of the value that forecasts provide to water resources managers and the public, it is an important measure, and current forecasts are generally understood to fall short of the maximum possible skill on SI time scales (*e.g.*, <http://www.clivar.org/organization/wgsip/spw/spw_position.php>). Schaake *et al.* (2007) describe the SI hydrologic prediction process for model-based prediction in terms of several components: (1) development, calibration and/or downscaling of SI climate forecasts; (2) estimation of hydrologic initial conditions, with or without data assimilation; (3) SI hydrologic forecasting models and methods; and (4) calibration of the resulting forecasts. Notable opportunities for forecast skill improvement in each area are discussed here.

2.4.1 Improving Seasonal-to-Interannual Climate Forecast Use for Hydrologic Prediction

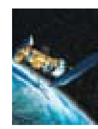
SI climate forecast skill is a function of the skill of climate system models, the efficacy of model combination strategies if multiple models are used, the accuracy of climate system conditions from which the forecasts are initiated, and the performance of post-processing approaches applied to correct systematic errors in numerical model outputs. Improvements are sought in all of these areas.

2.4.1.1 Climate forecast use

Many researchers have found that SI climate forecasts must be downscaled, disaggregated and statistically calibrated to be suitable as inputs for applied purposes (*e.g.*, hydrologic prediction, as in Wood *et al.*, 2002). Downscaling is the process of bridging the spatial scale gap between the climate forecast resolution and the application's climate input resolution, if they are not the same. If the climate forecasts are from climate models, for instance, they are likely to be at a grid resolution of several hundred kilometers, whereas the application may require climate information at a point (*e.g.*, station location). Disaggregation is similar to downscaling, but in the temporal dimension—for exapmple, seasonal climate forecasts may need to be translated into daily or sub-daily temperature and precipitation inputs for a given application. Forecast calibration is a process by which the statistical properties (such as bias and spread errors) of a probabilistic forecast are corrected to match their observed error statistics (*e.g.*, Atger, 2003; Hamill *et al.*, 2006). These procedures may be distinct from each other, or they may be inherent parts of a single approach (such as the analogue techniques of Hamill *et al.*, 2006). These steps do not necessarily improve the signal to noise ratio of the climate forecast, but done properly, they do correct bias and reliability problems that would otherwise render impossible their use in applications. For shorter lead predictions, corrections to forecast outputs have long been made based on (past) model output statistics (MOS; Glahn and Lowry, 1972). MOS are sets of statistical relations (*e.g.*, multiple linear regression) that effectively convert numerical model outputs into unbiased, best climate predictions for selected areas or stations, where "best" relates to past performance of the model in reproducing observations. MOS corrections are widely used in weather prediction (Dallavalle and Glahn, 2005). Corrections may be as simple as removal of mean biases indicated by historical runs of the model, with the resulting forecasted anomalies superimposed on station climatology. More complex methods specifically address spatial patterns in climate forecasts based on specific inadequacies of the models in reproducing key teleconnection patterns or topographic features (*e.g.*, Landman and Goddard, 2002; Tippett *et al.*, 2003).

Seasonal-to-interannual climate forecasts must be downscaled, disaggregated and statistically calibrated to be suitable as inputs for applied purposes.

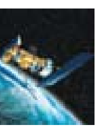
A primary limitation on calibrating SI forecasts is the relatively small number of retrospective forecasts available for identifying biases. Weather predictions are made every day, so even a few years of forecasts provide a large number of examples from which to learn. SI forecasts, in contrast, are comparatively infrequent and even the number of forecasts made over several decades may not provide an adequate resource with which to develop model-output corrections (Kumar, 2007). This limitation is exacerbated when the predictability and biases themselves vary between years and states of the global climate system. Thus, there is a clear need to expand current "reforecast" practices for fixed SI climate models over long historical periods to provide both for quantification (and verification) of the evolution of SI climate forecast skills and for post-processing calibrations to those forecasts.



2.4.1.2 Development of objective multi-model ensemble approaches

The accuracy of SI climate forecasts has been shown to increase when forecasts from groups of models are combined into multi-model ensembles (*e.g.*, Krishnamurti *et al.*, 2000; Palmer *et al.*, 2004; Tippett *et al.*, 2007). Multi-model forecast ensembles yield greater overall skill than do any of the individual forecasts included, in principle, as a result of cancellation of errors between ensemble members. Best results thus appear to accrue when the individual models are of similar skill and when they exhibit errors and biases that differ from model to model. In part, these requirements reflect the current uncertainties about the best strategies for choosing among models for inclusion in the ensembles used and, especially for weighting and combining the model forecasts within the ensembles. Many methods have been proposed and implemented (*e.g.*, Rajagopalan *et al.*, 2002; Yun *et al.*, 2005), but strategies for weighting and combining ensemble members are still an area of active research (*e.g.*, Doblas-Reyes *et al.*, 2005; Coelho *et al.*, 2004). Multi-model ensemble forecast programs are underway in Europe (DEMETER, Palmer *et al.*, 2004) and in Korea (APEC; *e.g.*, Kang and Park, 2007). In the United States, IRI forms an experimental multi-model ensemble forecast, updating monthly, from seasonal forecast ensembles run separately at seven centers, a "simple multi-model" approach that compares well with centrally organized efforts such as DEMETER (Doblas-Reyes *et al.*, 2005). The NOAA Climate Test Bed Science Plan also envisions such a capability for NOAA (Higgins *et al.*, 2006).

Seasonal-to-interannual climate forecast skill for most regions comes from knowledge of current sea surface temperatures or predictions of future sea surface temperatures, especially those in the tropics.



2.4.1.3 Improving Climate Models, Initial Conditions, and Attributions

Improvements to climate models used in SI forecasting efforts should be a high priority. Several groups of climate forecasters have identified the lack of key aspects of the climate system in current forecast models as important weaknesses, including underrepresented linkages between the stratosphere and troposphere (Baldwin and Dunkerton, 1999), limited processes and initial conditions at land surfaces (Beljaars *et al.*, 1996; Dirmeyer *et al.*, 2006; Ferranti and Viterbo, 2006), and lack of key biogeochemical cycles like carbon dioxide.

Because climate prediction is, by most definitions, a problem determined by boundary condition rather than an initial condition, specification of atmospheric initial conditions is not the problem for SI forecasts that it is for weather forecasts. However, SI climate forecast skill for most regions comes from knowledge of current SSTs or predictions of future SSTs, especially those in the tropics (Shukla *et al.*, 2000; Goddard and Dilley, 2005; Rosati *et al.*, 1997). Indeed, forecast skill over land (worldwide) increases directly with the strength of an ENSO event (Goddard and Dilley, 2005). Thus, an important determinant of recent improvements in SI forecast skill has been the quality and placement of tropical ocean observations, like the TOGA-TAO (Tropical Atmosphere Ocean project) network of buoys that monitors the conditions that lead up to and culminate in El Niño and La Niña events (Trenberth *et al.*, 1998; McPhaden *et al.*, 1998; Morss and Battisti, 2004). More improvements in all of the world's oceans are expected from the broader Array for Real-time Geostrophic Oceanography (ARGO) upper-ocean monitoring arrays and Global Ocean Observing System (GOOS) programs (Nowlin *et al.*, 2001). In many cases, and especially with the new widespread ARGO ocean observations, ocean data assimilation has improved forecast skill (*e.g.*, Zheng *et al.*, 2006). Data assimilation into coupled ocean-atmosphere-land models is a difficult and unresolved problem that is an area of active research (*e.g.*, Ploshay and Anderson, 2002; Zheng *et al.*, 2006). Land-surface and cryospheric conditions also can influence the seasonal-scale dynamics that lend predictability to SI climate forecasting, but incorporation of these initial boundary conditions into SI climate forecasts is in an early stage of development (Koster and Suarez, 2001; Lu and Mitchell, 2004; Mitchell *et al.*, 2004). Both improved observations and improved avenues for including these conditions into SI climate models, especially with coupled ocean-atmosphere-land models, are needed. Additionally, education and expertise deficiencies contribute to unresolved problems in data assimilation for geophysical modeling. The Office of the Federal Coordinator for Meteorology (2007) documents that there is a need for more students (either undergraduate or graduate) who have sufficient mathematics and computer science skills to engage in data assimilation work in the research and/or operational environment.

Finally, a long-standing but little explored approach to improving the value of SI climate forecasts is the attribution of the causes of

climate variations. The rationale for an attribution effort is that forecasts have greater value if we know why the forecasted event happened, either before or after the event, and why a forecast succeeded or failed, after the event. The need to distinguish natural from human-caused trends, and trends from fluctuations, is likely to become more and more important as climate change progresses. SI forecasts are likely to fail from time to time or to realize less probable ranges of probabilistic forecasts. Knowing that forecasters understand the failures (in hindsight) and have learned from them will help to build increasing confidence through time among users. Attempts to attribute causes to important climate events began as long ago as the requests from Congress to explain the 1930s Dust Bowl. Recently NOAA has initiated a Climate Attribution Service (see: <http://www.cdc.noaa.gov/CSI/>) that will combine historical records, climatic observations, and many climate model simulations to infer the principal causes of important climate events of the past and present. Forecasters can benefit from knowledge of causes and effects of specific climatic events as well as improved feedbacks as to what parts of their forecasts succeed or fail. Users will also benefit from knowing the reasons for prediction successes and failures.

2.4.2 Improving Initial Hydrologic Conditions for Hydrologic and Water Resource Forecasts

Operational hydrologic and water resource forecasts at SI time scales derive much of their skill from hydrologic initial conditions, with the particular sources of skill depending on seasons and locations. Better estimation of hydrologic initial conditions will, in some seasons, lead to improvements in SI hydrologic and consequently, water resources forecast skill. The four main avenues for progress in this area are: (1) augmentation of climate and hydrologic observing networks; (2) improvements in hydrologic models (*i.e.*, physics and resolution); (3) improvements in hydrologic model calibration approaches; and (4) data assimilation.

2.4.2.1 Hydrologic observing networks

As discussed previously (in Section 2.2), hydrologic and hydroclimatic monitoring networks provide crucial inputs to hydrologic and water resource forecasting models at SI time scales. Continuous or regular measurements of streamflow, precipitation and snow water contents provide important indications of the amount of water that entered and left river basins prior to the forecasts and thus directly or indirectly provide the initial conditions for model forecasts.

Observed snow water contents are particularly important sources of predictability in most of the western half of the United States, and have been measured regularly at networks of snow courses since the 1920s and continually at SNOTELs (automated and telemetered snow instrumentation sites) since the 1950s. Snow measurements can contribute as much as three-fourths of the skill achieved by warm-season water supply forecasts in the West (Dettinger, 2007). However, recent studies have shown that measurements made at most SNOTELs are not representative of overall basin water budgets, so that their value is primarily as indices of water availability rather than as true monitors of the overall water budgets (Molotch and Bales, 2005). The discrepancy arises because most SNOTELs are located in clearings, on flat terrain, and at moderate altitudes, rather than the more representative snow courses that historically sampled snow conditions throughout the complex terrains and micrometeorological conditions found in most river basins. The discrepancies limit some of the usefulness of SNOTEL measurements as the field of hydrologic forecasting moves more and more towards physically-based, rather than empirical-statisti-

The need to distinguish natural from human-caused trends, and trends from fluctuations, is likely to become more and more important as climate change progresses.

cal models. To remedy this situation, and to provide more diverse and more widespread inputs as required by most physically-based models, combinations of remotely sensed snow conditions (to provide complete areal coverage) and extensions of at least some SNOTELs to include more types of measurements and measurements at more nearby locations will likely be required (Bales *et al.*, 2006).

Groundwater level networks already are contributing to drought monitors and response plans in many states.

Networks of ground-water level measurements are also important because: (1) these data support operations and research, and (2) the networks' data may be critical to some aspects of future hydrologic forecast programs. Groundwater level measurements are made at thousands of locations around the United States, but they have only recently been made available for widespread use in near-real time (see: <http://ogw01.er.usgs.gov/USGSGWNetworks.asp>). Few operational surface water resource forecasts have been designed to use ground-water measurements. Similarly climate-driven SI groundwater resource forecasts are rare, if made at all. However, surface water and groundwater are interlinked in nearly all cases and, in truth, constitute a single resource (Winter *et al.*, 1998). With the growing availability of real-time groundwater data dissemination, opportunities for improving water resource forecasts by better integration and use of surface- and groundwater data resources may develop. Groundwater level networks already are contributing to drought monitors and response plans in many states.

Similarly, long-term soil-moisture measurements have been relatively uncommon until recently, yet are of potentially high value for many land management activities including range management, agriculture, and drought forecasting. Soil moisture is an important control on the partitioning of water between evapotranspiration, groundwater recharge, and runoff, and plays an important (but largely unaddressed) role in the quantities addressed by water resource forecasts. Soil moisture varies rapidly from place to place (Vinnikov *et al.*, 1996; Western *et al.*, 2004) so that networks that will provide representative measurements have always been difficult to design (Wilson *et al.*, 2004). Nonetheless, the Illinois State Water Survey has monitored soil moisture at about 20 sites in Illinois for many years (see: <http://www.sws.uiuc.edu/warm/soilmoist/ISWSSoilMoistureSummary.pdf>), but was alone in monitoring soil moisture at the state scale for most of that time. As the technologies for monitoring soil moisture have become less troublesome, more reliable, and less expensive in recent years, more agencies are beginning to install soil-moisture monitoring stations (*e.g.*, the NRCS is augmenting many of its SNOTELs with soil-moisture monitors and has established a national Soil Climate Analysis Network (SCAN; <http://www.wcc.nrcs.usda.gov/scan/SCAN-brochure.pdf>); Oklahoma's Mesonet micrometeorological network includes soil-moisture measurements at its sites; California is on the verge of implementing a state-scale network at both high and low altitudes). With the advent of regular remote sensing of soil-moisture conditions (Wagner *et al.*, 2007), many of these *in situ* networks will be provided context so that their geographic representativeness can be assessed and calibrated (Famligietti *et al.*, 1999). As with groundwater, soil moisture has not often been an input to water resource forecasts on the SI time scale. Instead, if anything, it is being simulated, rather than measured, where values are required. Increased monitoring of soil moisture, both remotely and *in situ*, will provide important checks on the models of soil-moisture reservoirs that underlie nearly all of our water resources and water resource forecasts, making hydrological model improvements possible.

Augmentation of real-time stream gauging networks is also a priority, a subject discussed in the Synthesis and Assessment Product 4.3 (CCSP, 2008).

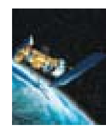
2.4.2.2 Improvements in Hydrologic Modeling Techniques

Efforts to improve hydrologic simulation techniques have been pursued in many areas since the inception of hydrologic modeling in the 1960s and 1970s when the Stanford Watershed Model (Crawford and Linsley, 1966), the Sacramento Model (Burnash *et al.*, 1973) and others were created. More recently, physically-based, distributed and semi-distributed hydrologic models have been developed, both at the watershed scale (*e.g.*, Wigmosta *et al.*, 1994; Boyle *et al.*, 2000) to account for terrain and climate inhomogeneity, and at the regional scale (Liang *et al.*, 1994 among others). Macroscale models (like the Sacramento Model and the Stanford Watershed Model) were partly motivated by the need to improve land surface representation in climate system modeling approaches (Mitchell *et al.*, 2004), but these models have also been found useful for hydrologic applications related to water management (*e.g.*, Hamlet and Lettenmaier, 1999; Maurer and Lettenmaier, 2004; Wood and Lettenmaier, 2006). The NOAA North American Land Data Assimilation Project (Mitchell *et al.*, 2004) and NASA Land Information System (Kumar *et al.*, 2006) projects are leading agency-sponsored research efforts that are focused on advancing the development and operational deployments of the regional/physically based models. These efforts include research to improve the estimation of observed parameters (*e.g.*, use of satellite remote sensing for vegetation properties and distribution), the accuracy of meteorological forcings, model algorithms and computational approaches. Progress in these areas has the potential to improve the ability of hydrologic models to characterize land surface conditions for forecast initialization, and to translate future meteorology and climate into future hydrologic response.

Aside from improving hydrologic models and inputs, strategies for hydrologic model implementation are also important. Model calibration—, the identification of optimal parameter sets for simulating particular types of hydrologic output (single or multiple)—has arguably been the most extensive area of research toward improving hydrologic modeling techniques (*e.g.,* Wagener and Gupta, 2005, among others). This body of work has yielded advances in the understanding of the model calibration problem from both practical and theoretical perspectives. The work has been conducted using models at the watershed scale to a greater extent than the regional scale, and the potential for applying these techniques to the regional scale models has not been explored in depth.

Data assimilation is another area of active research (*e.g.*, Andreadis and Lettenmaier 2006; Reichle *et al.*, 2002; Vrugt *et al.*, 2005; Seo *et al.*, 2006). It is a process in which verifying observations of model state or output variables are used to adjust the model variables as the model is running, thereby correcting simulation errors on the fly. The primary types of observations that can be assimilated include snow water equivalent and snow covered area, land surface skin temperature, remotely sensed or *in situ* soil moisture, and streamflow. NWSRFS has the capability to do objective data assimilation. In practice, NWS (and other agencies) perform a qualitative data assimilation, in which forecaster judgment is used to adjust model states and inputs to reproduce variables such as streamflow, snow line elevation and snow water equivalent prior to initializing an ensemble forecast.

Efforts to improve hydrologic simulation techniques have been pursued in many areas since the inception of hydrologic modeling in the 1960s and 1970s.



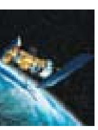
2.4.3 Calibration of Hydrologic Model Forecasts

Even the best real-world hydrologic models have biases and errors when applied to specific gages or locations. Statistical models often are tuned well enough so that their biases are relatively small, but physically-based models often exhibit significant biases. In either case, further improvements in forecast skill can be obtained, in principle, by post-processing model forecasts to remove or reduce any remaining systematic errors, as detected in the performance of the models in hindcasts. Very little research has been performed on the best methods for such post-processing (Schaake *et al.*, 2007), which is closely related to the calibration corrections regularly made to weather forecasts. Seo *et al.* (2006), however, describe an effort being undertaken by the National Weather Service for short lead hydrologic forecasts, a practice that is more common than for longer lead hydrologic forecasts. Other examples include work by Hashino *et al.* (2007) and Krzysztofowicz (1999). At least one example of an application for SI hydrologic forecasts is given in Wood and Schaake (2008); but as noted earlier, a major limitation for such approaches is the limited sample sizes available for developing statistical corrections.

2.5 IMPROVING PRODUCTS: FORECAST AND RELATED INFORMATION PACKAGING AND DELIVERY

There is wide support for a comparative and relative "now *versus* normal *versus* last year" form of characterizing hydrologic and climate forecasts.

The value of SI forecasts can depend on more than their forecast skill. The context that is provided for understanding or using forecasts can contribute as much or more to their value to forecast users. Several avenues for re-packaging and providing context for SI forecasts are discussed in the following paragraphs.



Probabilistic hydrologic forecasts typically represent summaries of collections of forecasts, forecasts that differ from each other due to various representations of the uncertainties at the time of forecast or likely levels of climate variation after the forecast is made, or both (Schaake *et al.*, 2007). For example, the "ensemble streamflow prediction" methodology begins its forecasts (generally) from a single best estimate of the initial conditions from which the forecasted quantity will evolve, driven by copies of the historical meteorological variations from each year in the past (Franz *et al.*, 2003). This provides ensembles of as many forecasts as there are past years of appropriate meteorological records, with the ensemble scatter representing likely ranges of weather variations during the forecast season. Sometimes deterministic forecasts are extended to represent ranges of possibilities by directly adding various measures of past hydrologic or climatic variability. More modern probabilistic methods are based on multiple climate forecasts, multiple initial conditions or multiple parameterizations (including multiple downscalings) (Clark *et al.*, 2004; Schaake *et al.*, 2007). However accomplished, having made numerous forecasts that represent ranges of uncertainty or variability, the probabilistic forecaster summarizes the results in terms of statistics of the forecast ensemble and presents the probabilistic forecast in terms of selected statistics, like probabilities of being more or less than normal.

In most applications, it is up to the forecast user to interpret these statistical descriptions in terms of their own particular data needs, which frequently entails (1) application of various corrections to make them more representative of their local setting and (2), in some applications, essentially a deconvolution of the reported probabilities into plausible examples that might arise during the future described by those probabilities. Forecast users in some cases may be better served by provision of historical analogs that closely resemble the forecasted conditions, so that they can analyze their own histories of the results during the analogous (historical) weather conditions. For example, Wiener *et al.* (2000) report that there is wide support for a comparative and relative "now *versus* normal *versus* last year" form of characterizing hydrologic and climate forecasts. Such qualitative characterizations would require careful and explicit caveats, but still have value as reference to historical conditions in which most current managers learned their craft and in which operations were institutionalized or codified. While "normal" is increasingly problematic, "last year" may be the best and most accessible analogue for the wide variety of relevant market conditions in which agricultural water users (and their competitors), for example, operate.

Alternatively, some forecast users may find that elements from the original ensembles of forecasts would provide useful examples that could be analyzed or modeled in order to more clearly represent the probabilistic forecast in concrete terms. The original forecast ensemble members are the primary source of the probabilistic forecasts and can offer clear and definite examples of what the forecasted future *could* look like (but not specifically what it *will* look like). Thus, along with the finished forecasts, which should remain the primary forecast products, other representations of what the forecasts are and how they would appear in the real world could be useful and more accessible complements for some users, and would be a desirable addition to the current array of forecast products.

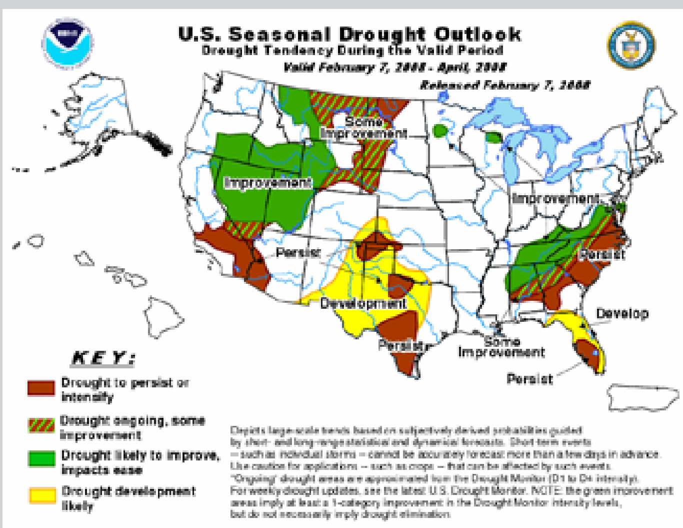
Another approach to providing context (and, potentially, examples) for the SI water resource forecasts involves placing the SI forecasts in the context of paleoclimate reconstructions for the prior several centuries. The twentieth century

BOX 2.3: The CPC Seasonal Drought Outlook

The CPC Drought Outlook (DO) is a categorical prediction of drought evolution for the three months forward from the forecast date. The product, which is updated once per month, comprises a map that is accompanied by a text discussion of the rationale for the categories depicted on the map.

The starting conditions for the DO are given by the current Drought Monitor (DM) (a United States map that is updated weekly showing the status of drought nationwide located: <http://www.drought.unl.edu/DM/monitor.html>), and the DO shows likely changes in and adjacent to the current DM drought areas. The DO is a subjective consensus forecast that is assembled each month by a single author (rotating between CPC and the National Drought Mitigation Center [NDMC]) with feedback from a panel of geographically distributed agency and academic experts. The basis for estimating future drought evolution includes a myriad of operational climate forecast products: from short- and medium-range weather forecasts to seasonal predictions from the CPC climate outlooks and the NCEP CFS outputs; consideration of climate tendencies for current El Nino–Southern Oscillation state; regional hydroclimatology; and medium-range to seasonal soil moisture and runoff forecasts from a variety of sources.

The DO makes use of the most advanced objective climate and hydrologic prediction products currently available, including not only operational, but experimental products, although the merging of the different inputs is based on expert judgment rather than an objective system. The DO is verified by comparing the DM drought assessments at the start and end of the DO forecast period; verification skill scores have been tracked for the last seven years. The DO is the primary drought-related agency forecast produced in the United States, and is widely used by the drought management and response community from local to regional scales.

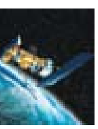


The DO was developed in the context of new drought assessment partnerships between the CPC, U.S. Department of Agriculture and the NDMC following the passage of the National Drought Policy Act of 1998. The DM was released as an official product in August, 1999, with the expectation that a weekly or seasonal drought forecast capacity would be added in the future. A drought on the Eastern Seaboard in the fall of 1999 required briefings for the press and the Clinton Administration; internal discussions between DM participants at the CPC led to the formation of the first version of the DO (maps and text) for these briefings. These were released informally to local, state and federal agency personnel throughout the winter of 1999 to 2000, and received positive feedback.

The CPC decided to make the products official, provided public statements and developed product specifications, and made the product operational in March 2000. The initial development process was informal and lasted about six months. In November 2000, the first Drought Monitor Forum was held, at which producers and users (agency, state, private, academic) came together to evaluate the DM in its first year and plan for its second, providing, in addition, a venue for discussion of the DO. This forum still meets bi-annually, focusing on both DM- and DO-relevant issues. Developmental efforts for the DO are internal at CPC or within NCEP, and the primary avenues for feedback are the website and at presentations by DO authors at workshops and conferences. The DO authors also interact with research efforts funded by the National Oceanic and Atmospheric Administration (NOAA) Climate Program Office and other agency funding sources, and with NOAA research group efforts (such as at NCEP), as part of the ongoing development effort. URL: <http://www.cpc.noaa.gov/products/expert_assessment/drought_assessment.shtml>.

has, by and large, been climatically benign in much of the nation, compared to previous centuries (Hughes and Brown, 1992; Cook *et al.*, 1999). As a consequence, the true likelihood of various forecasted, naturally-occurring climate and water resource anomalies may best be understood in the context of longer records, which paleoclimatic reconstructions can provide. At present, approaches to incorporating paleoclimatic information into responses to SI forecasts are uncommon and only beginning to develop, but eventually they may provide a clearer framework for understanding and perfecting probabilistic SI water resource forecasts. One approach being investigated is the statistical synthesis of examples (scenarios) that reflect both the long-term climate variability identified in paleo-records and time-series-based deterministic long-lead forecasts (Kwon *et al.*, 2007).

The true likelihood of various forecasted, naturally-occurring climate and water resource anomalies may best be understood in the context of longer records, which paleoclimatic reconstructions can provide.



2.6 THE EVOLUTION OF PROTOTYPES TO PRODUCTS AND THE ROLE OF EVALUATION IN PRODUCT DEVELOPMENT

Studies of what makes forecasts useful have identified a number of common characteristics in the process by which forecasts are generated, developed, and taught to and disseminated among users (Cash and Buizer, 2005). These characteristics include: ensuring that the problems that forecasters address are themselves driven by forecast users; making certain that knowledge-to-action networks (the process of interaction between scientists and users which produces forecasts) are end-to-end inclusive; employing "boundary organizations" (groups or other entities that bridge the communication void between experts and users) to perform translation and mediation functions between the producers and consumers of forecasts; fostering a social learning environment between producers and users (*i.e.*, emphasizing adaptation); and providing stable funding and other support to keep networks of users and scientists working together.

This Section begins by providing a review of recent processes used to take a prototype into an operational product, with specific examples from the NWS. Some examples of interactions between forecast producers and users that have lead to new forecast products are then reviewed, and finally a vision of how user-centric forecast evaluation could play a role in setting priorities for improving data and forecast products in the future is described.

2.6.1 Transitioning Prototypes to Products

During testimony for this Product, heads of federal operational forecast groups all painted a relatively consistent picture of how most in-house innovations currently begin and evolve. Although formal and quantitative innovation planning methodologies exist (see Appendix A.3: Transitioning NWS Research into Operations and How the Weather Service Prioritizes the Development of Improved Hydrologic Forecasts), for the most part, the operational practice is often relatively *ad hoc* and unstructured except for the larger and longer-term projects. The Seasonal Drought Outlook is an example of a product that was developed under a less formal process than that used by the NWS (Box 2.3).

Climate and water resource forecasters are often aware of small adjustments or "tweaks" to forecasts that would make their jobs easier; these are often referred to as "forecasts of opportunity". A forecaster may be aware of a new dataset or method or product that he/she believes could be useful. Based on past experience, production of the forecast may seem feasible and it could be potentially skillful. In climate forecasting in particular, where there is very high uncertainty in the forecasts themselves and there is marginal user adoption of existing products, the operational community often focuses more on potential forecast skill than likely current use. The belief is that if a product is skillful, a user base could be cultivated. If there is no skill, even if user demand exists, forecasting would be futile.

Attractive projects may also develop when a new method comes into use by a colleague of the forecaster (someone from another agency, alumni, friend or prior collaborator on other projects). For example, Redmond and Koch (1991) published the first major study of the impacts of ENSO on streamflow in the western United States. At the time the study was being done, a NRCS operational forecaster was one of Koch's graduate students. The student put

Koch's research to operational practice at the NRCS after realizing that forecast skill could be improved.

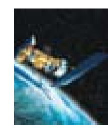
Efficiency is also often the inspiration for an innovation. A forecaster may be looking for a way to streamline or otherwise automate an existing process. For example, users frequently call the forecaster with a particular question; if it is possible to automate answering that question with a new Internet-based product, the forecaster may be freed up to work on other tasks. While most forecasters can readily list several bottlenecks in the production process, this knowledge often comes more from personal experience than any kind of structured system review.

At this stage, many ideas exist for possible innovations, although only some small subset of them will be pursued. The winnowing process continues with the forecaster and/or peers evaluating the feasibility of the innovation: Is the method scientifically defensible? Are the data reliably available to support the product? Are the computers powerful enough to complete the process in a reasonable time? Can this be done with existing resources, would it free up more resources than it consumes, or is the added value worth the added operational expense? In other words, is the total value of the advance worth the effort? Is it achievable and compatible with legacy systems or better than the total worth of the technology, installed base and complementary products?

If it is expected to be valuable, some additional questions may be raised by the forecaster or by management about the appropriateness of the solution. Would it conflict with or detract from another product, especially the official suite (*i.e.*, destroy competency)? Would it violate an agency policy? For example, a potential product may be technically feasible but not allowed to exist because the agency's webpage does not permit interactivity because of increasingly stringent congressionally-mandated cyber-security regulations. In this case, to the agency as a whole, the cost of reduced security is greater than the benefit of increased interactivity. It is important to note that if security and interactivity in general are not at odds, the issue may be that a particular form of interactivity is not compatible with the existing security architecture. If a different security architecture is adopted or a different form of interactivity used (*e.g.*, written in a different computer language), then both may function together, assuming one has the flexibility and ability to change.

Additionally, an agency policy issue can sometimes be of broader, multi-organizational scope and would require policy decisions to settle. For example, no agency currently produces water quality forecasts. Which federal agency should be responsible for this: the U.S. Department of Agriculture, Environmental Protection Agency, U.S. Geological Survey or National Weather Service? What of soil moisture forecasts? Should it be the first agency to develop the technical proficiency to make such forecasts? Or should it be established by a more deliberative process to prevent "mission creep?" Agencies are also concerned about whether innovations interfere with the services provided by the private sector.

No agency currently produces water quality forecasts.

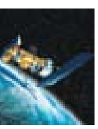


If appropriate, the forecaster may then move to implement the solution on a limited test basis, iteratively developing and adapting to any unforeseen challenges. After a successful functional prototype is developed, it is tested in-house using field personnel and/or an inner circle of sophisticated customers and gradually made more public as confidence in the product increases. In these early stages, many of the "kinks" of the process are smoothed out, developing the product format, look and feel; and adapting to initial feedback (*e.g.*, "please make the map labels larger") but, for the most part, keeping the initial vision intact.

There is no consistent formal procedure across agencies for certifying a new method or making a new product official. A product may be run and labeled "experimental" for one to two years in an evaluation period. The objectives and duration of the evaluation period are sometimes not formalized and one must just assume that if a product has been running for an extended period of time with no obvious problems, then it succeeds and the experimental label removed. Creating documentation of the product and process is often part of the transition from experimental to official, either in the form of an internal technical memo, conference proceedings or peer-reviewed journal article, if appropriate.

If the innovation involves using a tool or technique that supplements the standard suite of tools, some of the evaluation may involve running both tools in parallel and comparing their performance. Presumably, ease of use and low demand on resources are criteria for success (although the task of running models in parallel can, by itself, be a heavy demand on resources). Sometimes an agency may temporarily stretch its resources to accommodate the product for the evaluation period and if additional resources are not acquired by the end of the evaluation (for one of a number of reasons, some of which may not be related to the product but, rather, are due to variability in budgets), the product may be discontinued.

There is no consistent formal procedure across agencies for certifying a new method or making a new product official.



Sometimes skill is used to judge success, but this can be a very inefficient measure. This is because seasonal forecast skill varies greatly from year to year, primarily due to the variability of nature. Likewise, individual tools may perform better than other tools in some years but not others. In the one to two years of an evaluation period the new tool may be lucky (or unlucky) and artificially appear better (or worse) than the existing practice.

If the agency recognizes that a tool has not had a fair evaluation, more emphasis is placed on "hindcasting", using the new tool to objectively and retrospectively generate realistic "forecasts" for the last 20 to 30 years and comparing the results to hindcasts of the existing system and/or official published forecasts. The comparison is much more realistic and effective, although hindcasting has its own challenges. It can be operationally demanding to produce the actual forecasts each month (*e.g.*, the agency may have to compete for the use of several hours of an extremely powerful computer to run a model), much less do the equivalent of 30 years worth at once. These hindcast datasets, however, have their own uses and have proven to be very valuable (*e.g.*, Hamill *et al.*, 2006 for medium range weather forecasting and Franz *et al.*, 2003 for seasonal hydrologic forecasting). Oftentimes, testbeds are better suited for operationally realistic hindcasting experiments (Box 2.4).

During the evaluation period, the agency may also attempt to increasingly "institutionalize" a process by identifying and fixing aspects of a product or process that do not conform to agency guidelines. For example, if a forecasting model is demonstrated as promising but the operating system or the computer language it is written in does not match the language chosen by the agency, a team of contract programmers may rewrite the model and otherwise develop interfaces that make the product more user-friendly for operational work. A team of agency personnel may also be assembled to help transfer the research idea to full operations, from prototype to project. For large projects, many people may be involved, including external researchers from several other agencies.

During this process of institutionalization, the original innovation may change in character. There may be uncertainty at the outset and the development team may consciously postpone certain decisions until more information is available. Similarly, certain aspects of the original design may not be feasible and an alternative solution must be found. Occasionally, poor communication between the inventor and the developers may cause the final product to be different than the original vision. Davidson *et al.* (2002) found success in developing a hydrologic database using structured, iterative development involving close communication between users and developers throughout the life of the project. This model is in direct contrast to that of the inventor generating a ponderous requirements document at the outset, which is then passed on to a separate team of developers who execute the plan in isolation until completion.

2.6.2 Evaluation of Forecast Utility

As mentioned in Section 2.1, there are many ways to assess the usefulness of forecasts, one of which is forecast skill. While there are inherent limitations to skill (due to the chaotic nature of the atmosphere), existing operational systems also fall short of their potential maximum skill for a variety of reasons. Section 2.4 highlighted ways to improve operational skill, such as by having better models of the natural system or denser and more detailed climate and hydrologic monitoring networks. Other factors, such as improved forecaster training or better

BOX 2.4: What Role Can a "Testbed" Play in Innovation?

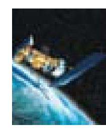
For an innovation to be deemed valuable, it must be able to stand on its own and be better than the entire existing system, or marginally better than the existing technology, if it is compatible with the rest of the framework of the existing system. If the innovation is not proven or believed likely to succeed, its adoption is less likely to be attempted. However, who conducts the experiments to measure this value? And who has the resources to ensure backwards-compatibility of the new tools in an old system?

This model lacks any direct communication between user and producer and leaves out the necessary support structure to help users make the most of the product (Cash *et al.*, 2006). Similarly, testbeds are designed as an alternative to the "Loading Dock Model" of transferring research to operations. A loading dock model is one in which scientists prepare models, products, forecasts or other types of information for general dissemination, in somewhat of a vacuum, without consulting with and/or understanding the needs of the people who will be using that information, with the anticipation that others will find these outputs useful.

Previously, a researcher might get a short-term grant to develop a methodology, and conduct an idealized, focused study of marginal operational realism. The results might be presented at research conferences or published in the scientific literature. While a researcher's career may have a unifying theme, for the most part, this specific project may be finished when publication is accomplished and the grant finishes. Meanwhile, the operational forecaster is expected to seek out the methodology and attempt to implement it, although, often, the forecaster does not have the time, resources or expertise to use the results. Indeed, the forecaster may not be convinced of the incremental advantage of the technique over existing practices if it has not endured a realistic operational test and been compared to the results of the official system.

Testbeds are intermediate activities, a hybrid mix of research and operations, serving as a conduit between the operational, academic and research communities. A testbed activity may have its own resources to develop a realistic operational environment. However, the testbed would not have real-time operational responsibilities and instead, would be focused on introducing new ideas and data to the existing system and analyzing the results through experimentation and demonstration. The old and new system may be run in parallel and the differences quantified. The operational system may even be deconstructed to identify the greatest sources of error and use that as the motivation to drive new research to find solutions to operations-relevant problems. The solutions are designed to be directly integrated into the mock-operational system and therefore should be much easier to directly transfer to actual production.

NOAA has many testbeds currently in operation: Hydrometeorological (floods), Hazardous Weather (thunderstorms and tornadoes), Aviation Weather (turbulence and icing for airplanes), Climate (ENSO, seasonal precipitation and temperature), and Hurricanes. The Joint Center for Satellite Data Assimilation is also designed to facilitate the operational use of new satellite data. A testbed for seasonal streamflow forecasting does not exist. Generally, satisfaction with testbeds has been high, rewarding for operational and research participants alike.



visualization tools, also play a role. This Section addresses the role of forecast evaluation in driving the technology development agenda.

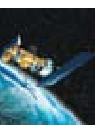
Understanding the current skill of forecast products is a key component to ensuring the effectiveness of programs to improve the skill of these products. There are several motivations for verifying forecasts including administrative, scientific and economic (Brier and Allen, 1951). Evaluation of very recent forecasts can also play a role in helping operational forecasters make mid-course adjustments to different components of the forecast system before issuing an official product.

Of particular interest to forecasting agencies is administrative evaluation because of its ability to describe the overall skill and efficiency of the forecast service in order to inform and guide decisions about resource allocation, research directions and implementation strategies (Welles, 2005). For example, the development of numerical weather prediction (NWP) forecasting models is conducted by numerous, unaffiliated groups following different approaches, with the results compared through objective measures of performance. In other words, the forecasts are verified, and the research is driven, not by *ad hoc* opinions postulated by subject matter experts, but by the actual performance of the forecasts as determined with objective measures (Welles *et al.*, 2007). The most important sources of error are identified quantitatively and systematically, and are paired with objective measures of the likely improvement resulting from an innovation in the system.

Recently, the NWS adopted a broad national-scale administrative initiative of hydrologic forecast evaluation. This program defines a standard set of evaluation measures, establishes a formal framework for forecast archival and builds flexible tools for access to results. It is designed to provide feedback to local forecasters and users on the performance of the regional results, but also to provide an end-to-end assessment of the elements of the entire system (HVSRT, 2006). Welles *et al.* (2007) add that these activities would be best served by cultivating a new discipline of "hydrologic forecast science" that engages the research community to focus on operational-forecast-specific issues.

Forecast evaluation should be more broadly defined than skill alone; it should also include measures of communication and understandability, as well as relevance.

While administrative evaluation is an important tool for directing agency resources, innovation should ultimately be guided by the anticipated benefit to forecast users. Some hydrologists would prefer not to issue a forecast that they suspect the user could not use or would misinterpret (Pielke, Jr., 1999). Additionally, evaluations of forecasts should be available and understandable to users. For instance, it might be valuable for some users to know that hydrologic variables in particular regions of interest lack predictability. Uncertainty about the accuracy of forecasts precludes users from making more effective use of them (Hartmann *et al.*, 2002). Users want to know how good the forecasts are so they know how much confidence to place in them. Agencies want to focus on the aspects of the forecast that are most important to users. Forecast evaluation should be more broadly defined than skill alone; it should also include measures of communication and understandability, as well as relevance. In determining these critical aspects, agencies must make a determination of the key priorities to address given the number and varied interest of potential forecast users. The agencies can not fully



BOX 2.5: The Advanced Hydrologic Prediction Service

Short- to medium-range forecasts (those with lead times of hours to days) of floods are a critical component of National Weather Service hydrological operations, and these services generate nearly $2 billion of benefits annually (NHWC, 2002). In 1997 the NWS Office of Hydrologic Development began the Advanced Hydrologic Prediction Service (AHPS) program to advance technology for hydrologic products and forecasts. This 16-year multi-million dollar program seeks to enhance the agency's ability to issue and deliver specific, timely, and accurate flood forecasts. One of its main foci is the delivery of probabilistic and visual information through an Internet-based interface. One of its seven stated goals is also to "Expand outreach and engage partners and customers in all aspects of hydrologic product development" (NRC, 2006).

Starting in 2004, the National Research Council reviewed the AHPS program and also analyzed the extent that users were actually playing in the development of products and setting of the research agenda (NRC, 2006). The study found that AHPS had largely a top-down structure with technology being developed at a national center to be delivered to regional and local offices. Although there was a wide range of awareness, understanding and acceptance of AHPS products inside and outside the NWS, little to no research was being done in early 2004 on effective communication of information, and some of the needs of primary customers were not being addressed. From the time the NRC team carried out its interviews, the NWS started acting on the perceived deficiencies, so that, by the time the report was issued in late 2006, the NWS had already made some measurable progress. This progress included a rigorous survey process in the form of focus groups, but also a more engaged suite of outreach, training, and educational activities that have included presentations at the national floodplain and hydrologic manager's conferences, the development of closer partnerships with key users, committing personnel to education activities, conducting local training workshops, and awarding a research grant to social scientists to determine the most effective way to communicate probabilistic forecasts to emergency and floodplain managers.

satisfy all users. The Advanced Hydrologic Prediction System (AHPS) of the NWS provides a nice case study of product development and refinement in response to user-driven feedback (Box 2.5).

There is another component to forecast skill beyond the assessment of how the forecast quantities are better (or worse) than a reference forecast. Thinking of forecast assessment more broadly, the forecasts should be evaluated for their "skill" at communicating their information content in ways that can be correctly interpreted both easily and reliably—*i.e.*, no matter what the quantity (*e.g.*, wet, dry, or neutral tercile) of the forecast, the user can still correctly interpret it (Hartmann *et al.*, 2002).

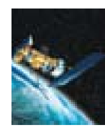
Finally, it seems important to stress that agencies should provide for user-centric forecast assessment as part of the process for moving prototypes to official products. This would include access to user tools for assessing forecast skill (*i.e.*, the Forecast Evaluation Tool, which is linked to by the NWS Local 3-month Temperature Outlook [Box 2.6]), and field testing of the

BOX 2.6: National Weather Service Local 3-Month Outlooks for Temperature and Precipitation

In January 2007, the National Weather Service made operational the first component of a new set of climate forecast products called Local 3-Month Outlooks (L3MO). Accessible from the NWS Weather Forecast Offices (WFO), River Forecast Centers (RFC), and other NWS offices, the Local 3-Month Temperature Outlook (L3MTO) is designed to clarify and downscale the national-scale CPC Climate Outlook temperature forecast product. The corresponding local product for precipitation is still in development as of the writing of this Product. The local outlooks were motivated by ongoing National Oceanic and Atmospheric Administration, NWS activities focusing on establishing a dialog with NWS climate product users <http://www.nws.noaa.gov/directives/>. In particular, a 2004 NWS climate product survey (conducted by Claes Fornell International for the NOAA Climate Services Division) found that a lack of climate product clarity lowered customer satisfaction with NWS CPC climate outlook products; and presentations and interactions at the annual Climate Prediction Application Science Workshop (CPASW) highlighted the need for localized CPC climate outlooks in numerous and diverse applications.

In response to these user-identified issues, CSD collaborated with the NWS Western Region Headquarters, CPC, and the National Climatic Data Center (NCDC) to develop localized outlook products. The collaboration between the four groups, which linked several line offices of NOAA (*e.g.*, NCDC, NWS), took place in the context of an effort that began in 2003 to build a climate services infrastructure within NOAA. The organizations together embarked on a structured process that began with a prototype development stage, which included identifying resources, identifying and testing methodologies, and defining the product delivery method. To downscale the CPC climate outlooks (which are at the climate division scale) to local stations, the CSD, and WR development team assessed and built on internal, prior experimentation at CPC that focused on a limited number of stations. To increase product clarity, the team added interpretation, background information, and a variety of forecast displays providing different levels of data density. A NWS products and services team made product mockups that were reviewed by all 102 WFOs, CPC and CSD representatives and a small number of non-agency reviewers. After product adjustments based on the reviews, CSD moved toward an experimental production stage, providing NWS staff with training and guidelines, releasing a public statement about the product and writing product description documentation. Feedback was solicited via the experimental product website beginning in August 2006, and the products were again adjusted. Finally, the products were finalized, the product directive was drafted and the product moved to an operational stage with official release. User feedback continues via links on the official product website <http://www.weather.gov/climate/l3mto.php>.

In general, the L3MO development process exhibited a number of strengths. Several avenues existed for user needs to reach developers, and user-specified needs determined the objectives of the product development effort. The development team, spanning several parts of the agency, then drew on internal expertise and resources to propose and to demonstrate tentative products responding to those needs. The first review stage of the process gave mostly internal (*i.e.*, agency) reviewers an early opportunity for feedback, but this was followed by an opportunity for a larger group of users in the experimental stage, leading to the final product. An avenue for continued review is built into the product dissemination approach.



communication effectiveness of the prototype products. Just as new types of forecasts should show (at least) no degradation in predictive skill, they should also show no degradation in their communication effectiveness.