THE ROLE OF CLIMATE VARIABILITY IN

OPERATIONAL WATER SUPPLY FORECASTING FOR

THE WESTERN UNITED STATES

by

Thomas Christopher Pagano

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF HYDROLOGY AND WATER RESOURCES

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2 0 0 5

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by **Thomas Christopher Pagano** entitled **"The Role of Climate Variability in Operational Water Supply Forecasting for the Western United States"** and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of **Doctor of Philosophy**

Soroosh Sorooshian Date: November 29,2004

Juan Valdez Date: November 29,2004

James Shuttleworth Date: November 29,2004

Don Davis Date: November 29,2004

Steven Mullen Date: November 29,2004

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Soroosh Sorooshian November 29,2004
Dissertation Director Date

STATEMENT BY AUTHOR

ACKNOWLEDGEMENTS

Writing a dissertation is, to say the least, a challenging experience that is at times inspiring and exhausting. Clearly without the support of others, such work is not possible.

I appreciate the useful feedback and guidance of my committee members: James Shuttleworth, Don Davis, Juan Valdez, and Steven Mullen. It took some time and distance to fully appreciate how privileged I was to work with these individuals.

I also would like to thank my friends and colleagues at the University of Arizona: Jiming Jin, Doug Boyle, Steven Fassnacht, Michael Thiemann, Ali Farid, Dave Gochis, Soni Yatheendradas, and Hamid Moradkhani. With my mother as a doctor at a time when few women entered into science, I admire the female hydrologists in the department: Kristie Franz, Terri Hogue, Shayesteh Mahani, Feyzan Misirli, Anne Kramer, Martha Whitaker, and Newsha Khodatalab. I appreciate graduate advisor Terrie Thompson's patience and help.

My officemate Holly Hartmann provided most of the ideas for my research, was a candid critic and an excellent sounding board. Similarly, I value my interactions with Bob Maddox and Katie Hirschboeck; their unending encouragement was an inspiration.

I am grateful to the larger scientific community, especially but certainly not limited to Roy Koch, Edwin Welles, George Leavesley, John Schaake, Tom Croley II, Martyn Clark, Lauren Hay, Alan Hamlet, Phil Mote, Dennis Lettenmaier, Andy Wood, Klaus Wolter, Tom Piechota, Dan Cayan, Mike Dettinger, Ed O'lenic, Huug van den Dool, Simon Mason, Dan Wilks, Kelly Redmond, Gregg Garfin and others.

Most of this research is directed towards the operational hydrology community. There are many people within the NRCS to thank including but not limited to Tom Perkins, Jolyne Lea, Jennifer Erxleben, Jim Marron, Chris Pacheco, Greg Johnson, Dan Murray, Richard Armijo, Mike Gillespie, Rick Eastlund, Paul Gallegos, Tony Tolsdorf, Mary Vigil, Nathaniel Todea, Larry Martinez, Randy Julander, Jerry Beard, Jon Lea, Ray Wilson, Tim Bardsley, Bob Nault, Ron Abramovich, Roy Kaiser, Rick McClure, Scott Pattee, Dan Greenlee, Sheila Strachan, Melissa Clark and Marianne Hallet. Water and Climate Services branch leader Phil Pasteris was kind enough to let me finish my research while working and Dave Garen was an excellent advisor away from the university. I also thank Dallas Reigle, Brenda Alcorn, Dave Brandon, Greg Smith, Jeff Smith, Bill Reed, Brent Bernard, Paul Greer, John Schmidt, Tony Anderson, Esther Vincent, Charlie Liles, Chris Cutler and Steve Vandiver.

What can be said to fully sum up one's appreciation for and admiration of one's advisor? Soroosh Sorooshian is a charismatic leader with an unparalleled management style. Although coming to work for him was something of an accident, I feel I will go places I would have never been able to go had I never met him.

Finally I would like to thank my family, Matteo, Ruth and Matthew Pagano, my partner A'lisha Leisek and my familiar Neko.

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS - Continued

LIST OF TABLES

ABSTRACT

The single greatest uncertainty in seasonal water supply forecasts is the amount of precipitation falling after the forecast issue date. There has been a long history of attempting to incorporate seasonal climate forecasts into operational water supply forecasts. The skill of these precipitation forecasts remains low especially compared to highly confident snow-based streamflow forecasts. Early in the season (e.g., September-December), however, large-scale climate indices are the best available predictors of future water supplies. This dissertation suggests practical methods for issuing climate-based operational streamflow forecasts.

This study also documents the existence of strong decadal trends in water supply forecast skill. Across the Western US, 1 April forecast skill peaked in the 1960-1970s and has been on the decline more recently. The high skill period was a very calm period in the Western US, with a near absence of extreme (wet or dry) spring precipitation events. In contrast, the period after 1980 has had the most variable, persistent, and skewed spring and summer streamflows in the modern record. Spring precipitation is also now more variable than it has been since at least the 1930s. This rise in spring precipitation variability in the Colorado/Rio Grande Basins and the Pacific Northwest is the likely cause behind the recent decline in water supply forecast skill.

1. INTRODUCTION AND MOTIVATION

Informed decisions about water resources are key to the sustainability of the Western US. As the region's population grows, legislation attempts to reduce dependence on non-renewable groundwater supplies, and environmental and tribal water rights gain recognition, stresses will be placed on already overcommitted surface water supplies. The vulnerability of this region to variability in water supplies will be higher in the near future than at any other time since major reservoirs were constructed. Therefore, it is imperative to manage these supplies with increasing sophistication.

The scientific community is strongly encouraged to provide useful information to decision-makers, which, in turn, may be used to improve natural resource management practices. Currently the primary interface between the hydroclimatic scientific community and Western US water managers is the operational Water Supply Outlooks (WSOs) issued cooperatively by the Natural Resources Conservation Service, the National Weather Service River Forecast Centers and other agencies where applicable (e.g., the Salt River Project in central Arizona). Some water management agencies are required by law to use these forecasts to determine releases from reservoirs for flood control and water supply. The question remains: what are the greatest opportunities for using state-of-the-art scientific information to improve these operational forecasts? One such opportunity is the incorporation of climate information and seasonal climate forecasts into the water supply forecasts.

Recent years have been an active period in climate research, particularly in the United States. Although seasonal climate forecasts have been issued in one form or another since the 1950s, it was not until the mid 1980s that researchers identified and diagnosed the impacts of large scale climate phenomenon, such as the El Niño/Southern Oscillation (ENSO), on hydroclimatic variability in the Western US. The climate community has improved its objective tools for quantifying and forecasting this variability with lead times of several months or more. The ability to monitor the ocean, atmosphere and land-surface has increased dramatically because of remote sensing and a high-density network of Tropical Ocean monitoring buoys. Likewise, General Circulation Models (GCMs), numerical descriptions of the earth's climate, are being run on some of the world's most powerful computers, at spatial resolutions several orders of magnitude higher than those run less than ten years ago. Finally, the study and practice of ensemble simulations have flourished, enhancing the ability to make reliable probabilistic forecasts of future conditions.

Coupled with these technical advances is a heightened societal awareness of the potential benefits of using climate forecasts in natural resource management. For example, while the impacts of ENSO on streamflow in Arizona have been relatively well known since the 1980s, it wasn't until the 1997-98 El Niño that the Salt River Project modified their reservoir operations in anticipation of a wet winter based on a climate forecast (Pagano et al. 2001). Public awareness of El Niño is very high; a LexisNexis general search of major newspapers (1972-2003) for “El Niño” returned over 11,000 matches, over 70% of which occurred in connection with or after the 1997-98 event.

There is increased organizational commitment within the National Ocean and Atmospheric Administration to facilitate the creation, distribution and use of climate forecasts. For example, many local interdisciplinary Regional Integrated Science Assessment (RISA) projects are conducting stakeholder driven research and are engaging resource managers about their use of climate forecasts (Hartmann 2001). The International Research Institute for Climate Prediction (IRI) is a new entity that uses research-grade simulation modeling technology to produce global climate forecasts (Barnston et al. 2003). The IRI supports an entire branch devoted to user interaction, ranging from hydropower operators in Taiwan to subsistence farmers in rural Africa. The recently formed National Weather Service Climate Services Division is devoted to the improvement, creation, dissemination and development of applications related to climate information and seasonal climate forecasts (NRC 2001). They also have an ambitious program to provide a broad array of NWS field office personnel with competence in the technical aspects of climate prediction and climate services. While precipitation and temperature prediction has been the focus of the climate community, streamflow prediction is a natural next step to increasing the societal relevance of climate information.

However, surveys of stakeholders and decision-makers have repeatedly found that there are significant barriers to more effective climate forecast use in natural resource management (e.g., Pulwarty and Redmond 1997; NRC 1999). Part of this is due to a lack of familiarity with the concepts and information involved - such unfamiliarity will undoubtedly diminish in time. Part of the barrier is due to concerns with the accuracy of the forecasts (addressed by Hartmann et al. 2002a). Lastly, and perhaps most intimidating and intractable to researchers are the institutional barriers associated with using forecasts.

Many natural resource management issues, particularly those related to water, are too complex to consider climate forecasts in isolation. For example, the management of water supplies and restoration of salmon stocks in the Pacific Northwest involves at least 20 different agencies (Pulwarty and Redmond 1997). Determining the precise adjustment of management practices in any given year based on a climate forecast would involve a prohibitively lengthy deliberative process to ensure that no party perceived it was being slighted under the guise of "better management through modern science".

This is not to say that water management agencies do not adjust practices in anticipation of interannual variability in surface supplies. As mentioned earlier, several water management agencies are required by law to use the WSOs in their operations. There is a specific and established course of action defined for particular water managers depending on the volume of anticipated streamflow (e.g Burke and Stevens 1984). However, when faced with both a climate and water supply forecast, it may be difficult for a water manager to reconcile them and determine the optimal course of action, or as one water manager put it succinctly, "the man with two watches never knows what time it is" (Pagano et al. 2001). In all fairness, some may counter that the man with two watches has a higher chance of having at least one watch with the correct time, but the discrepancy remains a concern for water managers.

Rather than attempting to have water managers use climate forecasts in an ad hoc way, it would be useful for the seasonal climate forecasting community to view the WSOs as the "front door" to positively affecting water management practices. It is mutually beneficial for all parties to integrate seasonal climate forecasts into the WSOs;

the WSOs may improve in accuracy and federal climate research funding agencies would be able to fulfill part of their mission involved with the transfer of research products to operations. This dissertation addresses this issue of linking climate information and operational seasonal water supply outlooks.

This dissertation also discusses some of the issues related to long term climate variability and trends. The scientific community and water managers are sounding a rising chorus of concern about natural and anthropogenically-induced climate change. Although long-range climate projections are uncertain, any potential impacts on water supply forecasting cannot be discounted. If the climate is changing, hydrologists should investigate any vulnerabilities in their water supply forecasting techniques. Climate stationarity (i.e., that the future will be like the past) remains a core assumption in the current water supply forecasting environment despite increasing evidence that the climate may not be stationary even on interannual and decadal timescales.

This research builds upon the works of previous authors who have outlined the theoretical underpinnings of linking climate and water supply forecasts. It also draws from past studies of climate variability and change in the Western US. The unique contribution of this dissertation is its attention to developing practical methods for climate-based streamflow forecasting in an operational setting, and testing these methods on a wide variety of basins throughout the Western US. It also highlights the social and institutional issues specific to the operational hydrology environment. Finally, this study explores previously unexamined aspects of climate change that are particularly relevant to water supply forecasting.

1.1 Objective

This study develops and tests practical methods of integrating climate forecasts and information into operational water supply forecasts for the Western US. This dissertation is a foundation document for developing climatological literacy among hydrologists who are interested in issuing very long lead seasonal water supply outlooks. It analyzes for forecasters and users the expected skill of these forecasts and develops a framework for effectively communicating highly uncertain forecast information. It explores and quantifies the role of climate variability in water supply forecast error, a first step towards reducing such error.

1.2 Structure of dissertation

Several different chapters within this dissertation use a variety of different hydroclimatic datasets. Therefore, all of the data used throughout are described in the following section (Chapter 2). Although this study attempts to be general, more in-depth analysis is instructive at certain junctures. Therefore, chapter 2.7 also describes the selection of 29 special basins used for in-depth analyses. Chapter 3 reviews the various climate phenomena and their teleconnections to the Western US. It summarizes past studies and fills gaps in the existing scientific literature with new analyses.

Chapter 4 begins with an overview of the major operational climate and water supply forecasts, specifically their format and interpretation. Sections 4.2.2 and 4.3.2

describe the evolution of operational-oriented research and compares climate and water supply forecasting techniques. This chapter provides the reader with an understanding of the current generation of products and a general flavor of how they may evolve in the future. Chapter 5 documents past efforts to link climate and water supply forecasts from the 1950s to present.

Chapter 6 analyzes the baseline performance of the current forecasts and identifies the major factors that contribute to their skill. This chapter is necessary in order to measure the relative improvement gained by using existing climate information and to explore the possible theoretical maximum skill that could be gained under ideal conditions.

Chapter 7 develops an objective system to mimic the behavior of the historical official water supply forecasts. This system is used to conduct sensitivity tests to measure the relative benefits, at various times of the year, of using climate information in a variety of ways.

Chapter 8 addresses the role of decadal climate variability in water supply forecasts. It begins by testing a method to account for decadal climate variability in the water supply forecasts (chapter 8.2). Sections 8.4-8.7 recognize and diagnose observed long term trends in water supply forecast skill and long term changes in observed streamflow variability.

Chapter 9 addresses the operational hydrologist, providing practical recommendations about the development of a climate-based streamflow forecast. It also identifies institutional and communication issues related to these highly uncertain probabilistic forecasts. It offers some alternative forecast formats that water supply forecasters may want to test.

Chapter 10 summarizes the above results and makes recommendations for future research.

Most of the material in this dissertation appears in several other publications by the author. Chapter 3, the review of past studies of the impacts of climate on the hydroclimatology of the Western US is contained in Pagano et al. (1999). Parts of chapter 3.6 on the impacts of climate warming on water supply forecasting are elaborated on in Pagano et al. (2004a). Chapters 5 and 9 on the history of efforts linking climate and water supply forecasts, and a discussion of practical aspects of generating and communicating operational climate-based streamflow forecasts is primarily based on Pagano and Garen (2005a). Chapter 6, the baseline evaluation of water supply forecasts corresponds to Pagano et al. (2004b). Finally, chapter 8 on the observed trends in streamflow variability and persistence is derived from Pagano and Garen (2005b).

Relevant figures are included at the end of each chapter. Figures are numbered sequentially within each chapter. For example, figure 4.3 is the third figure of chapter four and can be found at the end of chapter four, before the start of chapter five. Tables follow a similar numbering scheme but are embedded within each chapter, near to the relevant reference.

2. DATA

2.1 Introduction

This dissertation investigates the relationship between climate and the hydrology of the Western US. As such, several hydroclimatic datasets are needed to explore these relationships. Five primary variables are studied: snow water equivalent, precipitation, climate indices, streamflow, and water supply forecasts. This chapter describes the source and aspects of these data. Later chapters describe how these data are aggregated for analysis, in their specific context.

2.2 Snow water equivalent

In order to settle a dispute about the effects of logging on snowpack in the Tahoe Basin, Dr James Church began the first program of systematic western snow surveys in 1906. Snow surveys consist of repeated visits to snow courses, fixed locations where the snowpack is measured at many locations along a transect. The first few snow courses consisted of hundreds of samples along the entire slope of a mountain, although it was soon recognized that fewer (5-20) samples at "index" sites would suffice. Inspired by the design of a butter corer, the classics professor developed a snow sampler consisting of a long metal tube with a ring of cutting teeth on the end. After coring the snow, the tube is weighed, its length was measured and its snow water equivalent (SWE) and density calculated. Although the design is somewhat modified from the original, the "Federal Sampler" remains the standard snow measuring device at snow courses today.

SWE is the liquid water equivalent of the snow if the snow were completely melted. SWE is more relevant than snow depth to water supplies as the density of snow can widely vary, from less than 10% for new snow to as much as 60% for very ripe snow (Army Corps of Engineers 1956). At a limited number of inaccessible snow courses, "aerial markers" are used to visually measure the depth of snow while flying by in a small aircraft. SWE measurements at those sites are estimated using snow water density values derived from other nearby manual snow courses.

Beginning in the late 1970s the NRCS began replacing manual snow courses with automated SNOTEL (SNOw TELemetry) sites. SNOTEL sites typically consist of a large storage ("rocket") precipitation gage and a large metal or rubber pillow filled with a non-freezing liquid. Overlaying snow exerts pressure on the pillow, as measured by an electronic transducer. This pressure measurement is then converted into SWE. The snow is sampled sub-daily and the data telemetered via a meteorburst system to one of several master stations around the Western US. Typically the midnight reading is used to indicate the daily value.

It is not the place of this dissertation to argue the merits of SNOTEL sites versus snow courses. Some hydrologists believe that the benefit gained from having daily (as opposed to monthly) snow data at SNOTEL sites is somewhat offset by a degradation in the quality of the snow measurement caused by sensor drift, diurnal temperature fluctuations, and malfunction. In many places, a SNOTEL site replaced an existing snowcourse; several years of overlapping measurements often permitted linear-regression back-estimation of the monthly SNOTEL data, if a stable relationship existed between the snowcourse and SNOTEL measurements. This study used back-estimated monthly SNOTEL data as it existed in the NRCS CDBS database.

Snow water equivalent data were obtained from the NRCS website at ftp://ftp.wcc.nrcs.usda.gov/data/snow/snow_course/. This website contains snow course data, aerial marker data, SNOTEL and back-estimated SNOTEL data. Recent publications using SNOTEL and snow course data include, among others, Bohr and Aguado (2001), Cayan (1996), Clark et al. (2001), Fassnacht et al. (2003), McCabe and Dettinger (2002), McCabe and Legates (1995), and Serreze et al. (1999, 2000). Close to 3000 unique sites have existed in the Western US and Alaska, approximately 2150 of which have 20 or more years of available data. Of the sites with 20 or more years of data, 674 lie within the 35-km buffer around the study basins used herein (as described in chapter 7.2).

A battery of internal consistency and data quality tests was applied to the data values, as described in Table 2.1. If a data value failed any of the tests, the entire year's worth of data for this station was removed from the analysis. In total, of the over 135,000 station-years of data in the Western US and Alaska, about 1.7% failed one or more of the tests. There was no obvious spatial pattern to the flags raised. The analysis favored the removal of snow course data over SNOTEL data because, for example, no depth measurements are taken at most SNOTEL sites and thus there was no opportunity to fail the density test. Additionally, SNOTEL sites are always measured on exactly the first of the month and cannot have an internally inconsistent measurement date.

Table 2.1. Snow measurement data quality tests. Data must pass all tests to be used in the analysis. If any month's data is flagged, the entire year for that site is disqualified.

Description
Is the entry not malformed? (i.e., not characters or symbols where integers are expected)
Is the snow depth a contiguous, right justified integer (e.g., not "1 4")?
Is the snow density from January-April between 7.7% and 60%? In May and June is the snow density between 10% and 60%?
Is the SWE physically realistic (i.e., positive and not extraordinarily large)
If the data value is identified as a "first of the month" measurement, is the actual measurement date of the snow course within ~11 days of the first day of that month?
Other: there is a prevalence of sites whose SWE measurements are exactly 99.9. These data are disqualified.

Figure 2.1 shows a time series of the number of valid snow measurements for various months. Few sites existed before the 1930s and the network steadily grew at 30-40 sites per year from the 1950s to the 1980s. The sharp increase in sites in 1961 is an artifact of the efforts to back-estimate data for snow courses installed after 1961; personnel in Idaho, Montana and northwest Wyoming were seeking a serially complete dataset with which calculate the 1961-90 "normal". The installation of new sites slowed in the 1980s due to shifting priorities and limited financial resources. The sharp drop after 1989 represents the "snow course reduction plan", a program to discontinue snow courses in favor of automated SNOTEL sites. The number of samples measured at each site was

also reduced (from ~20-40) to where a typical snow course today involves 5 measurements.

This analysis does not include “legacy” snow courses, those discontinued before the 1980s. In the early 1980s, forecast activities were collected into a national center in Portland (the National Water and Climate Center), as opposed to being state based. One of the Center’s first tasks was the development of a centralized digital database. Given computer storage and time constraints, the decision was made to digitize only those snow courses relevant to water supply forecasting (i.e., those that were still active). Although an exact measurement is unavailable, roughly 20% of the monthly snow course records remain undigitized in paper copies at the Portland office. Previous generations of hydrologists would have had access to these data and would have used them in the water supply forecasts issued before 1980.

Figure 2.1 shows that snow course measurements are taken more often in certain months than others. Snow courses are most commonly measured in March and April, less so in February and least in January. May measurements are about as frequent as February measurements and June data are as common as January data (not shown). SNOTEL sites, in contrast, provide daily measurements year-round.

Figure 2.2 is a map of all of the snow measurement sites in the NRCS network. Sites tend to cluster around mountain ranges, away from valley bottoms and low elevation regions. Most sites are located in regions of maximum snow accumulation versus those with ephemeral snowpack or near the snow line.

2.3 Precipitation

The NWS maintains a network of precipitation and temperature instruments measured by cooperative observers (the "CO-OP network"). Over 18 000 stations have existed in the United States although the network is comparatively sparse in the Western US. These sites are typically located where people live in cities at relatively low elevations in valley bottoms. The Western US is actually more "urbanized" than many eastern US states; its vast empty landscapes are punctuated with high density population centers as opposed to evenly spread continuous stretches of medium density suburban communities. Therefore fewer co-op stations tend to be sited in "water supply relevant" locations such as headwaters and mountains. With SNOTEL sites located specifically at high elevations and wilderness areas, the two networks are complementary (Redmond 2003).

Co-op data were obtained from the National Climatic Data Center (NCDC) internet site http://hurricane.ncdc.noaa.gov/pls/plclimprod/somdmain.somdwrapper?datasetabbv=TD3220&countryabbv=&GeoRegionabbv=&Forceoutside=. This dataset is known as the "TD3220" and it contains daily precipitation depths (from "TD3200") that have been accumulated into monthly values with additional quality control applied. Reek et al. (1992) developed a quality control procedure that flagged nearly 400 000 data discrepancies. NCDC often estimated data to complete months of data that were partially missing. Given the extensive sophisticated spatial, temporal and cross-sensor quality controls done by NCDC, no additional quality control was performed for this study.

Figure 2.3 is a time series of the number of precipitation sites within the 35 km boundary of the study basins (as described in chapter 7.2). A site is counted if it has at least one valid measurement in a calendar year. Roughly, these sites represent about 3-5% of the entire COOP network, which had 8232 sites in 2002. Few sites existed before the 1930s (or NCDC may not have digitized that data). The network steadily expanded until its peak in the 1950s and has steadily contracted ever since. Figure 2.4 is a map of the COOP network. Sites along coastal and Northern California are not shown in this map because these regions were not included in this study. Sites tend to cluster around major urban centers; Salt Lake City, Phoenix, Denver and Albuquerque are readily evident.

Precipitation measurements have historically been made year round for all months, as opposed to snow measurements which have a historical abundance of 1 April measurements over 1 January measurements. It is generally perceived that frozen precipitation is less accurately measured than liquid rainfall although Eischeid et al. (2000) show that spatially heterogeneous convective summer rainfall is not measured well either.

In the chapter describing the correlation between various climate indices and precipitation (Chapter 3) a subset of high quality stations are used. These “Global HydroClimatic Data Network” sites have had additional quality control tests applied to them (Peterson and Vose 1997). These data are available at http://lwf.ncdc.noaa.gov/cgi-bin/res40.pl

Finally, spatially distributed climatological average precipitation data are obtained from the PRISM system. This hybrid statistical-geographical approach blends information about topography and point estimates of precipitation from 1961-1990 to derive gridded fields of average annual precipitation at a four-km resolution for the entire US (Daly et al. 1994). The project was funded and coordinated by the National Water and Climate Center of the NRCS and data are available on the Internet at http://www.ftw.nrcs.usda.gov/prism/prism.html

2.4 Climate indices

Atmospheric behavior is highly complex with many microscale transient weather features. However the atmosphere also has several large-scale features that slowly vary and persist over many months. Climatologists use global Sea Surface Temperature (SST) or atmospheric pressure data sets to describe these long-lived features, and derive time series indices to describe the behavior of the climate features from year to year. Seven of the most commonly encountered climate indices are used in this study, as listed in Table 2.2.

This roster of indices captures what could be considered the mainstream of current climate science. The El Niño/Southern Oscillation (SOI, Niño3.4) is a well-known and well-described phenomenon with no less than 4,500 citations in the scientific literature (source: Meteorological and Geoastrophysical Abstracts search on “El Nino”). The PNA index has found use for more than 20 years, as has the North Atlantic

Table 2.2, Climate indices used in this study

Index	Name	Period of Record	Describes
CPCnao[1]	CPC North Atlantic Oscillation	1950-2002	North Atlantic pressures (Barnston and Livezey, 1987)
PNA[1]	Pacific North American Pattern Index	1950-2002	North Pacific and North America pressures (Barnston and Livezey, 1987)
SOI[1]	Southern Oscillation Index	1900-2002	Tropical Pacific pressures at Tahiti and Darwin (Walker and Bliss, 1934)
WP[1]	West Pacific	1950-2002	Western Pacific pressures (Barnston and Livezey, 1987)
NAO[2]	North Atlantic Oscillation	1900-2002	North Atlantic pressures at Gibraltar and Stykkisholmur (Jones et al. 1997)
Niño3.4[1,3]	El Niño 3.4 Index	1900-2002	Tropical Pacific Ocean temperatures (Ropelewski and Halpert, 1987)
PDO[4]	Pacific Decadal Oscillation Index	1900-2002	North Pacific Ocean temperatures (Mantua et al., 1997)

1. Source: Climate prediction center ftp://ftpprd.ncep.noaa.gov/pub/cpc/wd52dg/data/indices/
2. Source: East Anglia Climate Research Unit http://www.cru.uea.ac.uk/cru/data/nao.htm
3. Source: Columbia University http://ingrid.ldgo.columbia.edu/SOURCES/.Indices/.nino/
4. Source: University of Washintgon ftp://ftp.atmos.washington.edu/mantua/pnw_impacts/INDICES/PDO.latest

Oscillation. The Pacific Decadal Oscillation is a relatively new index, although low frequency North Pacific indices (such as WP) were identified earlier by Barnston and Livezey (1987) and Wallace and Gutzler (1981). The PDO index is somewhat controversial in that official climate forecasters do not explicitly consider it in their forecasts, citing studies (e.g., Newman et al. 2003) that the PDO is not a "true" independent oscillation but rather an artifact of low-frequency variability associated with

El Niño. Several of these indices contain redundant information, such as Niño3.4 and SOI. Table 2.3 documents the cross-correlation among various indices computed over their common periods (1900-2002 or 1950-2002).

Table 2.3 Cross correlation coefficient of September-November values of the various climate indices used in this study

	CPC_PNA	SOI	WP	NAO	Niño3.4	PDO
CPC_NAO	-0.10	-0.13	-0.07	+0.50**	+0.09	-0.05
PNA		-0.20	+0.14	+0.08	+0.21	+0.29*
SOI			-0.21	+0.17	-0.78**	-0.39**
WP				-0.07	+0.22	+0.50**
NAO					-0.11	-0.06
Niño3.4						+0.48**

* <5% significance level ** <1% significance level for a student's T-test

2.5 Streamflow

The United States Geological Survey (USGS) has collected streamflow data at almost ten thousand sites across the Western US (west of 104 west longitude). These locations are generally critical to water management operations (e.g., inflows to major reservoirs, interstate compact points, major diversions).

Many forecast points are regulated, in that the observed streamflow is significantly altered by human activity such as irrigation diversions and reservoir releases. Naturalization of streamflow values to remove human influences is a difficult task, and even the best efforts cannot completely remove human effects. In reality, there are differences between true natural flow and unregulated flow data (which account for a limited number of measured reservoirs and losses). As a result of these complications, regulated streamflow locations were avoided in this study (see also Dracup et al. 1985).

Slack and Landwehr (1992) identified a subset of Hydro-Climatic Data Network (HCDN) streamgages as being relatively free of significant human influences and therefore appropriate for climate studies. In the continental Western US, there are 481 such points west of 104 west longitude (figure 2.5). Excluding Alaska, 151 of the HCDN gages are currently water supply forecast locations. All of the streamgages chosen for this study are HCDN locations. The HCDN gages were identified in 1992, and this study assumes that these locations remain unregulated; this is a relatively safe assumption considering the diminishing number of new large water projects since the 1970s. Monthly streamflow data were obtained from the US Geologic Survey online database (http://waterdata.usda.gov/nwis/sw).

For the 29 study basins (see chapter 2.7), missing observed data are estimated by linear regression between the streamflow for the location of interest and data from upstream or nearby streamgages. Only 1.1% of the observed streamflow data values needed to be estimated, and based on the strength of the correlation coefficients of the regression equations, the estimated values are likely to differ less than 5% from the true values. Therefore, the estimation procedure should not significantly affect the following analysis. This study assumed that the differences between the forecasts and observations were entirely due to forecast error and were in no part due to the quality of the observations. The USGS estimate that approximately 95% of their daily discharge measurements are within 10% of the true value.

Adjusted (unregulated) streamflow data were used only at one location, the Tongue River near Dayton, Wyoming. Although an HCDN location, significant diversions for irrigation occur during the summer months. The accounted-for diversion (Highline Ditch nr Dayton, Wyoming, USGS station number 06297500) amounts to 3-5% of the seasonal streamflow volume. In the 28 other study basins, the NRCS currently calibrates its statistical forecast equations using observed flow data, without any adjustments.

2.6 Water supply forecasts

The contents of and the methodologies to create water supply forecasts are described in chapter 4.3. This chapter describes the forecasts archived for this study.

The 4841 historical forecasts used in this study were drawn from a variety of existing sources. The primary source of forecasts was paper versions of the historical state "Basin Outlook Reports" and the "Water Supply Outlook for the Western United States", housed at the NRCS National Water and Climate Center, from which the values were manually digitized. Forecasts after 1990 were available in electronic versions of the same reports. A secondary source of forecasts was the NRCS Forecast Error Analysis Routine (FEAR) electronic database, as used by Shafer and Huddleston (1984). Third, the University of Arizona Department of Hydrology maintains an electronic archive of water supply forecasts for the Colorado River Basin, as they appeared in NWS publications. Finally, paper archives of the publication "Runoff Forecasts", in the *Western*

Construction News (1947-1954) were used to obtain a very limited number of early forecasts in the 1940s and 1950s.

The NRCS, as an entity, has produced seasonal water supply outlooks since mid 1930s, although some pre-NRCS forecast activities started in the 1910s and earlier. The number of WSO forecast points has increased dramatically over the years. In 1922, a limited set of forecasts was available in California and Nevada. After the mid-1930s, forecasts increased in number steadily, adding a net 11 forecast locations per year on average. Figure 2.6 (top) shows a map of the current roster of Water Supply Outlook locations. Before the 1940s, water supply forecasts were almost exclusively issued after 1 April. In time, the demand for longer lead-time information grew. The NRCS began issuing March forecasts in the early 1950s, with the start of February forecasts following in the 1960s. January forecasts outside of Arizona began at the NRCS in 1980 (figure 2.7). According to these trends, the advent of December forecasts almost seems overdue.

The calibration errors of the NRCS water supply forecast regression equations have been used to compute the confidence intervals corresponding to 10%, 30%, 70%, and 90% exceedance probabilities associated with the median (50% exceedance probability) forecasts. These five probability bounds have appeared in NRCS publications since 1989. Before 1986, only median forecasts were published. The publications also generally included values for the historical (e.g., 30-year) average streamflow for each basin along with the median forecast streamflow as a percent of the historical average. For the remainder of this text, “forecast” will refer exclusively to the median forecast. In some publications, this value is referred to as the “most probable” forecast, although this is not a statistically rigorous term and is not the preferred terminology.

While the deterministic forecast is the focus of this study, Blanchard (1955) and many others since have demonstrated that an optimal decision maker, e.g., an irrigator or reservoir operator, gets more value from a probabilistic forecast than a deterministic forecast. The author recognizes this issue and believes that the deterministic evaluation here is a positive first step towards a fuller probabilistic evaluation. Given that the forecasts were developed using statistical tools, and the forecast distribution width was proportional to expected forecast skill, a probabilistic evaluation should not paint a radically different picture from this analysis unless an improper shape was assumed for the forecast distribution. In comparison, simulation models have a well-known tendency to produce overconfident forecasts with narrow forecast distributions in part because they do not account for the uncertainty due to model calibration and data errors (Barnston et al. 2003). Such overconfidence would be penalized in a probabilistic evaluation.

Many forecast points had multiple target seasons. For example, forecasters predicted the April-June, April-July, and April-September flow volume for the Big Lost River, Idaho to serve the needs of different users. Before the 1950s, forecasts almost exclusively had a target period of April-September, the period that corresponds to the irrigation and snowmelt season around most of the Western US. In recent years, to isolate the effects of the relatively unpredictable summer monsoon, Upper Colorado Basin forecasts have been for April-July. Other locations may begin snowmelt earlier, such as the Pecos River in New Mexico, which had a forecast target of March-July. Arizona forecasts were unique in that the target period shrank throughout the season. In January, the forecast target was January-May, in February it was February-May and so on until April-May. Arizona also has a long history of issuing mid-month forecasts (e.g. issued March 15, predicting March 15- May flow) although these forecasts were not evaluated in this study.

For this study, some forecasts' target seasons were changed by multiplying a forecast for a different target season by the ratio of the long-term average flow for the target seasons, as published at the time. For example, multiplying the April-September forecast by the April-July long-term average and dividing by the April-September long-term average creates an estimated April-July forecast. This technique was chosen because it preserved the forecast as a percent of average, and forecasters commonly developed a forecast for one target period and applied the percent of average to the other periods. In some situations in this study, concurrent averages for different periods were not available, preventing such a transformation. In these cases, the forecasts were estimated using regressions between the observed streamflow values for the various target periods, excluding the observed flow for the year the forecast is being estimated. In a very limited number of cases, a streamgage had been permanently moved to a nearby location within the basin, and these forecasts were adjusted to remove the effect of changing the gage. Of the 4841 unique forecast values, 13% were estimated by one of the means just described. Almost half of the estimated values were on the Weber, Pecos, and Beaver Rivers because of the changing target periods throughout their history. Of these target period changes, the Pecos was the most uncertain, with an R^2 = 0.952 relationship between

March-July and April-September flows. When an estimated value appeared obviously out of line with what a forecaster reasonably would have issued, the forecast was listed as missing. For example, if a forecast was being estimated using forecasts from a nearby basin and there was an unusual sharp gradient in snowpack between the basins, a hydrologist would have been aware of this gradient and would have reflected this difference in the forecasts.

In many instances, forecasts were cross-checked for consistency among multiple sources. The most common discrepancies were due to keying errors. Discrepancies were resolved on a case-by-case basis, almost always favoring the value that appears in a paper publication. Based on the frequency of discrepancies discovered (and corrected), the author estimates that at least 99.6% of the forecast values used in this study were identical to the actual forecast. In the instances where forecasts from the NRCS and NWS disagree, the NRCS forecasts were used. Visual inspection of forecast and observation time series and maps ensured that any remaining data entry errors were not gross enough to affect the following analysis significantly.

2.7 Selection of study basins

All of the selected study basins are HCDN basins and, as such, are free from significant human regulation. Human regulation adds an unpredictable element of noise to the flow data; forecasts in heavily regulated basins are expected to be less skillful than forecasts in unregulated basins. The difference in water supply skill is likely to be inversely proportional to the quality of the naturalized flow data and the extent to which most of the regulations are accounted for. Selecting HCDN basins for this study was necessary to isolate and emphasize meaningful climatic signals. Operationally, the use of poor quality unregulated flow data may obscure this information.

To assess possible trends in forecast accuracy, a long history of forecasts is necessary, limiting the number of basins eligible for analysis. Generally, the selection of basins for this study favored those with a continuous record of forecasts and observations from 1955-2002, with streamgages that are still active today. One may assume that basins with a long period of record of forecasts also have many years of historical streamflow and snowpack data. Data-rich basins have better forecasts than, for example, basins with less than 10 years of historical streamflow data, where it is difficult to reliably estimate the relationship between snowpack and future streamflow. Although they are very rare, forecasts on ungaged basins are the least reliable.

To ensure relatively complete geographic coverage and a range of basin sizes and types, some basins with shorter forecast records were selected (e.g., the Sandy River near Marmot, Oregon, 1971-2002). Alaskan forecasts were omitted because of their short period of record and compressed forecasting season (i.e., they are only issued in March, April, and May). Finally, the number of basins chosen was limited by the resources available to digitize the historical forecasts manually. Table 2.4 details the characteristics of the 29 forecast points used in this study, and their locations are shown in Figure 7.2.

Table 2.4 Study basins and their characteristics. Latitude and longitude are the location of the USGS streamgage. Sites with circles by the name are log-transformed (see chapter 6.3.2). Sites with diamonds have shrinking forecast target seasons. See also figure 7.2.

USGS SITE NAME	USGS Code	Lat (North)	Lon (West)	Basin Area (km^2)	Forecast Target Season
1 Yellowstone, MT	06191500	45.1	110.8	6,794	Apr-Sep
2 Clarks Fk Yellowstone, MT	06207500	45.0	109.1	2,989	Apr-Sep
3 Tongue nr Dayton, WY	06298000	44.9	107.3	528	Apr-Sep
4 Pecos nr Pecos, NM	08378500	35.7	105.7	490	Mar-Jul
5 East at Almont, CO	09112500	38.7	106.9	749	Apr-Sep
6 Green at Warren Bridge, WY	09188500	43.0	110.1	1,212	Apr-Sep
7 White nr Meeker, CO	09304500	40.0	107.9	1,955	Apr-Sep
8 Animas at Durango, CO	09361500	37.3	107.9	1,792	Apr-Sep
9 Little Colorado, AZ	09384000	34.3	109.4	1,829	Jan-Jun
10 Virgin at Littlefield, AZ	09415000	36.9	113.9	13,183	Apr-Jun
11 San Francisco at Clifton, AZ	09444500	33.1	109.3	7,164	Jan-May
12 Salt nr Roosevelt, AZ	09498500	33.6	110.9	11,153	Jan-May
13 Verde blw Tangle Creek, AZ	09508500	34.1	111.7	15,175	Jan-May
14 Weber nr Oakley, UT	10128500	40.7	111.3	420	Apr-Sep
15 Beaver nr Beaver, UT	10234500	38.3	112.6	236	Apr-Jul
16 West Walker nr Coleville, CA	10296000	38.4	119.5	469	Apr-Jul
17 Carson nr Ft. Churchill, NV	10312000	39.3	119.3	3,372	Apr-Jul
18 Lamoille Ck nr Lamoille NV	10316500	40.7	115.5	65	Apr-Jul
19 Martin Creek, NV	10329500	41.5	117.4	454	Apr-Jul
20 Dungeness nr Sequim, WA	12048000	48.0	123.1	404	Apr-Sep
21 North Fork Flathead, MT	12355500	48.5	114.1	4,009	Apr-Sep
22 Stehekin at Stehekin, WA	12451000	48.3	120.7	831	Apr-Sep
23 Big Lost at Howell Ranch, ID	13120500	44.0	114.0	1,166	Apr-Sep
24 Bruneau nr Hot Spring, ID	13168500	42.8	115.7	6,812	Mar-Sep
25 Malheur nr Drewsey, OR	13214000	43.8	118.3	2,357	Apr-Sep
26 Salmon at Whitebird,ID	13317000	45.8	116.3	35,094	Apr-Sep
27 Umatilla nr Gibbon, OR	14020000	45.7	118.3	339	Apr-Sep
28 Sandy nr Marmot, OR	14137000	45.4	122.1	681	Apr-Sep
29 Rogue abv Prospect, OR	14328000	42.8	122.5	808	Apr-Sep

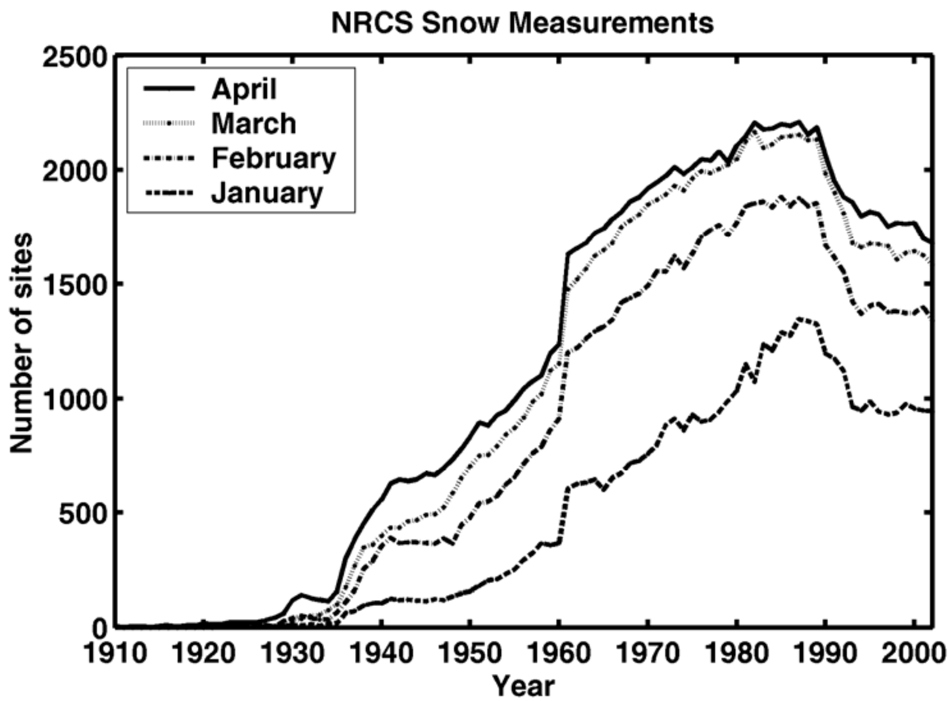


Figure 2.1. Time series of NRCS snow measurements by year and by measurement month.

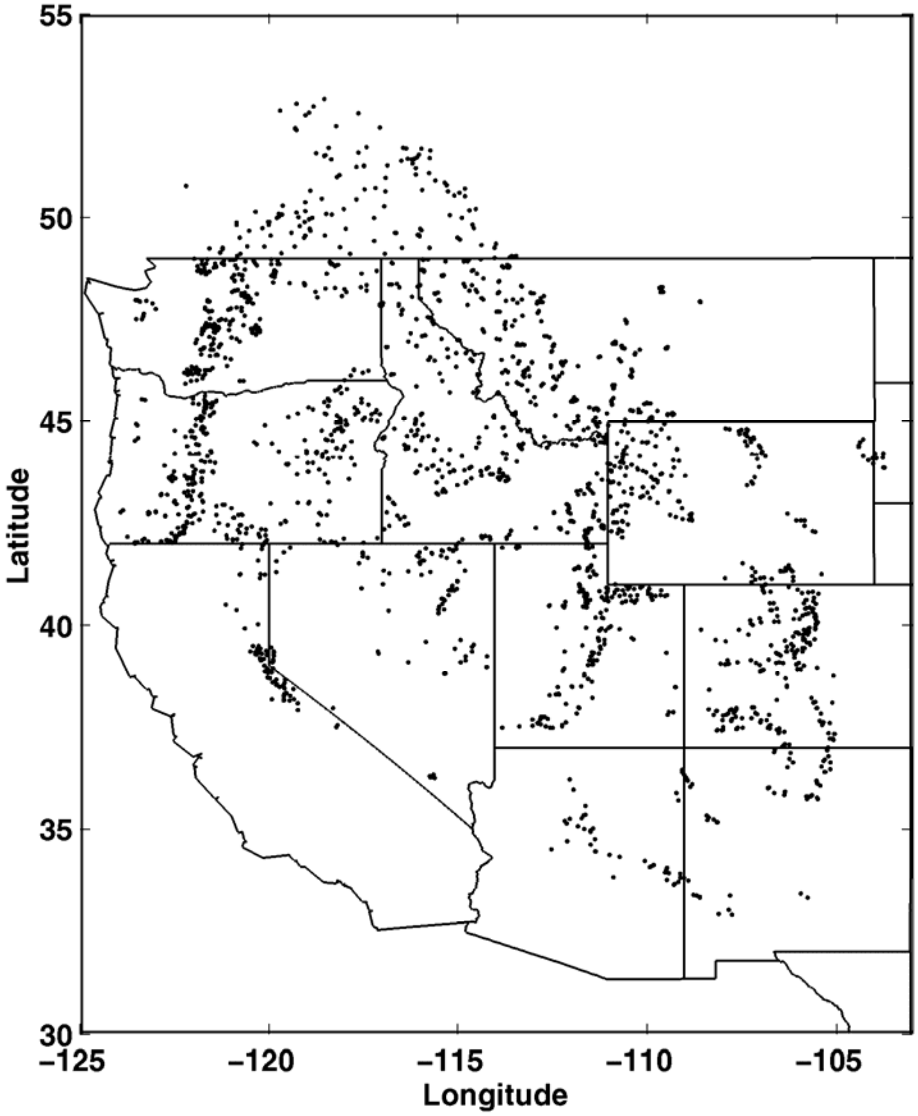


Figure 2.2. Map of NRCS snow measurement network. This map includes both snow course and SNOTEL sites as well as sites discontinued after 1980.

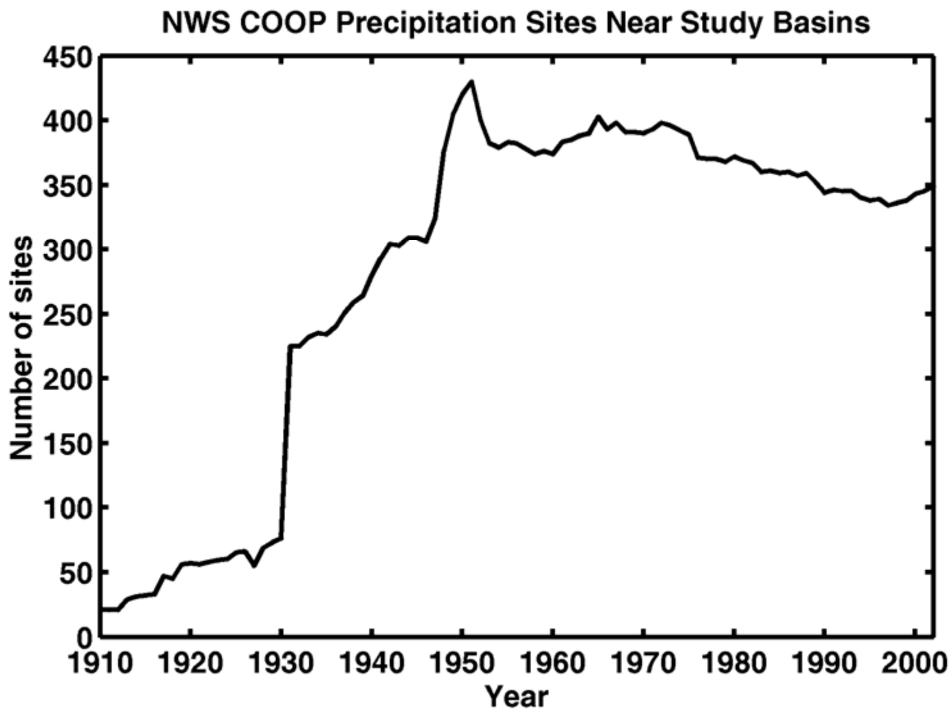


Figure 2.3. Time series of NWS cooperative observer (COOP) precipitation sites within 35 km of the study basins described in section 2.7 and shown in figure 7.2. A site is counted if it has at least one valid measurement in a calendar year. These sites represent approximately 3-5% of the entire United States COOP network, which had 8,232 sites in 2002.

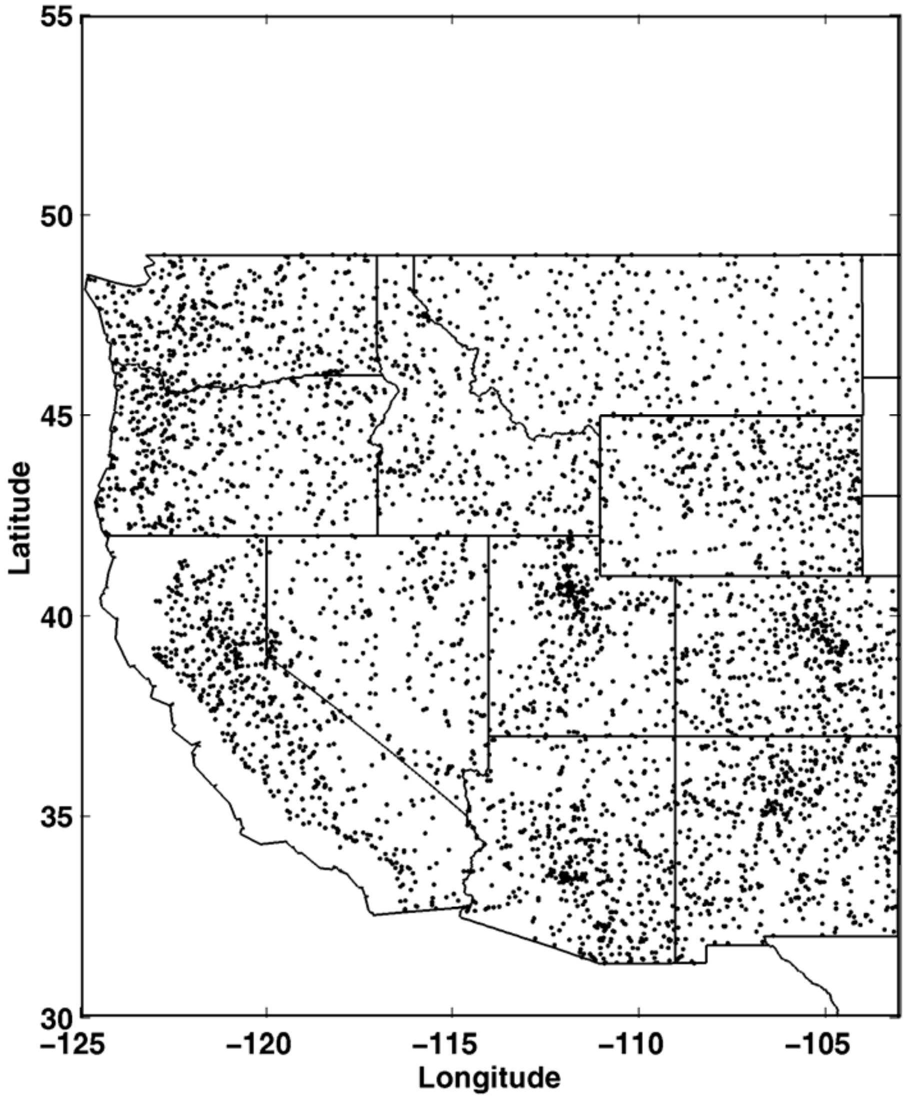


Figure 2.4. Map of the NWS cooperative observer (COOP) precipitation network. Sites along coastal and Northern California are not shown because they are not used in this study. The density of sites in that region is typical of western Oregon.

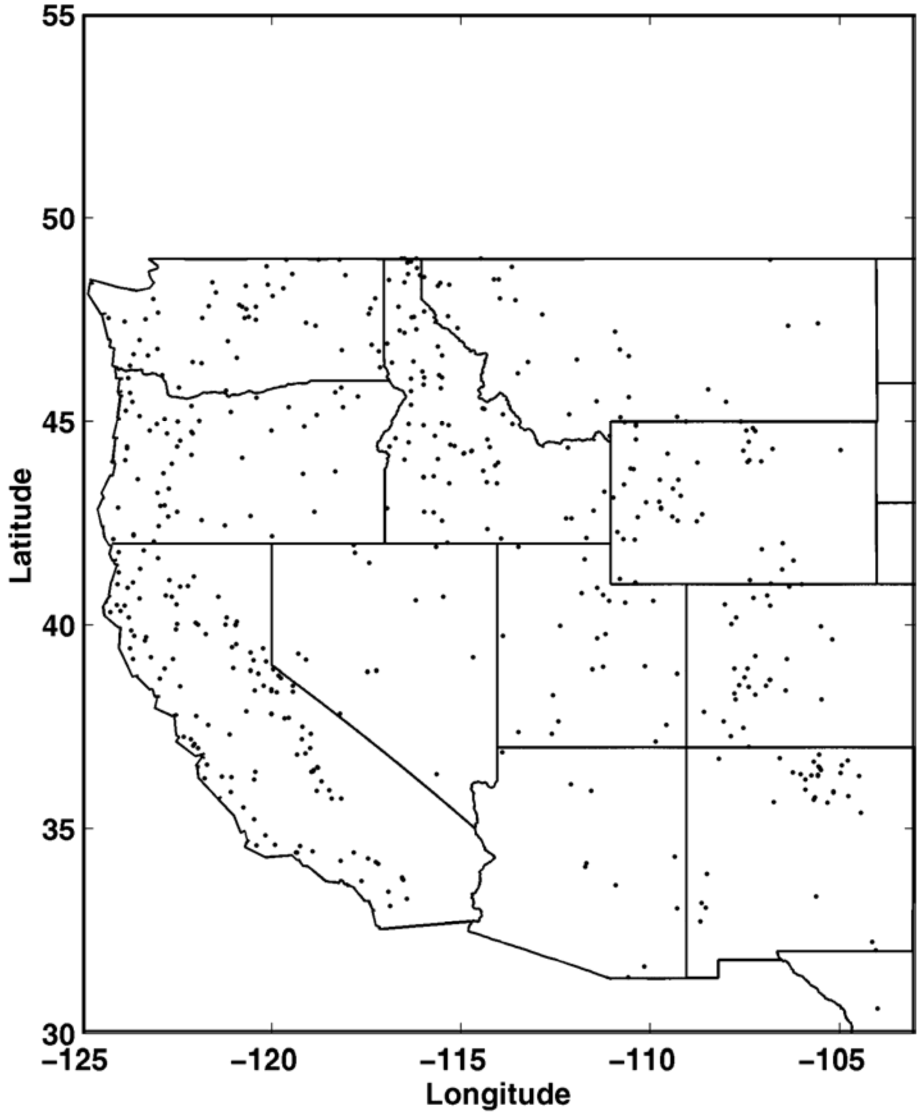


Figure 2.5. Map of the USGS Hydroclimatic Data Network for streamflow. These sites are a subset of the fuller network of the nearly 10,000 Western US USGS stations. These locations are unaffected by significant human regulation and have not experienced major land use change.

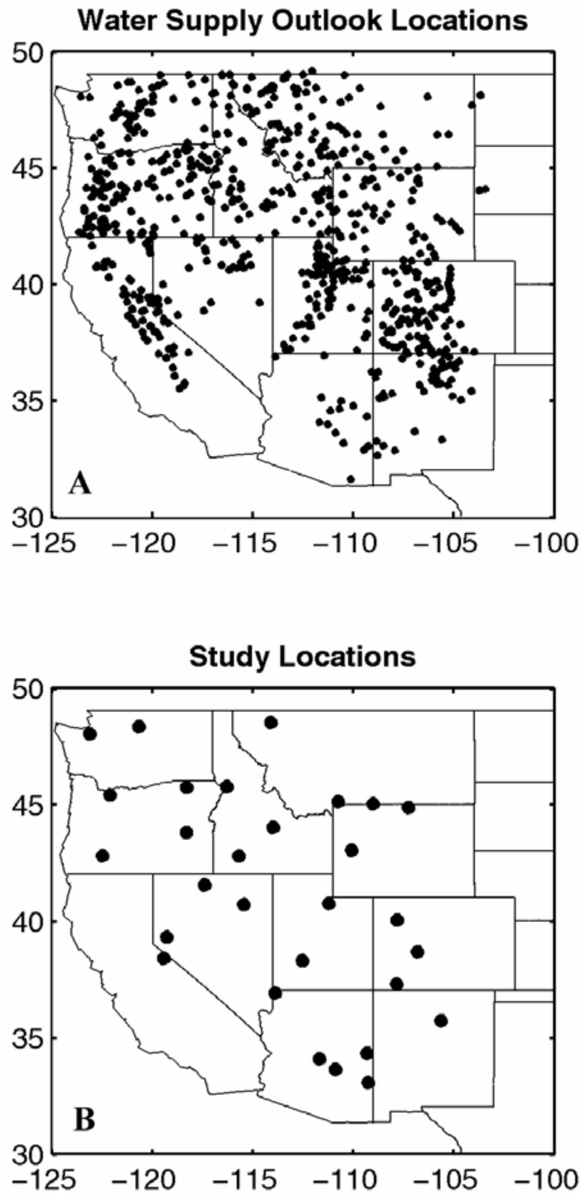


Figure 2.6 Top: Map of water supply forecast locations in 2003. Bottom, map of study basins used in this study. See also figure 7.2.

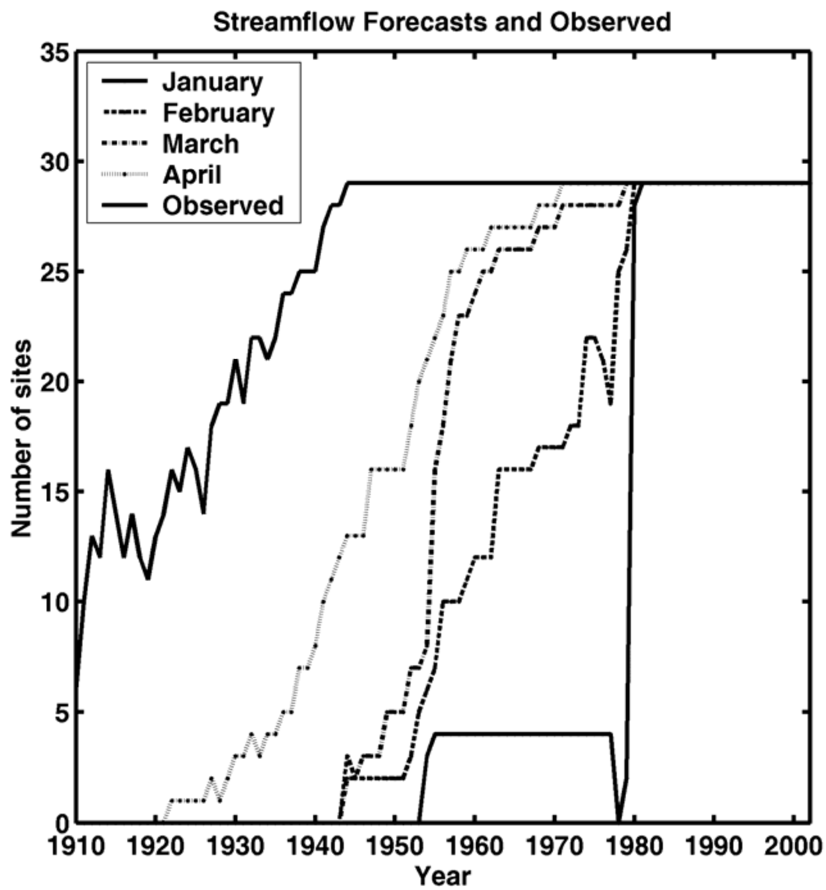


Figure 2.7 Time series of available forecasts and observations for the 29 study basins described in section 2.7. The top most solid line represents the number of observations available and the bottom four lines represent the forecasts, separated by forecast issue month. Forecasts and observations are serially complete beginning in 1980.

3. CLIMATE TELECONNECTIONS AND THE WESTERN US

3.1 Introduction

This chapter provides an overview of the relationship between large-scale climate variability and the hydroclimatology of the Western US. It focuses primarily on the surface hydrology (namely precipitation and streamflow) of the region, although undoubtedly other climate aspects have the potential to affect water management (such as temperature and rain/snow partitioning). However, the intent of this chapter is to develop a general understanding of the region's primary teleconnections, and establish a scientific foundation for the inclusion of climate forecasts in water supply forecasts. The origins and physical descriptions of the various climate phenomena are described by several review publications available on El Niño (Delecluse et al. 1998; Graham and White 1988; Philander 1992; Rasmusson 1984; Rasmusson and Carpenter 1982), the Pacific Decadal Oscillation (Mantua and Hare 2002; Mantua et al. 1997), and the North Atlantic Oscillation (Hurrell et al. 2003; Marshall et al. 2002). Further, Pagano et al. (1999, pp 12-79) includes a comprehensive literature review of teleconnections and hydroclimatic impacts in the southwestern US.

3.2 The El Niño/Southern Oscillation

El Niño is the oceanic component of a coupled atmosphere-ocean interaction occurring in the tropical Pacific. It is characterized by anomalously warm Sea Surface

Temperatures (SSTs) in the eastern Tropical Pacific Ocean. In contrast, anomalously cold SSTs are referred to as La Niña (or a "cold event"). These shifts in ocean temperatures induce a regional atmospheric response that, in turn, has an impact on global atmospheric circulation. The local atmospheric response is referred to as the "Southern Oscillation" and the entire phenomenon is collectively known as the El Niño/Southern Oscillation or ENSO. For this study, the terms "El Niño" and "warm ENSO events" are used interchangeably to indicate warm ocean conditions and the accompanying atmospheric response. The phenomenon as a whole is the dominant global climate signal on interseasonal and interannual timescales and is the subject of an extensive body of research dating back to the early part of this century. Figure 3.1 shows a time series of the September-November averaged values of the Niño3.4 Sea Surface Temperature Index. High values of this index indicate El Niño/warm ENSO conditions and negative values indicate La Niña/cold ENSO conditions.

Early research indicated that ENSO could be responsible for changes in precipitation and temperature patterns in North America (Namias and Cayan 1984; Ropelewski and Halpert 1986). Ropelewski and Halpert recognized, among other things, the tendency for above normal precipitation in northern Mexico, the Great Basin and High Plains (New Mexico and Colorado) in October to March during warm ENSO conditions. They noted an opposite signal (dry) in the Pacific Northwest. Andrade and Sellers (1988) uncovered a relationship between warm ENSO events and wetter than normal fall and spring (but not winter) conditions in Arizona. A similar mid-winter "lull" in El Niño impacts has also been found in New Mexico (Lee et al. 2002). El Niño's correlation with precipitation is regionally strongest over the Salt and Upper Gila River basins. Andrade and Sellers attribute this relationship partly to anomalously warm SSTs off the coast of California and western Mexico during El Niño causing more and stronger west-coast troughs. Warm Californian coastal SSTs also favor the low-level transport of moisture into the region. Only during the strongest warm ENSO events do Tropical Pacific storms (like Nora and Octave) impact Arizona and/or California. In September, the peak month for Tropical Pacific hurricane recurvature (i.e., when storm tracks become increasingly northward instead of westward), 3.4 tropical cyclones per year are generated, on average, during warm ENSO years versus 2.3 developing during non-ENSO years (Webb and Betancourt 1992).

El Niño's impact is not limited to the southwest. Redmond and Koch (1991) synthesized a series of past works that identify an out-of-phase relationship between the climate of the Pacific Northwest and the Southwestern US. During warm El Niño events, the wintertime storm track (jet stream) is displaced to the south and intensified, increasing the likelihood that storms will affect the Southwestern US (and, in turn, reduce the likelihood that they will affect the Pacific Northwest). El Niño favors dry conditions throughout the northwest (Cayan et al. 1999; Clark et al. 2001; Hamlet and Lettenmaier 1999; Kahya and Dracup 1993; Piechota and Dracup 1996; Redmond and Koch 1991; Ropelewski and Halpert 1987; Smith and O'Brien 2001) especially in northern Idaho (Harsburger et al. 2002), Washington/British Columbia (Bitz and Battisti 1999), and the Oregon Cascades (Koch and Fisher 2000). The signal of El Niño in the Upper Colorado River Basin is almost non-existent except in the Upper Green and San Juan basins

(NRCS 1997; Brandon 1998). If precipitation stations in the Upper Colorado basin with poor correlations are removed, a handful of high-elevation stations with modest correlations remain (El Niño favors wet, Hidalgo and Dracup 2003).

Figure 3.2 shows the correlation of the Niño3.4 index with seasonal precipitation. These diagrams say that El Niño favors wet conditions in the fall in Arizona and New Mexico and dry conditions north of 42 degrees North latitude except southern Wyoming. In winter, El Niño's wet signal shifts to the west including most of California and central Utah. The dry signal also moves to the north, with more of a focus in the Cascades, northern Idaho and Montana. During spring, El Niño still favors wet in California, Arizona, New Mexico and parts of southern Colorado. There is no signal for dry in the Pacific northwest in spring and there is a weak signal for wet in Idaho in the spring. El Niño's signal in summer precipitation is generally weak across most of the country except the northern Great Plains. Although there is a weak signal for wet in summer in Washington during El Niño, this is generally a dry time of year for that region.

Not surprisingly, the impacts of El Niño on snow bear a strong resemblance to its impacts on precipitation. El Niño favors snowy conditions in the southwest and less snow in the Pacific northwest (Cayan 1996; Clark et al. 2001). La Niña conditions are more reliably wet than El Niño conditions are dry in the Northwest (McCabe and Dettinger 2002). In the southwest, the correlation between El Niño and snowpack grows from weak to strong between January and April (McCabe and Dettinger 2002; Lee et al. 2002). In early winter, the Pacific Northwest and Great Basin (Western Colorado to Lake Tahoe) tend to be snowier during El Niño (Smith and O'Brien 2001), as do the northern

Cascades and Northern Rockies (Idaho, Montana, Northern Wyoming) in mid-winter. Smith and O'Brien find that during late winter Northern Utah and Wyoming tend to have more snow during both El Niño and La Niña compared to "non-Niño" years whereas just the opposite happens in nearby southwest Montana and eastern Idaho (more snow during "non-Niño" years). Few of these studies take into account the observed strong trends in snowpack mentioned in chapter 3.6. Only when the trend was removed from the snowpack data did Lee et al. (2002) detect any relationship between snowpack and El Nino in the Upper Rio Grande basin.

Figure 3.3 shows a map of the correlation of 1 April snowpack with September-November Niño3.4 index. At least 40 years of valid data must exist for a value to be shown. Circles indicate positive correlation, triangles indicate negative, with filled symbols having significance at the 0.1% level ($|R| > 0.5$). El Niño moderately favors light snow conditions in the Pacific Northwest and Wyoming, and weakly favors greater snow conditions in Arizona and New Mexico.

The influence of ENSO on streamflow is more robust than its influence on precipitation. Streamflow integrates the response of an entire watershed and it can be thought of as a low-pass filter for atmospheric variability. Typically, point measurements of rainfall possess significant noise because they do not capture the spatial variability of precipitation. The Western US rivers with the strongest correlation between seasonal SOI and seasonal streamflow are located in Arizona (the Virgin, Salt, Gila and San Pedro) and the Pacific Northwest (Cayan and Peterson 1989). Rivers in Arizona have correlation coefficients on the order of –0.5 (indicating El Niño implies high flow conditions), whereas the Pacific Northwest rivers have coefficients of 0.5 (El Niño/low flow). Redmond and Koch (1991) also investigate streamflow and generally confirm Cayan and Peterson's results. The results differ somewhat in the Southwest; Cayan and Peterson's analysis is seasonally based whereas Redmond and Koch investigated water year streamflow totals. El Niño also favors low flow conditions in British Columbia (Hsieh et al. 2003).

Piechota et al. (1997) used principal components analysis and cluster analysis to detect groups of streamflow stations that tended to seasonally co-vary. They aggregated streamflow in these regions and then composited the streamflow during El Niño years. Of the eight coherent regions of variability in the West, four possessed an El Niño signal. During El Niño, the Pacific Northwest (all of the Columbia river basin east of the Cascade mountains) favors dry in May-September. The Northern Rockies (a small region at the intersection of southeast Idaho and North Utah) have high flows for the 10 months prior to the El Niño winter (a preceding relationship), and dry soon after from April to July. Northern New Mexico tends to have high flows throughout almost all of the water year in which the El Niño winter occurs. Finally, southeast Arizona and southwest New Mexico have high flows during March through June. The results of this study should not leave the reader with the impression that the above regions are the **only** regions where El Niño has an impact; the dataset used by Piechota et al. had sparse coverage. For example, Arizona, Nevada and Wyoming were collectively represented by five streamgages, whereas California, alone, contained almost 20 gages. The authors noted a shift in the way Western US rivers responded to ENSO in 1976, an allusion to decadal variability in the Pacific (this topic will receive more attention in chapter 3.3).

3.3 The Pacific Decadal Oscillation

A dominant pattern of Pacific climate variability is the Pacific Decadal Oscillation (PDO). While ENSO describes tropical Pacific variability on timescales of 1-5 years, the PDO describes extratropical Pacific Ocean variability on time scales of 20-30 years. Paleoclimate reconstructions suggest a second periodicity of around 70 years. The "warm" phase of the PDO is associated with warm ocean temperatures in the eastern equatorial Pacific and along the west coast of North America. During this phase, cool temperatures dominate the central and western region of the Northern Pacific Ocean. The "cool" phase of the PDO indicates the opposite. The PDO "index" is defined (by Mantua et al. 1997) as the leading principal component of North Pacific monthly sea surface temperature variability (poleward of 20 N) (see figure 3.4).

The general sentiment of researchers is that cool PDO prevailed from 1890-1924 and 1947-1976 and warm PDO occurred in 1925-46 and 1977 to the mid 1990s. There is controversy about the exact date of the "regime shifts", and there is particular uncertainty about whether a shift from warm PDO to cold PDO occurred in recent years. Although the causes of PDO variability are not well understood, and the PDO index can exhibit wild short term swings (e.g., 1989-91 and 1957-59), the phenomenon exhibits remarkable multi-year persistence. The autocorrelation of the September-November PDO index is significant (positive) at the 95% level out to 12 months and it remains positive for 7 years. In comparison, the autocorrelation of September-November Niño3.4 is significant at 10 months, but is only positive to 1 year. Therefore, simply observing the current state of the PDO can assist in forecasting climate on seasonal and possibly intra-annual timescales.

The impacts of PDO on surface precipitation and temperature in the Northern Hemisphere are generally similar but not identical to those of ENSO. Warm PDO brings dry winter conditions across the northern tier of states from the Pacific Northwest and Southwestern Canada to the Great Lakes. It also brings wet conditions to Arizona, New Mexico, Colorado, Oklahoma, Kansas and Mexico. The reverse is true of the cold PDO phase (Mantua and Hare 2002).

Liles (1999, see also Maxwell and Holbrook [2002] for Arizona where similar if weaker results were found) studied the impacts of PDO on annual precipitation totals for New Mexico climate divisions, and the seasonal impacts on Albuquerque precipitation. He performed an analysis on PDO epochs (1944-76 vs 1977-97) and compared it to categorization of individual PDO years into strong positive, weak positive, neutral, weak negative or strong negative years. For annual statewide precipitation totals, the average of strong negative (cool) PDO years was only 62% of the average of strong positive (warm) years. During negative PDO years, dry years outnumber wet years roughly five to one. "Wet" and "Dry" was defined as precipitation departures from the long-term average of more than 10%. The signal is strongest in the southwest part of New Mexico and during transition seasons (Mar-May, Sept-Nov). He reports no signal during winter. Maxwell and Holbrook (2002) reported no signal during summer.

The PDO has a modulating effect on El Niño's signal in Alaska and many other places (Koch and Fisher 2000; Gutzler et al., 2002; McCabe and Dettinger 1999), enhancing it during certain PDO/ENSO states, and deconstructively interfering in other years (Papineau 2001). The net effect of PDO on streamflow in Southeast Alaska is that the overall volumes are not changed, but warm PDO yielded earlier streamflow and cool PDO brought more streamflow later in the season (Neal et al. 2002).

Dettinger et al. (2000) found enhanced water year (October-September) streamflow in the subtropics and diminished streamflow in the midlatitudes during warm PDO periods. According to Mantua et al. (1997), annual water year discharge in the Skeena, Fraser and Columbia Rivers (in the Pacific Northwest) are diminished on average 8%, 8% and 14% during warm PDO years compared to cool PDO years. In several works, Hare (e.g., Hare and Mantua 2000) has generally confirmed this result in the Columbia River Basin, Washington State and British Columbia using hydro-climatological, and ecological data.

As with figures 3.2-3.3, figures 3.5-3.6 show the correlation of PDO with precipitation and snowpack, respectively. The overall strength and spatial pattern of PDO's impacts are very similar to El Niño's impacts during fall and winter. PDO has a relatively stronger impact in New Mexico in spring, with almost no signal in Arizona or California during this period. PDO also has no impact anywhere in the summertime precipitation. Interestingly, even though the correlations of PDO with precipitation are slightly weaker than El Niño's correlation with precipitation, the correlation of PDO with snow is stronger than that of El Niño, especially in the Pacific Northwest. This suggests that PDO may also be affecting snow through its influence on temperatures.

3.4 Atlantic variability

The North Atlantic Oscillation (NAO) concerns interactions between the ocean and atmosphere in the North Atlantic region, modulating the strength and position of the subtropical high and polar low air masses (Wallace and Gutzler 1981). It is often measured by the pressure difference between Iceland and Gibraltar and this difference is strongest in winter (see figure 3.7). "Low" NAO years favor the passage of winter frontal cyclones over Southern Europe. In "High" NAO years, these storm tracks are pushed farther to the north, causing Southern Europe to dry out and wettening northern Europe. NAO also influences the climate of the eastern US (e.g., Joyce 2002). Much as ENSO and PDO impact the same regions but operate on different timescales, NAO and Multi-decadal Atlantic variability impact the Atlantic sector similarly but on different timescales. Indeed, Enfield and Mestas-Nuñez (1999) correlated indices of Multi-decadal Atlantic variability with North American seasonal climate and detected a strong signal in the Southwest US, specifically Arizona. Much of this correlation can be attributed to the overlap of decadal variability "epochs" in the Atlantic (1860-1880, 1905-1925, 1940-1960, 1970-1990) with epochs in the North Pacific. For long timescales, therefore, PDO captures much of the Atlantic decadal variability. On month-to-month timescales, the

NAO index suffers from a lack of temporal persistence. Therefore, as an independent predictor, is not expected to have significant predictive ability with respect to Western US seasonal water supplies.

As with figures 3.2-3.3, figures 3.8-3.9 show the correlation of NAO with precipitation and snowpack respectively. In the fall, high NAO seems to weakly favor dry conditions in Arizona/western New Mexico and wet conditions in California, the Cascades and northern Idaho. The correlations are very weak, almost all falling in the range between –0.2, +0.2. In winter, spring and summer there is effectively no signal for NAO in Western US precipitation. Snowpack seems to have a uniform positive correlation with NAO especially in the Upper Colorado River basin although the correlations are typically less than 0.3. As with PDO, to rectify the maps of NAO's impacts on precipitation and snowpack, one may need to also consider temperature effects.

3.5 The Pacific North American pattern

The Pacific North American (PNA) pattern is one of the most prominent modes of low-frequency climate variability in Northern Hemisphere winter (Wallace and Gutzler 1981). It is measured by pressure variations over the Northern Hemisphere (see figure 3.10). High values indicate a deeper than normal Aleutian low, increased ridging over the Western US and low pressures in the eastern US. Low values indicate the opposite. PNA is highly correlated with Pacific Northwest temperatures but has weaker correlations with precipitation (Leathers et al. 1991). Woodhouse (1997, following Keables 1992) developed a modified PNA index to maximize its precipitation signal in the southwest US. El Niño and this modified PNA together explained over 63 percent of the variability of the number of cold-season rainy days in the Sonoran Desert (Arizona).

Redmond and Koch (1991) discuss the impact of PNA on the Western US, finding that the relationship is not as clear as with El Niño. This is probably because the PNA index lacks the temporal autocorrelation of ENSO indices, occasionally "flipping" mid-season. While it may describe coincident climate variability well, it may have limited value as a predictive index. The Colorado Basin River Forecast Center has studied the PNA index as a candidate variable in their seasonal streamflow forecasting regression equations, with modest success (Brent Taylor, CBRFC, personal communication, October 1999). When PNA and ENSO are considered together, there appears to be a non-linear effect enhancing the usefulness of PNA as a forecasting variable (i.e., when PNA is high and SOI is low, streamflow is dramatically enhanced, but the variables do not interact otherwise). Finally, the short time series of upper-air measurements limits the amount of historical research that can be done on PNA's impacts.

As with figures 3.2-3.3, figures 3.11-3.12 show the correlation of PNA with precipitation and snowpack respectively. Positive PNA yields dry conditions in the southwest in fall and dry conditions in the Pacific Northwest in winter. In spring, high PNA favors wet in the southwest outside of Arizona. In general, except for fall, the PNA's predictive signal in precipitation is spatially incoherent and weak.

3.6 Other climate phenomena and trends

Although this study has attempted to blend the state of the art in climate understanding into streamflow forecasting, there are many climate phenomenon that have not been considered for a variety of reasons. This chapter addresses some of the better-known aspects of climate variability that a hydrologist may encounter. Climate research is a rapidly expanding and evolving field and its breadth should not be underestimated. The Climate Explorer (http://climexp.knmi.nl/), an interactive Internet-based analysis tool for hydroclimatic research, offers no less than 70 major climate indices including the Madden Julian Oscillation (Madden and Julian 1994), Quasi-Biennial Oscillation (Naujokat 1986), and the Arctic Oscillation (Thompson and Wallace 1998). This does not consider the 12 ocean temperature climate indices of Drosdowsky and Chambers (1998), or the recent swell of interest in Indian Ocean temperature variability. All of these climate phenomena are scientifically sound although most of them are not relevant to western streamflow prediction.

Few climate prediction techniques are as controversial as sunspots and solar cycles. Prior to the 1970s, solar variations were the most common natural explanation for year-to-year variations in climate and hydrology. Streamflow of certain rivers could be correlated with various sunspot cycles, such as the 11-year cycle, or even short-term (2-3 week) episodes (e.g., Landscheidt 2000a). The behavior of every major climate index has been linked to solar activity (Landscheidt 2000b, 2001a,b).

Coincidentally, the earliest attempts at climate forecasting at the turn of the century were born out of attempts to link sunspot cycles with the periodicities found in climate time series, such as Indian monsoon rainfall (the variability of which is now known to be influenced by ENSO). For example, the methods used to produce the seasonal forecasts of the "Farmer's Almanac" are a combination of solar, astronomical and numerology techniques. However, mainstream seasonal forecasters generally view solar-climate connections as "black science" and frown upon their usage. For example, in the 1980s, an internal memorandum at the Climate Prediction Center strongly discouraged forecasting and research personnel from participating in solar-climate symposia (Tony Barnston, International Research Institute, personal communication, 24 October 2001). The predictive skill of sunspot-climate relationships is very low, and the relationships are unstable in space and time (Korzun 1978; Allan et al. 1996). For these reasons, solar cycles will not be considered in this study.

Hydrologists are keenly interested in the expected impacts of long-term climate change on Western US water supplies. Observed long-term trends in streamflow and precipitation are discussed further in chapters 8.4-8.7, and studies suggest that there are significant trends in Western US snowpack. Northern Hemisphere snow cover is on the decline (Brown 2000). Temperatures throughout the Pacific Northwest have risen, on average 1.5 deg F/century since 1920 causing snowpack in the Cascades to decline almost 50% since the 1950s and as much as 30% in Idaho (Mote 2003a,b). These trends are most pronounced at moderate to low elevations, especially below 1600 meters (5250 feet) elevation. Preliminary analysis by the author (not shown) suggests that Mote's snowpack declines also extend into Montana and Wyoming (the entire Western US north of 42 deg North). If the analysis were extended into the 1930s, Mote would have found that the snowpack time series in this region is convex ("n") shaped, low in the 1930s-1940s, high in the 1950s-1970s and low afterwards. In Colorado and Utah, there were not strong trends in snowpack from the 1930s-1960s. The late 1970s and early 1980s were very high snow years, followed by extended stretches of dry in the late 1980s and late 1990s. It is difficult to know if such variability constitutes a trend (see also Taylor et al 2004).

Nonetheless, Mote (2005) recently significantly expanded his previous analysis of snowpack trends. Using observed precipitation and temperature data forced into a land surface hydrology model, the snowpack record was estimated for all regions of the Western US from 1915-1997. The correspondence between the observed and simulated snowpack trends was very strong, reproducing the spatial extent and seasonality of the trends. This analysis also decomposed the snowpack trends into that forced by precipitation variability and temperature variability. While precipitation is the controlling factor for snow variability for much of the interior west, temperature is becoming increasingly dominant, particularly in transitional rain/snow regions such as the Cascade mountains in Oregon, Washington and northern California.

The Climate Prediction Center does account for some low frequency variability through its use of Optimal Climate Normals (see chapter 4.2.2). No similar activity exits at seasonal streamflow forecasting centers. Eventually, hydrologists will have to address the question of whether snow-based streamflow forecasting equations will still be relevant in a significantly warmer climate or if these strong trends in snow are interfering with the forecasts in other ways.

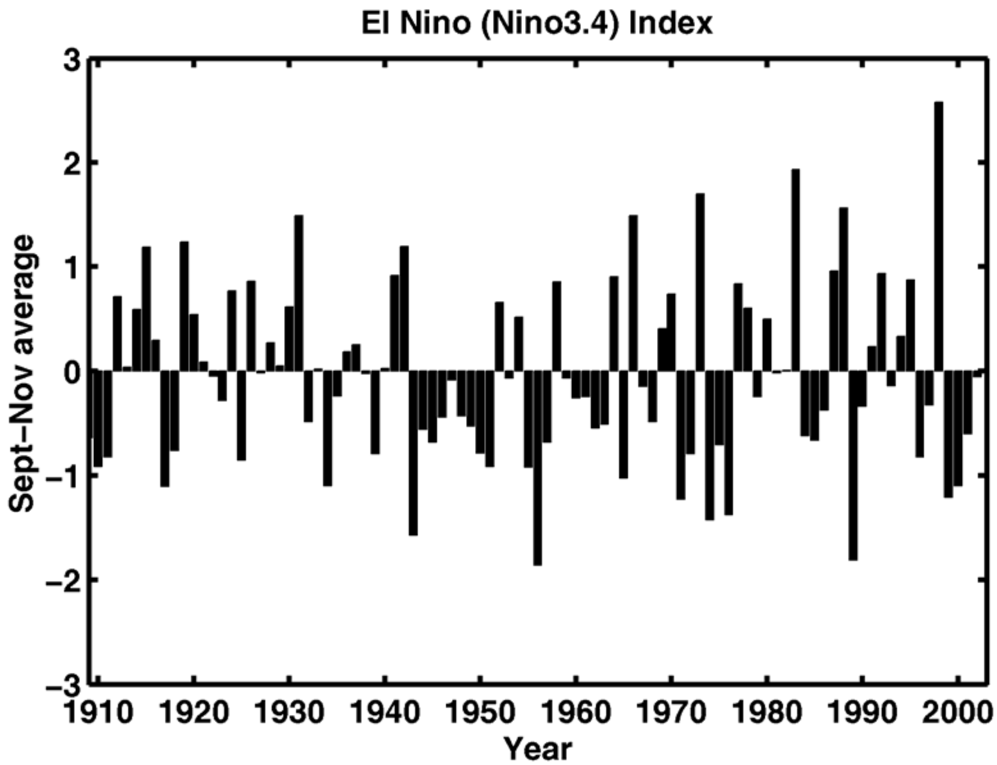


Figure 3.1. Time series of the Niño3.4 Index of equatorial Pacific ocean temperatures, averaged over September-November. High values indicate El Niño conditions and low values indicate La Niña.

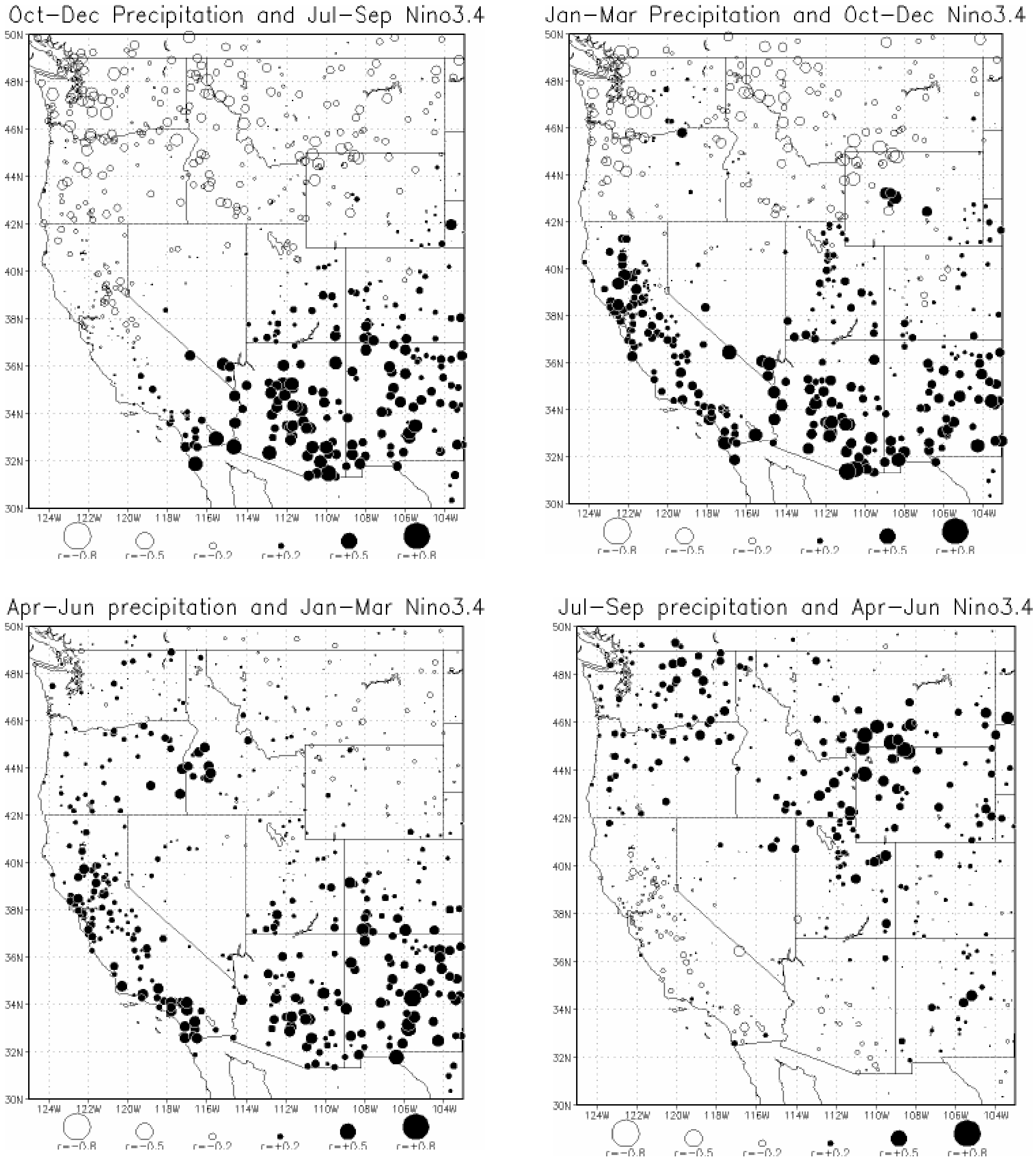


Figure 3.2. Correlation of precipitation with the Niño3.4 index. Each panel corresponds to a 3-month season of precipitation, correlated with the preceeding 3-month average of the Niño3.4 index, as follows: Top left (fall precipitation), top right (winter preciptation), lower left (spring precipitation), lower right (summer precipitation). Positive correlation (black dots) mean El Niño favors wet conditions whereas negative correlation (hollow dots) mean El Niño favors dry. The size of the dot is proportional to the strength of the correlation. Reference dots are provided along the bottom for R = -0.8, -0.5, -0.2, +0.2, +0.5 and +0.8.

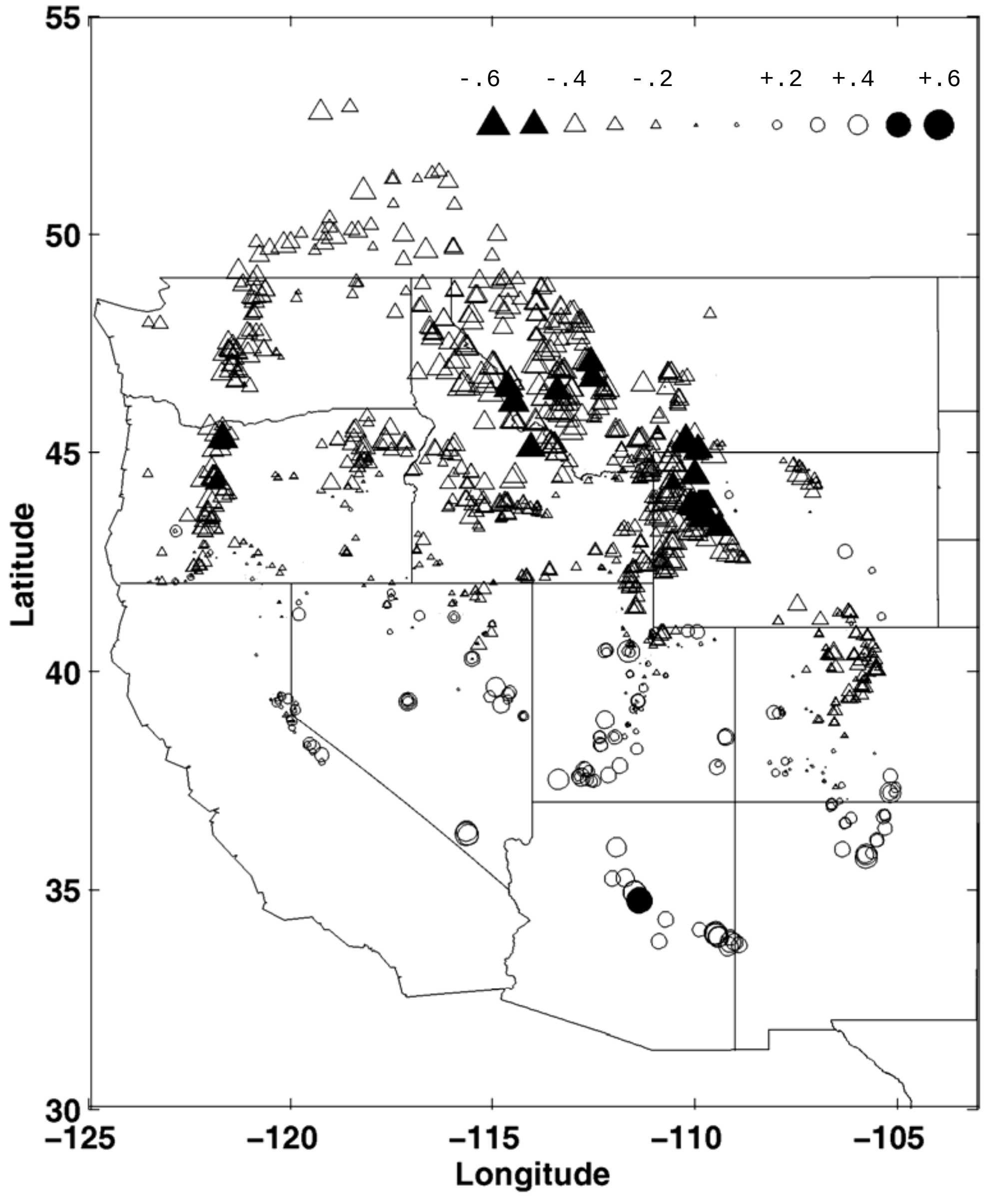


Figure 3.3. Correlation of 1 April snowpack and the September-November Niño3.4 index. At least 40 years of valid data must exist for a value to be shown. Circles indicate positive correlation, triangles indicate negative correlation, with filled symbols having significance at the 0.1% level (R > 0.5). Symbol diameter is linearlly proportional to the srength of the correlation. El Niño moderately favors dry conditions in the Pacific Northwest and Wyoming, and weakly favors wet conditions in Arizona.

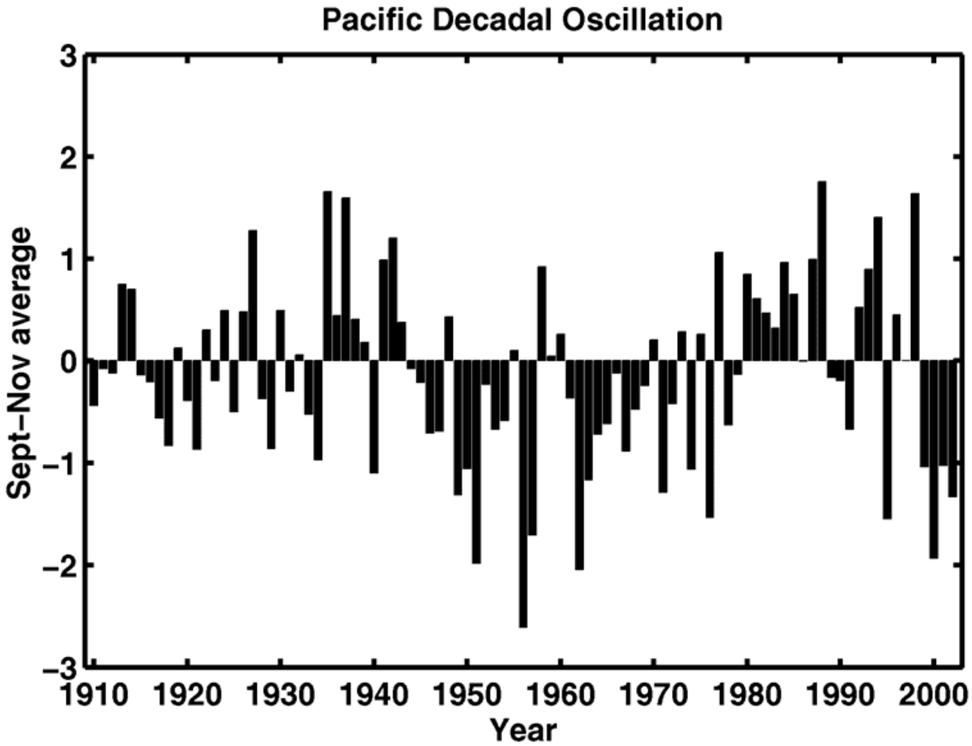


Figure 3.4. Time series of the Pacific Decadal Oscillation index averaged over September-November.

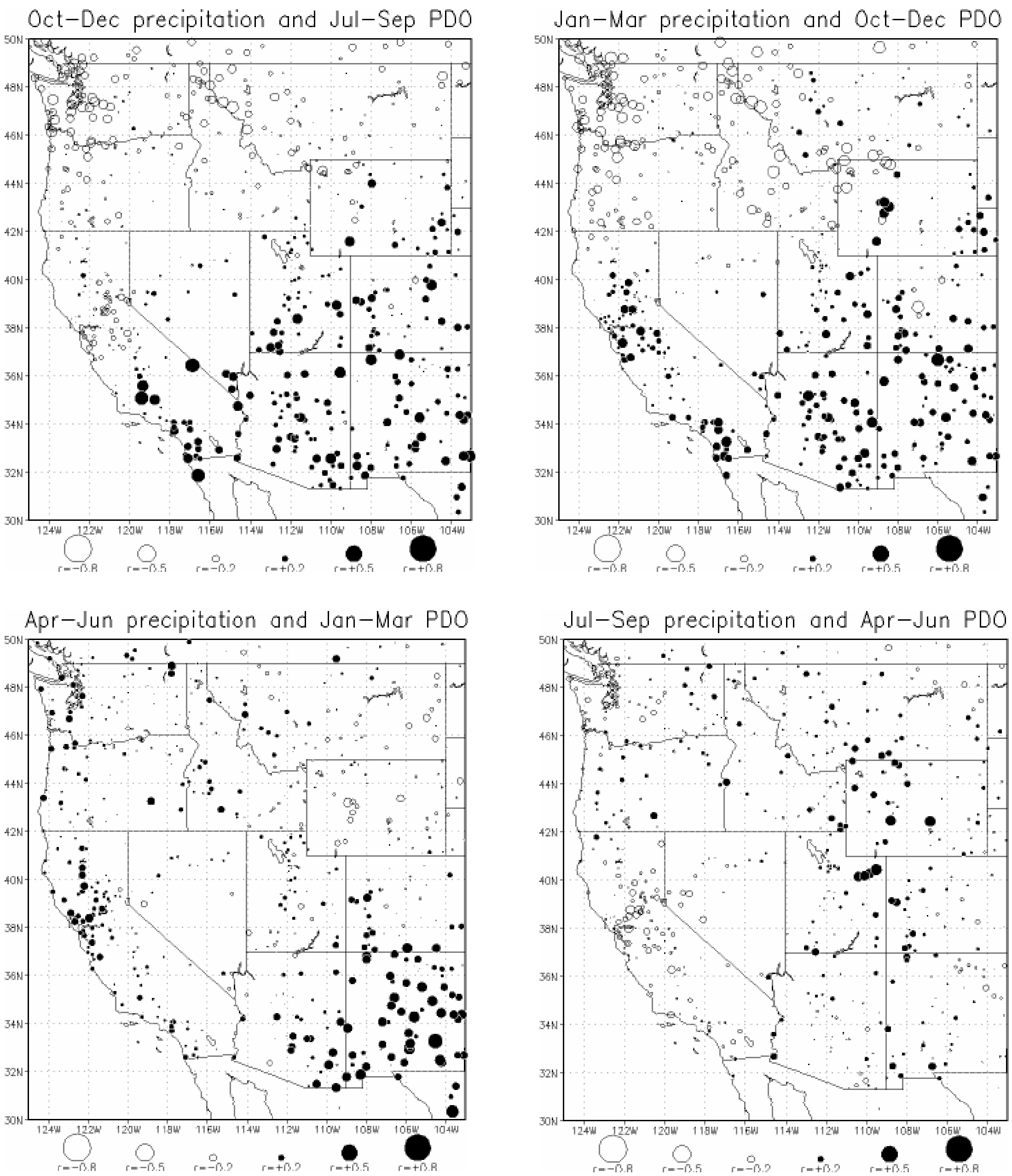


Figure 3.5. Correlation of seasonal precipitation with the PDO index. Compare with figure 3.2.

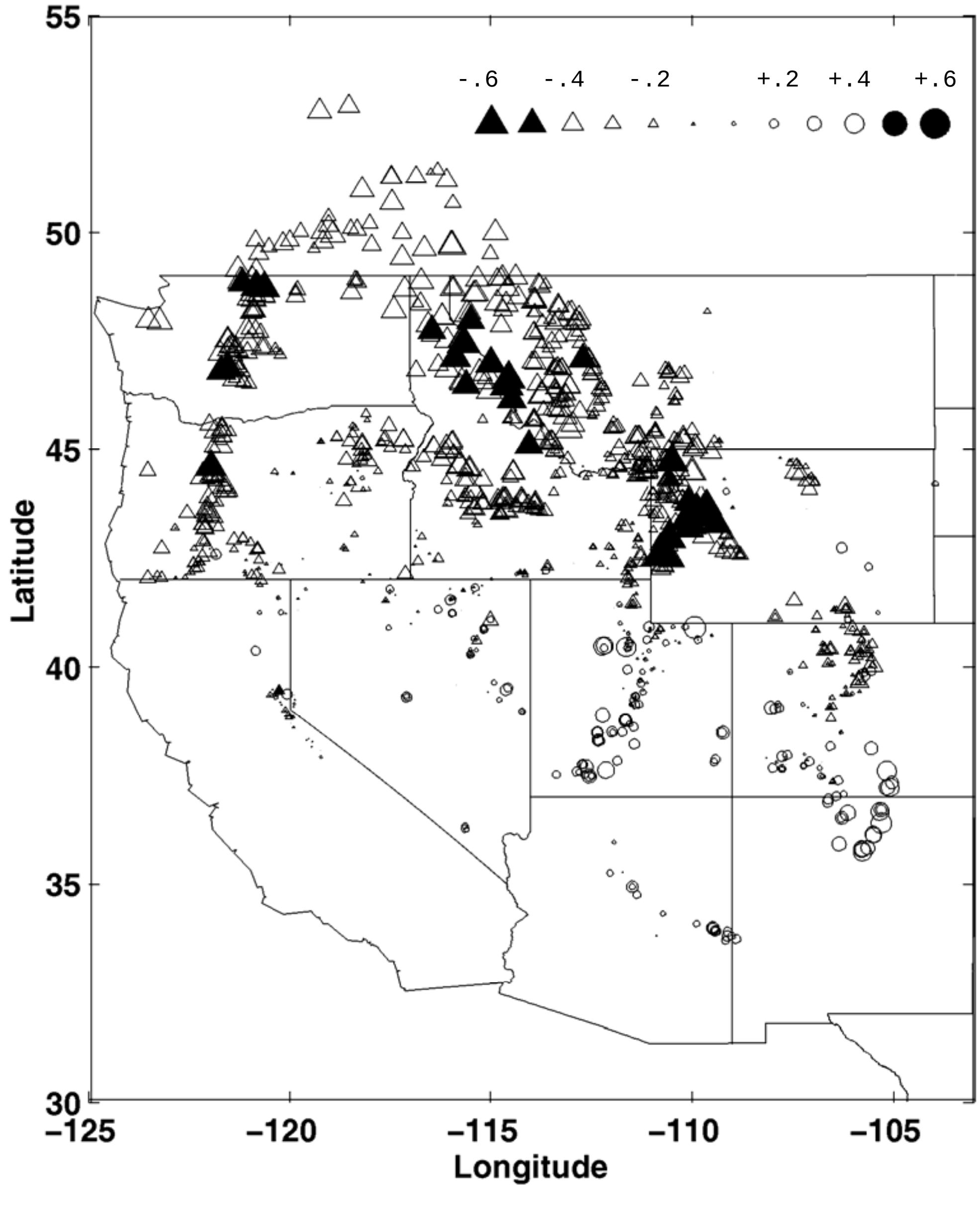


Figure 3.6. Correlation of 1 April snowpack and the September-November Pacific Decadal Oscillation Index. Compare with figure 3.3. Positive PDO moderately favors dry conditions in the Pacific Northwest and Wyoming, and weakly favors wet conditions in New Mexico.

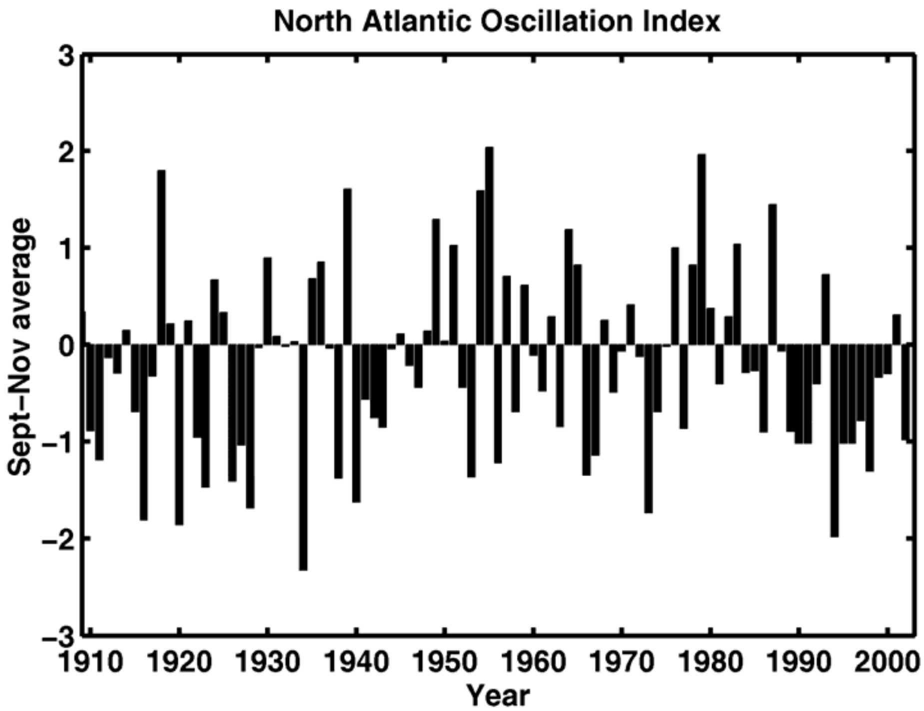


Figure 3.7. Time series of the North Atlantic Oscillation (Gibraltar Stykkisholmur) index, averaged over September-November.

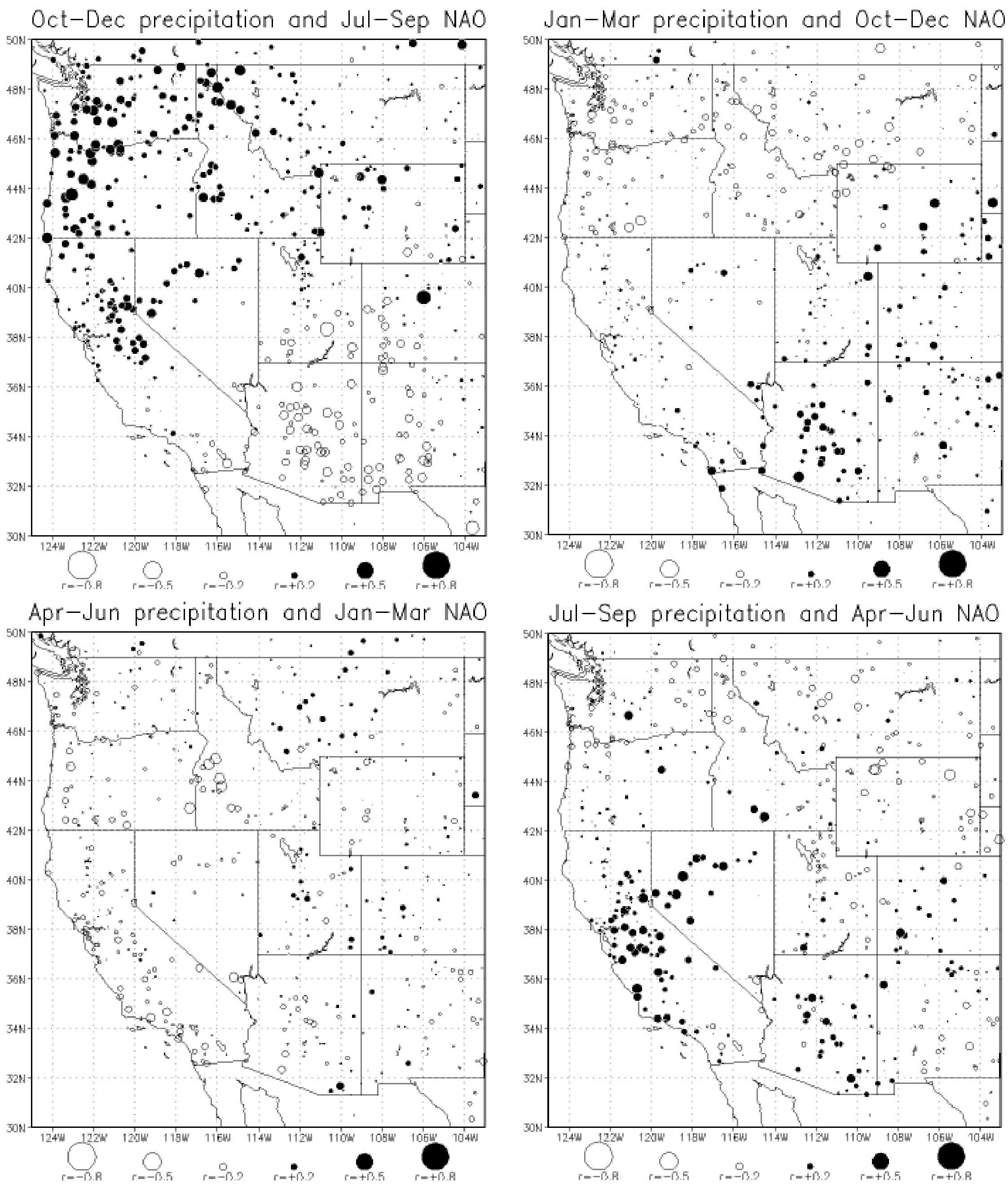


Figure 3.8. Correlation of precipitation with the NAO index. Compare with figure 3.2.

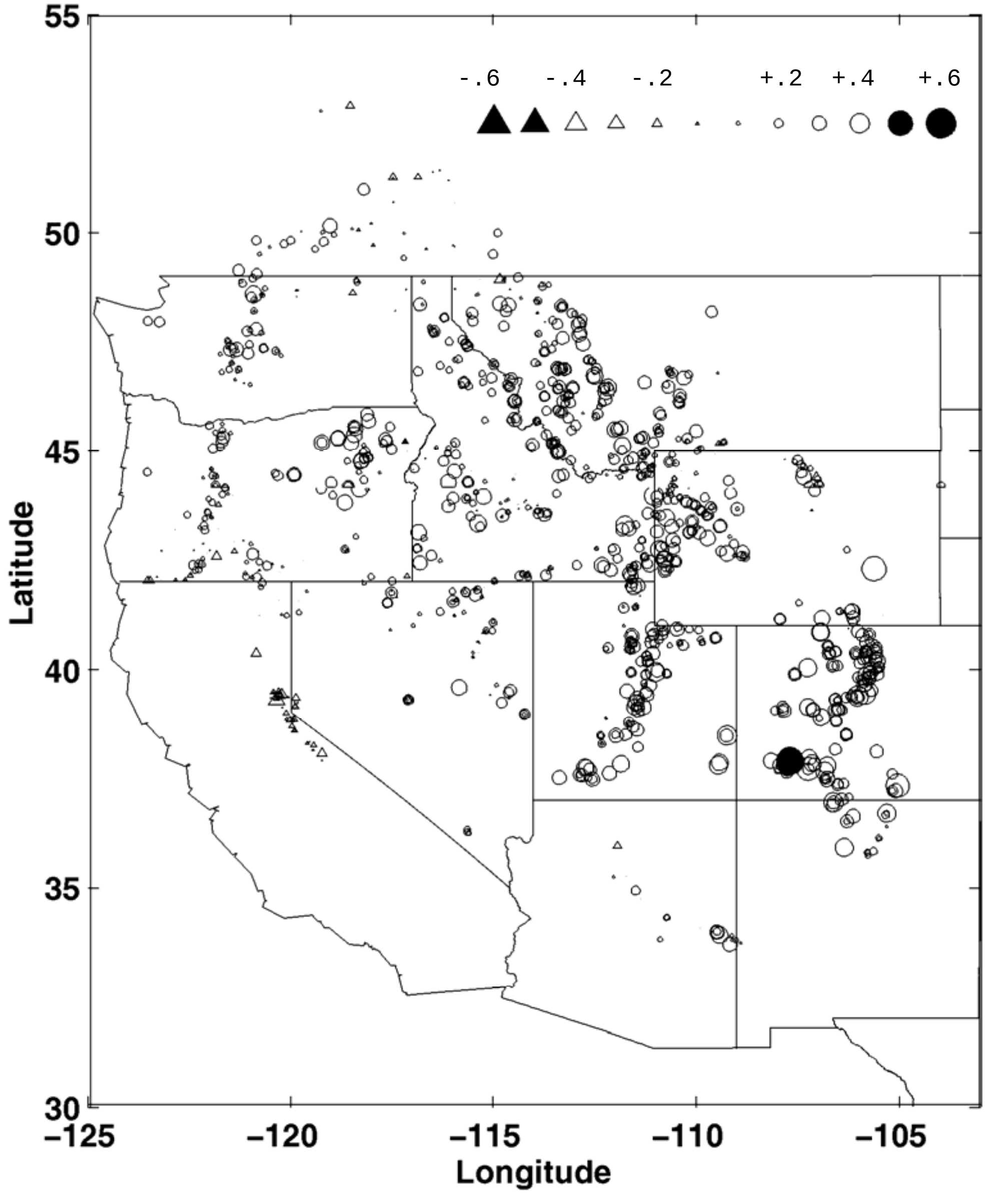


Figure 3.9. Correlation of 1 April snowpack and the September-November North Atlantic Oscillation. Compare with figure 3.3. Positive NAO weakly favors wet conditions throughout the Western US.

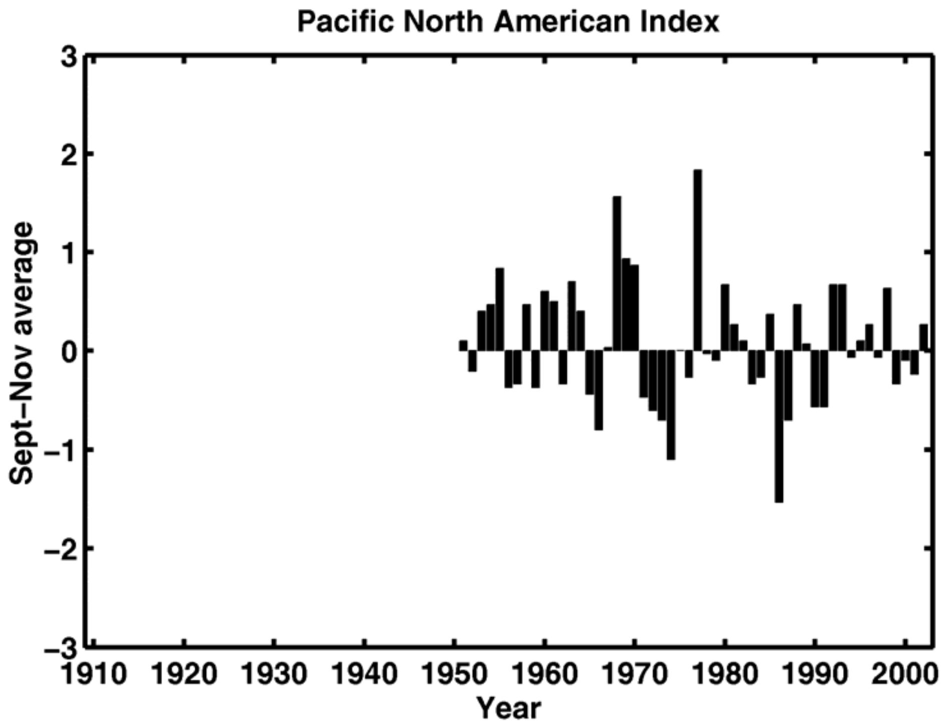


Figure 3.10. Time series of the Pacific North American index, averaged over September-November.

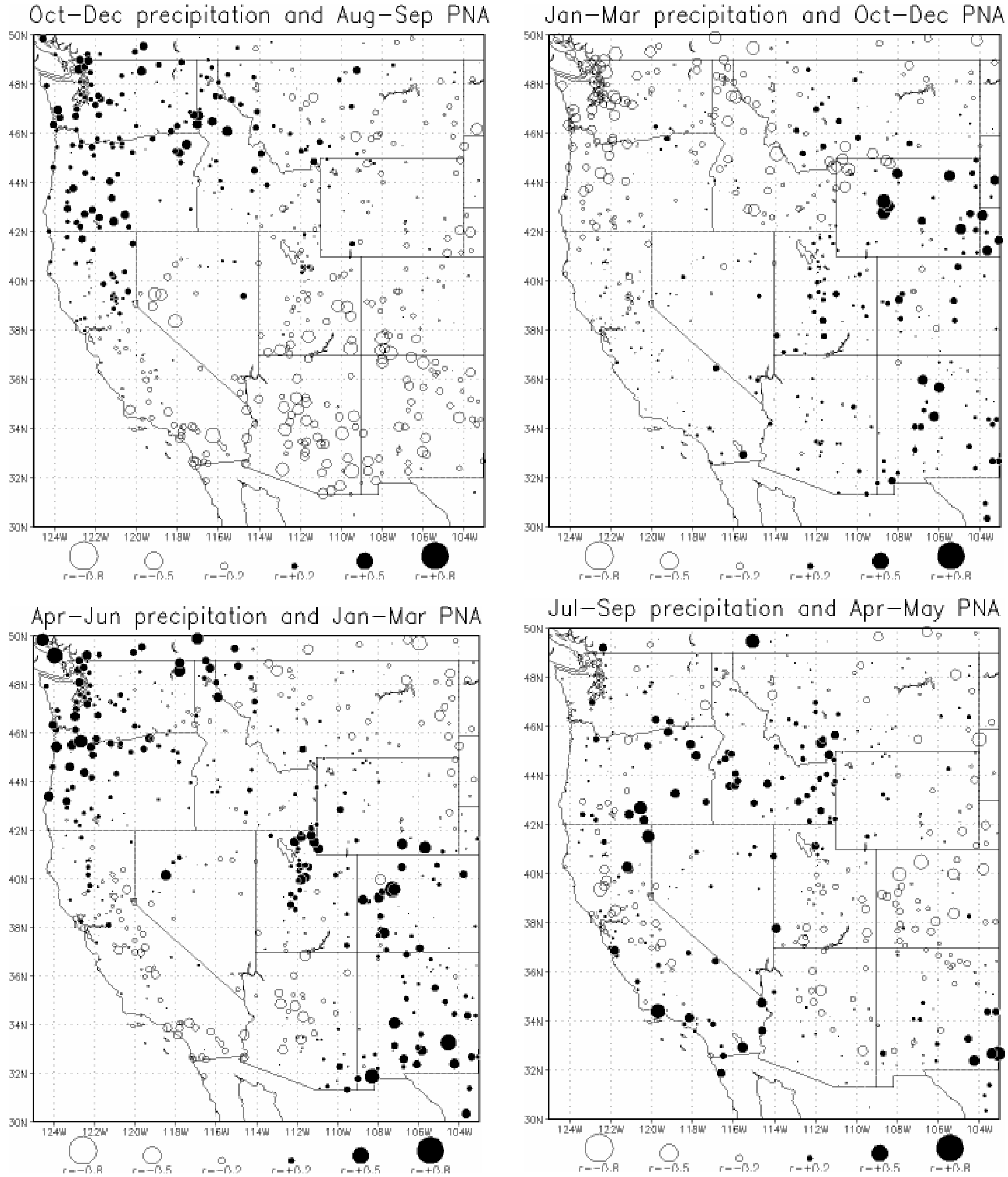


Figure 3.11. Correlation of precipitation with the PNA index. Compare with figure 3.2. The PNA index is not available in June-July. The overall predictive signal of PNA on precipitation is scattered and heterogeneous except in the fall.

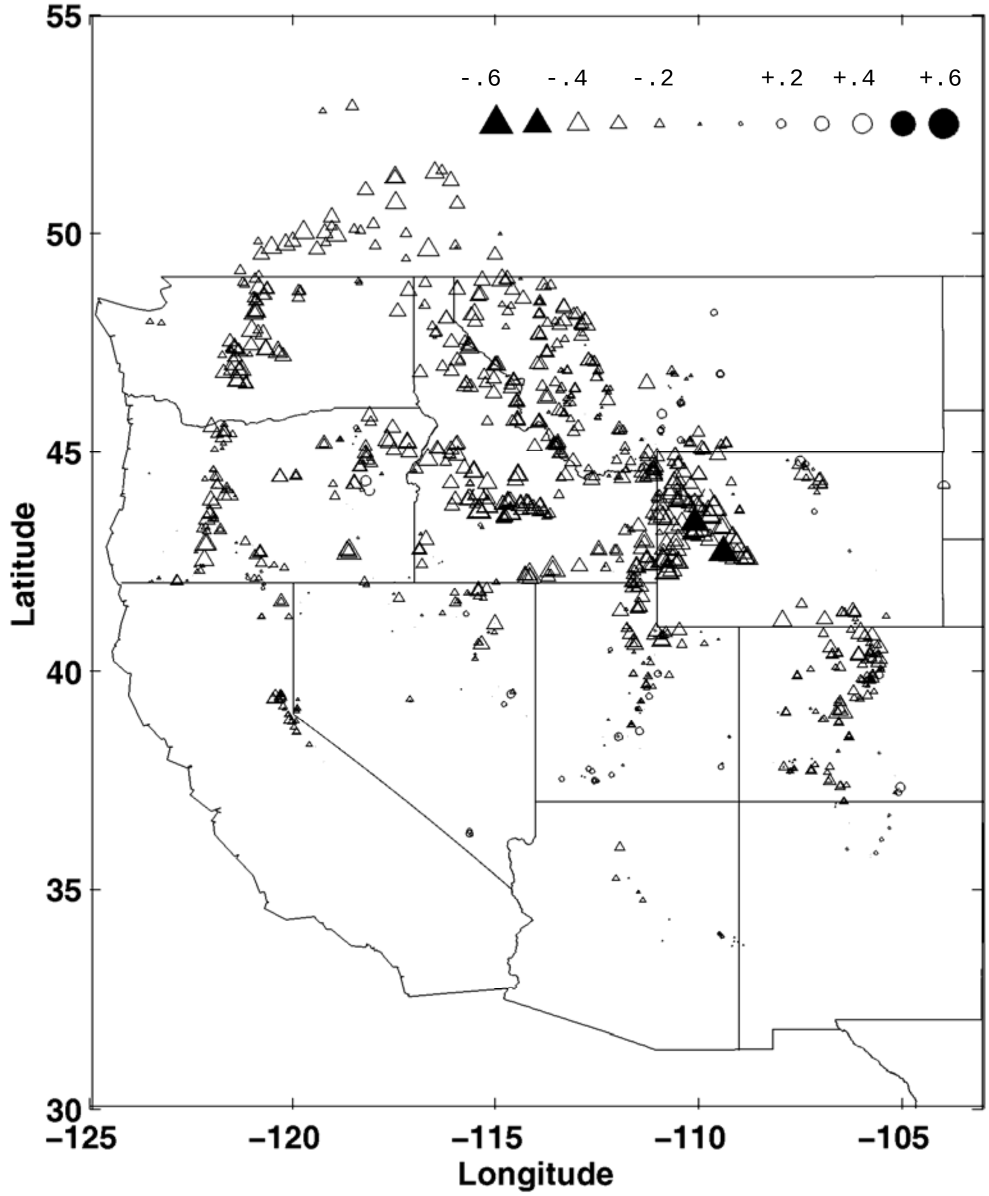


Figure 3.12. Correlation of 1 April snowpack and the September-November Pacific North American Index. Compare with figure 3.3. Positive PNA weakly favors wet conditions throughout the Western US.

4. EXISTING SEASONAL FORECASTS

4.1 Introduction

The previous chapter described the scientific link between several large scale climate features and Western US hydroclimatology. Along with other tools, climatologists use these linkages to create seasonal climate forecasts. A variety of such forecasts exist, produced by an equally varied number of agencies. This chapter describes the content, format and methods of production of the official climate and streamflow forecasts for the Western US. Descriptions of the operational environment are rare if existent at all in the peer-reviewed scientific literature; the forecast process is poorly documented as it is often passed down by oral tradition (Hartmann et al. 2002b). However, if one hopes to positively integrate scientific advancements into the operational environment, one must understand the forecaster's motivations, objectives, limitations and culture. This chapter is a key element in understanding the current generation of products as well as anticipating how they may evolve in the future. The water supply forecast information presented here is admittedly focused on the NRCS process, although other agencies are mentioned when applicable. Hartmann et al. (1999) provide a thorough review of all hydroclimatic forecast products applicable to the Southwestern US, including many not addressed here.

4.2 Climate forecasts

4.2.1 Current products

The current format of forecasts from the Climate Prediction Center (CPC) includes US maps of probability anomalies for the next single month and the thirteen consecutive three-month segments (eg. DJF, JFM, FMA) of average temperature and total precipitation (figure 4.1). These maps can be found on the Internet at http://www.cpc.ncep.noaa.gov/products/predictions/90day/. The contours of the map indicate the probability of a region experiencing conditions which would fall into the “above normal”, “normal” and “below normal” terciles of the historical record 1961-1990 (or 1971-2000 for forecasts after May 2001). In the absence of contours ("climatology" or “equal chances”), CPC indicates that there is an equal (33.3%) chance of the given forecast period falling into each of the terciles. A sophisticated user acting within a risk-based decision framework may choose to interpret CPC’s “climatology” forecast differently, as “total uncertainty” instead of an equiprobable tercile forecast.

The presence of contours indicates a shift of probabilities from one end of the distribution into the opposite extreme. For example, a contour signifying “B 40” indicates a 40% chance of falling in the below normal category, a 33% chance of near-normal and a 27% chance of above normal. This is true except for anomalies indicating very large probability shifts, in which case the remaining probability is removed from the middle (near-normal) tercile and 3.3% always remains in the least likely category.

Although forecasters currently have the option of issuing an increased probability of near-normal temperatures with a diminished probability of not-normal temperatures, they rarely exercise this option. Under the current format, a forecast for an increased chance of near-normal precipitation has never been issued, primarily because a lack of demonstrated skill in being able to forecast "near normal" conditions (Van den Dool and Toth 1991).

Although it is not yet a popular product among general users, CPC also issues "Probability of Exceedance" (POE) forecasts (Barnston et al. 2000). For each of the 90-day forecasts associated with the official seasonal climate outlook mentioned above, point forecasts are given for 102 "mega-climate divisions" around the US. These point forecasts describe the shift in the entire distribution relative to climatology (as opposed to probability shifts at the tercile boundaries). Although the cumbersome and busy format of the POE forecasts has intimidated uninitiated users, their format represents the maximum information content possible in a probabilistic forecast. The POE forecast is an indispensable product to highly sophisticated users who use the forecasts quantitatively. CPC also supports a POE forecast "downscaled" to individual cities (e.g., Phoenix, AZ airport).

Similar to CPC, the International Research Institute (IRI) issues probability anomaly forecasts for 3-month overlapping periods for temperature and precipitation. These forecasts can be obtained at http://iri.columbia.edu/climate/forecast/net_asmt/. The most significant differences between CPC and IRI forecasts are spatial (IRI's domain is global whereas CPC's is the United States) and methodological (CPC relies more heavily on statistical tools whereas IRI relies almost exclusively on simulation models). IRI also does not have a fixed rule on how the tercile probabilities are distributed. For example, although they would have little reason to do so, nothing prevents IRI from issuing a bimodal forecast of 40% chance of dry, 20% chance of near-normal and 40% chance of wet. IRI issues forecasts at lead times up to 4 months ahead (e.g., a forecast issued in October with a target period of February-April), compared to CPC's 12 month lead time. IRI explicitly "pixelates" its forecast map into 2x2 degree latitude and longitude grid cells whereas CPC allows smooth contours of an unspecified spatial resolution. Finally, IRI produces an additional "Extreme" forecast map which focuses on significant shifts in the tails of the forecast distribution. Here "Extreme" means in the lower or upper 15 percent of the climatological distribution. As discussed later, IRI participates in the development of CPC's forecasts and, as such, IRI tries to match its forecasts over the United States with CPC's depiction.

4.2.2 Operations research and evolution of forecast techniques

The history of climate research and forecasting tools is well documented by many review articles, including but not limited to Barnston et al. (1994, 1999, 2000), Epstein (1988), Gilman (1982, 1985), Goddard et al. (2000), Hartmann et al. (1999, 2002b), Mason et al. (1996, 1999), Namias (1968, 1985a), Nicholls (1980), and Wagner (1989).

This chapter provides a broad summary of these publications, traces various paths of progress in climate forecasting history, discusses changes within forecasting agencies and identifies likely advances in the future. Naturally the brevity of this chapter prevents an in-depth analysis and, as such, many events are omitted.

Climate forecasts of some form or another have existed for a very long time, including “almanac” style forecasts based on astronomical or solar cycles (see also Meinke and Stone 2003). Coincidentally, some of the earlier attempts at climate forecasting at the turn of the previous century were born out of attempts to link sunspot cycles with the periodicities found in climate time series, such as Indian monsoon rainfall (the variability of which is now known to be influenced by El Niño, see also chapter 3.6).

In the 1920s and 1930s researchers described the movement of large-scale atmospheric features in the context of fluid dynamics. Given an initial state, one might be able to describe how centers of action in the atmosphere will evolve and affect weather systems. The identification of slowly-evolving systems in the mid-troposphere provided a physical basis for the notion of “teleconnections”, action at a distance in the atmosphere. By 1941, the theoretical basis was sound enough for the Weather Bureau to begin issuing 5-day operational forecasts, primarily for military purposes. These forecasters were experts in the art of reading atmospheric pressure maps, understanding the physical processes at work and visualizing the future trajectories of features. Namias extended these ideas to begin producing experimental and routine 30-day forecasts in the mid-1940s. Official forecasts, issued to the public, began in 1953.

In time, the appreciation for the influence of ocean temperature patterns on seasonal climate grew. Features such as El Niño, the Southern Oscillation and the Pacific North American (PNA) pattern were recognized and described. By the late 1960s and certainly by the early 1980s, researchers and the operational community were aware of the impacts of these sometimes-remote phenomena on the United States (see chapters 3.2 and 3.5). Namias also sketched out two other phenomena important to seasonal forecasting, extratropical sea surface temperature anomalies (Namias 1965) and antecedent land surface conditions such as soil moisture or snow (Namias 1962, 1985b). All of these research areas are still active today.

In terms of forecasting tools, analogue and persistence forecasting techniques are among the oldest modern tools in use. The technique involves looking to the past to find one or more months or seasons that resemble the most recent month or season. The prediction is then similar to the climate that was observed in the periods following the chosen analogues. For example, Barnett and Preisendorfer (1978) used principal components analysis on ocean temperatures and atmospheric pressures to define the primary modes of variability (the “climate vector”). The current climate vector is then compared to past climate states to determine the year in the past that most resembles the present (e.g., January-March 2003 most resembles January-March 1973). The future should then resemble the season after the selected analogue (e.g., April-June 2003 should resemble April-June 1973). A recent advance is the use of “constructed” versus “natural” analogues. A natural analogue forecast map is derived from a single year (e.g., 1973) whereas a constructed analogue is pieced together from different years for different locations (e.g., 1975 in Chicago, 1983 in Boston, 1943 in Denver). In comparison to analogue forecasting, persistence forecasting involves the likelihood that, for example, a wet winter is followed by a wet spring. Along with subjective pressure height chart analysis, these techniques were most in use in operational climate forecasting from the mid-1960s to late-1980s.

Analogue selection techniques are not without their critics. Improper use of the approach involves developing, perhaps subliminally, a pre-conceived notion of what the forecast should be. Next the climatologist suffering from "confirmation bias" selects similarity criteria to produce an analogue that matches the pre-conceived forecast while ignoring other information that indicates the inappropriateness of that analogue. This climatologist may steer the analogue selection criteria towards extreme years to draw attention to the forecast (e.g., 1966 may be emphasized because it had the largest snow storm on record, even if it is only the 4th best analogue behind several uninteresting years).

Another pitfall involves the selection of physically irrelevant similarity criteria. For example, while floods have happened in Arizona in 1983 and 1993, one should be reluctant to tell users that they can wait until 2003 for the next flood because flood years end in "3". Although overt numerology has no place in climatology, subtler examples of correlation without physical mechanism abound (e.g., solar cycles). Consistent and objective selection criteria conceptually associated with physical mechanisms are generally favored over subjective, ad hoc techniques divorced from a causal mechanism. Finally, as the past never exactly repeats itself, deterministic analogue forecasting can be misleading to the user; statements such as "We're looking at a repeat of 1988" are bound not to verify (Nicholls 1999). Several authors, including Mark Twain, may have once said, "History does not repeat itself, at best it sometimes rhymes."

Climate forecasters recognized this flaw relatively early and moved away from deterministic analogue selection. Livezey and Barnston (1988), and Barnston and Livezey (1989) developed a technique that considers both analogues (good matches) and anti-analogues (poor matches). Instead of selecting one “best” analogue, the degree of similarity between the present and each past year determines the weighting each year will get in the forecast. Livezey and Barnston used as inputs principal components of Northern Hemisphere pressure heights, pressure level thicknesses, sea surface temperatures, surface temperatures over the United States and an index of the Southern Oscillation. Livezey et al. (1990) expanded the technique to blend analogue forecasting and persistence forecasting. The output of the model was also changed from categorical forecasts to deterministic anomaly forecasts.

Barnston and Ropelewski (1992) evolved the previous techniques by using Canonical Correlation Analysis (CCA), a form of multiple linear regression, to forecast the behavior of El Niño. Barnston’s (1994) application of CCA to US precipitation and temperature followed with widespread operational adoption in 1996. CCA is more advanced than simple linear regression in that it accepts more than one predictor and can predict a pattern or more than one variable. In essence, one derives the primary modes of variability among the predictors using principal components analysis. One then finds the primary modes of variability among the predictands using the same technique. The time series of these components are then cross correlated in a multiple linear regression framework. In real time, the predictor principal component values are computed, the predictand components are forecasted and then the predictand components are projected into their original space (e.g., temperature at various stations). Screen Multiple Linear Regression (SMLR) is similar to CCA (Unger 1996a,b) although it does not perform any principal components analysis on the predictand field. SMLR is good at describing fine spatial scale forecast features, such as coastal effects, whereas the CCA tends to produce smoother predictions in space. CCA finds more use by operational climate forecasters than SMLR, while SMLR most resembles what is in use in operational water supply forecasting circles.

Another forecast tool involves the feedback of the land surface on itself. Namais (1962) found that anomalously widespread snowcover in the central and southern US in 1959-60 caused as much as 10° F error in the temperature forecasts that were based on pressure level heights alone. The same study identified that precipitation anomalies, particularly in the Great Plains, can be self sustaining. A lack of precipitation desiccates soils that then have less moisture to evaporate to feed convective storms. Wet soils evaporate an abundance of water to convective systems that, in turn, rain more than they would otherwise. Building on notions of persistence described previously, van den Dool et al. (2003) developed an analogue precipitation and temperature forecasting technique that analyzes spatial patterns of a precipitation-based proxy of soil moisture. This “Soil Moisture Tool” (SMT) is mostly used in summer and in the Great Plains.

Climate forecasters have recognized that decadal variability and trends in climate exist and they try to compensate for this using the “Optimal Climate Normal” (OCN) tool. Official forecasts are currently expressed relative to a 30-year average, 1971-2000. Huang et al. (1996) found that the 30-year average is not necessarily the optimal or most skillful baseline for the forecasts. Instead, OCN subtracts the official 30-year climate average from the average over the past 10 years for temperature, and 15 years for precipitation. This anomaly then becomes the forecast, usually as a 5-10% probably anomaly centered on the largest normalized departures. For example, August-October temperatures around Las Vegas, NV have been in the range of the warmest 1/3 of years of the 30-year climate normal for twelve of the past twelve years (1991-2002). A forecaster should be highly reluctant to forecast anything but high probability of warm for this season and location. One can recognize an OCN based forecast feature by its appearance year after year. Currently, the only significant precipitation OCN signal in the Western US is a tendency for wet April-July in the Northern Cascades and dry July-September in the broader Pacific Northwest, as indicated by the persistent pattern in the recent operational maps.

The last major class of tools used in climate forecasting consists of dynamical numerical models or General Circulation Models (GCMs, Shukla et al. 2000). These resource intensive models consider the many complex physical processes that affect climate, including the behavior of the ocean, atmosphere and land surface. They are initialized with boundary conditions of the ocean (SSTs) and run forward in time with persisted SSTs or forecasted SSTs. The model is usually run several times with its initial conditions slightly modified each time. The result is an ensemble of forecasts from which a probabilistic forecast can be obtained or an average deterministic forecast derived.

The potential use of GCMs for seasonal climate forecasting has long been recognized. Researchers envisioned a day when incredibly powerful super computers would be able to blaze through countless tedious calculations to simulate the behavior of the atmosphere. A computer was used for the first time in 1950 to calculate weather information in the short range. GCMs grew in complexity in line with increases in computing power. Laurmann (1975) captured the belief by some that simulation modeling would soon dominate seasonal climate forecasting, despite its shortcomings at the time. Others expected statistical techniques to be competitive with simulation modeling for many years (Nicholls 1980). While popular among the research community, simulation models found little use in the operational climate forecasting environment until CPC adopted the model described in Ji et al. (1994). The formation of the IRI represents a dramatic shift in operational philosophy with its near total dependence on guidance from many different simulation models run in a variety of ensemble configurations.

IRI has also been very active in the practical research involved with use of simulation models in an operational environment. Research areas include how to combine the output of several models, how to downscale coarse continental predictions into high-resolution forecasts that consider complex topography, and how to compensate for large biases in the models. IRI has also come to appreciate that, compared to statistical techniques that are "damped" (i.e., forecast skill and forecast "bullishness" are related), simulation models tend to be overconfident and have high false alarm rates (Tony Barnston, International Research Institute, personal communication July 17, 2003).

IRI is especially active in the social science and applications aspects of climate forecasting. There is a long history of trying to understand the users, uses and perceptions of climate forecasts (e.g., Easterling 1986; Sonka et al. 1982). Ararbanel et al. 1980 conveyed from users that the categorical format of climate forecasts made them "virtually worthless"). However, this field has benefited from the surge in funding for climate change research and the formation of NOAA's Office of Global Programs (OGP). Research on societal vulnerability and adaptation to long-term climate change is highly related to topics dealing with interannual and seasonal climate variability. These projects have also focused on the effective communication of uncertain and/or probabilistic information, and the development of quantitative and qualitative decision support tools (NRC 1999; IRI 2001).

Along with IRI, OGP also funds the Regional Integrated Sciences and Assessments (RISA) program, a series of local interdisciplinary social and physical research projects, coupled with user outreach and communication activities ("stakeholder driven research"). One could easily envision a RISA program conducting intensive field tests of different climate forecast formats and displays to determine the strengths and weaknesses of, for example, various color schemes and layouts ("product testing"). RISAs will eventually learn from and contribute to the fields of usability engineering (Nielsen 1994) and information design (Tufte 2001) in the context of climate forecasts.

The remainder of the research community is focusing on a wide variety of topics (see also chapter 3). Along with the continued study of well known phenomena (e.g., El Niño), researchers are branching out into phenomena such as the Arctic Oscillation, the Pacific North American Pattern, the North Atlantic Oscillation, and others, some of which are interlinked. Researchers are also interested in multi-year to decadal variability such as the Pacific Decadal Oscillation and the Quasi-biennial Oscillation as well as intra-seasonal (30-60 day) variability such as the Madden Julian Oscillation. Finally, researchers are investigating complex climate features that typically have been difficult to predict, such as summer convective rainfall and the monsoon.

Operational climate forecasters generally consider information about El Niño and the output of their core tools generally tested and reliable. Information about decadal oscillations, the monsoon or other research topics are on the minds of forecasters but this information rarely finds its way into operational forecasts unless it is quantifiable and generally accepted. For example, information about the PDO is not formally used in forecasting, outside what may be picked up by the CCA or OCN tools; there is some skepticism on the part of forecasters that the PDO is a real phenomenon independent of El Niño (see also chapter 3.3). Nonetheless, there is an active, vibrant and well-funded operations-oriented climate research community interested in process studies, simulation models, statistical tool innovation, and the communication/use of products.

4.2.3 The operational climatology environment

Two major centers produce official climate forecasts over the United States: the NWS Climate Prediction Center (CPC) and the International Research Institute (IRI) for Climate Prediction. The IRI has a broader scope in that it produces forecasts for all land areas of the globe whereas the CPC is focused on the United States. Both agencies employ approximately 70 personnel, ¼ of which are involved with research and tool development, 1/5th engage in climate data monitoring, and 1/10th are operational forecasters although these boundaries are blurred as operational forecasters also do research, and researchers participate in forecast production. IRI has 30% of its staff devoted to user-oriented applications development, whereas the CPC is linked to the recently formed NOAA climate services division (10 employees). These figures do not include the broad real-time operational support these entities receive from academia, international agencies and groups such as the Climate Diagnostics Center in Boulder. Colorado. For example, 17 groups around the world contribute their predictions of El Niño Sea Surface Temperatures to IRI every forecast cycle.

As forecasts are produced every month of the year, climate forecast activities do not have a pronounced seasonality aside from the seasonal variation in forecast skill. The impact of El Niño on climate tends to be greatest in winter so if an El Niño event is present, there may be increased focus and visibility given to forecasts issued September to December. If anything, climate forecast activities vary more on interannual time scales; the mix of tools and techniques used during El Niño will be different than those used during “Non-Niño”, with heavier emphasis on statistical techniques during the former and dynamical techniques for the latter.

At the IRI, (Tony Barnston, Head of Forecast Operations IRI, personal communication July 17, 2003), the forecast cycle begins after the Thursday nearest to the middle of the month, with the anticipation of the Sea Surface Temperature (SST) anomaly forecasts. These anomalies are stitched together from separate dynamical and statistical predictions for the tropical Pacific, Indian, Atlantic Oceans and others, some of which come from external groups. Ocean temperature anomalies have strong persistence and it is difficult for short-range ocean temperature forecasts to outperform persistence. Therefore, in parallel, the IRI also gets the most recent observed SST anomalies and persists those forward in time.

The various sets of SSTs then form the boundary condition for runs in up to 6 atmospheric General Circulation Models (GCMs). Most of these GCM runs are done by external groups at the Scripps Institute for Oceanography, the Center for Ocean Land Atmosphere studies, the NASA Goddard Space Flight Center, and a climate forecasting group in Queensland, Australia. This stage of the process often takes up to a week as invariably one of the groups has technical difficulties and is delayed.

The data returns to IRI where the forecasters apply sophisticated automatic statistical post processing techniques to combine the various model outputs and to censor out regions on the maps where the models have a poor track record of skill. In total, over 60 forecast maps are combined to produce maps of results by individual model, and a single map of all models together. This last map serves as a rough draft for the final forecast product, although the individual model maps can be consulted if the forecaster is curious about what they contain. This part of the cycle may be different for the CPC, which relies heavily on statistical forecasting techniques. The conceptual underpinnings of CPC's tools are described in chapter 4.2.2 and in Hartmann et al., (1999).

Once a month, about 7 working days before the forecast issuance, the CPC hosts a conference call in which 5-10 of about 20 interested parties across the nation, including the IRI, discuss the conditions for the US. Participants are guided through a CPCmaintained webpage with over 35 links to realtime data and the results of forecast tools. A forecaster begins by discussing the performance of last month’s forecast, including physical explanations for any significant forecast errors. About 1/3 to ½ of the meeting then centers around the current state of El Niño and the predictions for how it may wax or wane in coming months. Aside from the random nature of climate, the expected state of El Niño is one of the greatest sources of forecast uncertainty. Eight tools may indicate the formation of a strong La Niña, seven tools may indicate “Non-Niño” and three may show an impending El Niño. Individual experts’ opinions also tend to diverge on this point.

After showing the recent behavior of large-scale atmospheric pressure patterns across the Northern Hemisphere, the lead forecaster will continue by showing the results of several different statistical tools. The individual may have a favorite tool he routinely consults or perhaps only one or two tools are giving a strong indication for wet or dry conditions. He weighs the strength of the forecast anomalies, the track record of skill for individual tools and if there is any consensus among the different tools. In some locations (i.e., temperatures in the southwest) all forecast tools might agree and depict a strong signal. In most locations and particularly for precipitation, many of the tools lack skill and those that do have skill have conflicting guidance. In general, the climatologist is focused on “forecasts of opportunity”; the forecast map, by default, is unfilled (“climatology”) until some clear signal emerges that a skillful forecast is possible. Human judgment does play a role, although operational climate forecasters tend to be uneasy about making a forecast that does not have any support from the existing array of objective tools.

At CPC, the specifics of the forecast map show the human hand, however. Aside from narrow coastal effects, forecasters generally draw features that are no smaller than the average Western US state. The "southwest" in the climate forecaster's statement "wet in the southwest" may mean Arizona, New Mexico, the southern halves of Colorado and Utah and a sliver of Texas and California. Wet anomalies and dry anomalies almost never share a border, in that a strip of climatology/indeterminacy will always separate them. Five percent probability anomalies usually indicate that there is limited consensus among tools, although the balance of evidence suggests that the precipitation may favor, say, wet as opposed to dry. Ten percent probability anomalies are often used to denote the epicenter of where a forecaster thinks that an effect may be strongest or where he is the most confident. For precipitation, probability anomalies greater than 10% rarely occur, if ever, outside of El Niño events. During El Niño, the specifics of these strong probability anomalies are often guided by historical frequencies of occurrence of various precipitation amounts.

In other words, unless a strong El Niño is underway, CPC forecasters are rarely literal about the meaning of the probabilities associated with the forecasts. Gradations of probability anomalies are used to indicate forecaster confidence (or lack thereof). Gilman (1982,1986) describes the semi-quantitative method for creating probabilistic operational climate outlooks in the 1980s. Probabilities are determined by balancing the observed performance of the official categorical forecasts issued 1959-82 with the degree of consensus among four forecasters (who independently drew up their own categorical forecast map). Maps of maximum potential predictability from Madden (1981) were used to reign in overconfident forecasts. While this technique is no longer in use, it captures the spirit of what is done today.

After the multi-agency conference call, three IRI forecasters sit down with the multi-model objective forecast maps and discuss if and how to make any changes. They consider very recent changes in SST that would have affected the model runs, had they been known at the time the runs were made. If an El Niño event is under way, they subjectively blend the model results with historical statistical composites and analysis. They also defer to local experts (e.g., CPC for the US or climate forecasters in India for their region).

Small groups from the IRI and CPC revisit by conference call two days before the forecast issuance to discuss the first draft of the actual forecast maps. The next morning, the lead forecast individuals from the IRI and CPC review finer points about the forecast, mostly about the state of El Niño and sub-seasonal (30-60 day) oscillations in tropical precipitation. This is then followed up by another IRI/CPC group conference call to review the final draft of the forecast maps over the US, after which the forecasts are sent to media and graphics support personnel. The final forecast maps are released to the public on the Thursday nearest to the middle of the month at 15:15 eastern time (although this may change in the future to earlier in the morning to accommodate domestic financial markets).

4.3 Western water supply outlooks

4.3.1 Current products

Currently, the Water Supply Outlooks (WSOs) are issued jointly by the NWS River Forecast Centers (RFCs), the USDA Natural Resources Conservation Service (NRCS) and, in certain basins, the Salt River Project (SRP). These forecasts are available in print "Basin Outlook Report" publications or on the Internet at http://www.wcc.nrcs.usda.gov/wsf (see also chapter 2.6).

These forecasts have been issued in generally the same format since the 1930s. The WSOs are issued from January through June and they predict the volume of naturalized streamflow for various target periods. In the Pacific and Interior Northwest the target period is April-September, at the West's midsection it is April-July. Due to the earlier melt in the Lower Colorado Basin, the forecast period begins sooner and decrements each month to include only the future. Specifically, a forecast issued 1 January has a target of January through May and a forecast issued 1 April has a target of April-May. Many locations have multiple target periods to satisfy the interests of multiple users (e.g., reservoir operators versus irrigators). A limited number of single month forecasts exist (e.g., January volume) as well as forecasts for exotic variables, such as the rise in elevation of a specific lake or the date that the daily flow will fall below a certain threshold (e.g., 200 cfs).

For a given location and target period, the forecast consists of the volume corresponding to each of the 10%, 30%, 50%, 70%, and 90% exceedence probabilities.

The 50% exceedence probability number (median forecast) is a popular "shorthand" for the entire distribution; In some publications, this value is referred to as the "most probable" forecast, although the term is not statistically accurate nor is it the preferred terminology. The publications also generally include values for the historical (e.g., 30-year) normal streamflow for each basin, and the median forecast streamflow expressed as a percent of the historical normal. The historical normal is the long term average except in Arizona where it is the median.

The NRCS produces several derivative products such as color-coded maps of the median forecast expressed as percent of the historical average. It also produces horizontal bar charts whose lengths are proportional to each of the probability of exceedence levels described above (see also figure 9.1). Another derivative product, the Surface Water Supply Index (Doesken, et al. 1991; Shafer and Dezman 1982) is supported in some states but not others. There are many variants to the SWSI, but a popular version is described in Garen (1993). The current reservoir contents are added to the water supply outlook volume, and the ranking of this total volume in the historical distribution is determined. This exceedence probability is then rescaled from –4.2 to 4.2 so that the index has a range similar to the Palmer drought indices (-4.2 indicates the driest year on record). The SWSI is particularly useful in placing the seasonal water supply forecasts in the context of the total water availability on the watershed.

In the near future, long-lead predictions of peak flows, low flows, and number of days to a particular flow threshold will be routinely produced through the NWS Advanced Hydrologic Prediction System (AHPS). The AHPS Ensemble Prediction

System (ESP) involves the calibration of a hydrologic simulation model, model initialization using current watershed states, and forcing based on a number of observed historical meteorological traces. The output is a series of “possible future” daily hydrographs, from which the above mentioned characteristics can be derived. The NRCS NWCC is also actively developing this kind of capability, including an advanced spatially distributed hydrologic simulation model. The delivery mechanism and visualization tools for this kind of forecast are yet to be determined. The Colorado Basin River Forecast Center is testing an operational prototype to display ensemble daily flow trace forecasts, although their visualization tools are in such rapid development that any description herein would soon be outdated.

4.3.2 Operations research and evolution of forecast techniques

As mentioned previously, Dr. James Church began the first program of systematic western snow surveys in 1906. Church continued his program of measurement when heavy snows in 1910-11 and the threat of flood concerned the Sierra Pacific Power Company. The company urged Church to translate the snow measurements into a seasonal streamflow forecast. This forecast, the first of its kind, involved determining the percent of normal of snowpack and directly relating that to the percent of normal streamflow. H.P. Boardman, professor of civil engineering at the University of Nevada, developed the forecasting equations. The forecasts were well received until 1915-1916: the first “busted” forecast. A near complete absence of spring precipitation caused a divergence between the forecast and observed streamflow of 50%. So too was born the first irate water supply forecast user. However, cooler heads prevailed and the venture continued on (Church 1937; Poulton 1964). Over the years, minor adjustments were made to the snow sampler (such as its dimensions) but Church's overall design remains the same.

Snow survey programs appeared in 6 regions from 1917-29 (Marr 1936). Each state produced its own water supply forecasts and delivered these products to users by post or radio address. In the mid-1930s, on the back of the Dust Bowl, the NRCS (then called the Soil Conservation Service) was formed and took the reins of snow surveying activities. Helms (1992) provides an excellent history of the evolution of the snow survey data-collection program after the formation of the NRCS.

Until the 1960s, statistical techniques that were primarily snow-based remained the only tool for water supply forecasting (Garen 1992). The earliest efforts before the 1940s based the forecasts on the components of a water balance with parameters derived from basin characteristics. In other words, these hydrologists tried to explicitly measure the volume of snowpack, the expected volume of evaporation, soil infiltration and other water balance components. Soon after, the merits of automatic regression techniques were recognized. "Graphical" procedures were popular in the Bureau of Reclamation and to a lesser extent in the NRCS. The hydrologist would draw and trace along a multi-step nomograph (i.e., begin with 1 April snow, follow across a graph until intersecting a line for fall precipitation, then trace down until intersecting a line for spring conditions then trace across for the streamflow forecast).

The operational literature after the 1940s then focused on three questions. First, which forecasting data are better, snow surveys or accumulated precipitation? The former was the mainstay of the NRCS, the latter was that of the NWS. In an operational sense, each agency had easy access to its own data. However, the NRCS contended that precipitation data accumulated in valley floors is not necessarily related to the water content on the watershed. Most of what will eventually become streamflow is temporarily stored in high elevation snowpack. The widespread homogeneous synoptic scales of winter precipitation systems make this a minor issue in practice. Ultimately, the answer is that both are valuable, but that one gets very similar overall skill using only one or the other data stream (see also chapters 7.2-7.7). Precipitation information does seem to be more important in far southern basins, such as Arizona and New Mexico (Lettenmaier and Garen 1979), while there are data quality issues with unshielded NWS gages in northern climates.

Second, what data besides winter snow and precipitation can be used to forecast streamflow? The issues of sublimation and soil moisture have vexed water supply forecasters for generations. While many studies of blowing snow and sublimation processes have been done, particularly on snow fences and forest-thinning practices, the interannual variability of sublimation has never been well measured or accounted for in forecasting. Wind fetch data were used but never caught on. Baseflow and fall precipitation are popular proxies for soil priming and soil moisture. The NRCS collected soil moisture data at its snow courses sporadically in the 1940s to 1960s. The data were difficult to obtain and the short time series of observations made them of little use in regression forecasting. The soil moisture data collection effort was eventually abandoned.

Degree days and spring temperature data piqued the interest of some forecasters, but this information had more of a coincident, rather than predictive, effect on streamflow amount and runoff efficiency. The fringe of forecasting research touched on sunspots in the 1970s to little avail (James Marron, NRCS National Water and Climate Center, personal communication, May 2003). Studies even showed that winter streamflow data in low-elevation rain-fed watersheds in western Washington and Oregon could compete with snow indices in forecasting seasonal flow on the Columbia River. All of the above issues are discussed in Army Corps of Engineers (1956) and CBIAC (1961,1964). Most analysis shows that, with respect to seasonal streamflow forecasting, observed snow and precipitation account for the lion's share of predictability. Using exotic variables gives diminishing returns on skill.

Third, some agencies used "future variables" in forecasting whereas others avoided it. For example, an equation for forecasting April-September streamflow would include 1 April snowpack, fall precipitation or baseflow, and an index of spring precipitation and/or temperature. On 1 April, one generally does not know the character of spring precipitation yet to come. The forecaster either assumes the long-term normal precipitation, or adjusts it away from normal based on expert judgment or a climate forecast. A forecast issued on 1 March also extrapolates the current snowpack to what it may be on 1 April, perhaps by assuming normal snowpack accumulation or a persisted trend. In contrast, other agencies would simply not use any "future variables" and use 1

March snowpack to predict streamflow directly without extrapolating the snow to its expected value on 1 April.

Both techniques have advantages and disadvantages. “Future variable” equations are convenient to be able to ask “what if?” questions about the forecast, such as, “how would 150% of average spring precipitation change the streamflow outlook?” or “What precipitation will we need to get normal runoff?” It was also perceived that the former technique was more consistent from month to month than the latter because one forecast equation was used throughout the season, and there wasn’t the possibility of forecast discontinuities that may arise from having different equations for each month. However, hydrologists have relatively low skill in assuming anything but normal about the future climate, especially if the adjustments are subjectively based (Pagano and Garen 2005a; Hartmann et al. 2002a). Garen’s (1992) analysis also allayed concerns about forecast consistency by showing that, in practice, each technique is just as likely as the other to produce forecast “waffles” (changes in forecast direction) throughout the season. Since the 1980s and particularly in the 1990s, “future variable” equations (i.e., equations that include quantitative estimates of future precipitation amounts) have waned in popularity. There was some use of “future variable” equations by the NRCS in Arizona and by the NWS in several other regions (Tom Perkins, NRCS National Water and Climate Center, personal communication, 5 August 2004). Today the NRCS does not use any “future variable” equations although this methodology does still find limited use in the NWS, particularly in the Pacific Northwest.

In the 1960s and 1970s computers entered into the realm of operational water supply forecasting, beginning with statistical experiments in Montana (Codd and Farnes 1960; Johnson 1960). In time, focus was turned to conceptual and physically based watershed simulation modeling. The primary priorities of the NRCS snow survey and water supply forecasting program in the mid 1980s were to discontinue snow courses in favor of automated SNOTEL sites, and to implement simulation modeling for forecasting. Remote sensing of hydrologic information was also very popular in the research community in the 1980s (e.g., Deutsch et al. 1979). While the NRCS has completely integrated computers into its operations, it has fallen short on simulation modeling and the use of remote sensing data. None of the many NRCS simulation modeling experiments (e.g., Jones et al. 1981; Marron 1986; Perkins 1988) have been sustainable. However, the NWS has implemented the Sacramento model in its operations. Neither agency uses satellite data because of the inability to see snow under clouds and forests. Satellite information is also not a convenient or timely data stream in the operational environment.

Most recent operations-oriented research (as presented at the annual Western Snow Conference) focuses on one or more of four topics. First, efforts are under way to develop high quality spatially distributed maps of snowpack (and to a lesser extent precipitation), based in part on remotely sensed satellite data. Next, the descriptions of snowpack processes in simulation models are being refined through data gathered in field experiments (e.g., Brubaker et al. 1996) This also includes studies that set up and test simulation models on individual basins. Third, GIS technology is being used in the context of spatial snow maps and hydrologic modeling. Finally, several researchers are looking at the relationship between climate, snowpack and streamflow, with an eye also on long-term trends and decadal variability, similar to this dissertation. Occasionally, research on statistical methodology (e.g., Garen 1992), non-linear, fuzzy logic, or neural network forecasting techniques is reported.

The Western Snow Conference is also a forum for presenting research results of natural resource studies, such as on the relationship between water yields and forest management practices, which are of secondary interest to water supply forecasters. A small amount of NRCS-sponsored research is being done on advanced in-situ snow sensor technology, in hopes of eventually replacing SNOTEL snow pillows with fluidless sensors that are more environmentally benign and are less sensitive to diurnal temperature variations. Very little research, most of which is anecdotal, focuses on the use of water supply forecasts or how users perceive and interpret them.

Regrettably, until the NRCS adopts a spatially distributed simulation model, its water supply forecasters cannot take much advantage of many research advances, such as spatially distributed snow maps. However, it cannot adopt a simulation model until the spatially distributed data are of sufficient quality, operationally attainable and easily processed. Until this gridlock is broken, results from these major thrusts of the research community, besides climate, are practically inaccessible to NRCS operations.

4.3.3 The operational hydrology environment

In 2002, four NRCS hydrologists were responsible for forecasts at over 700 locations. One hydrologist is responsible for the Missouri Basin and Platte River, while another is responsible for the Great Basin in Utah and the Columbia. The third hydrologist is responsible for coastal basins from Lake Tahoe to the Olympic Peninsula and the Cascades flowing west. The author is responsible for 206 locations in the Upper and Lower Colorado River Basin, the Rio Grande, the Arkansas River, the Yukon and Alaska.

Each hydrologist coordinates with counterpart hydrologists in developing forecasts. For example, the author coordinates with four NWS hydrologists in the Upper Colorado Basin, one NWS hydrologist and the Salt River Project hydrologist in the Lower Colorado, one NWS hydrologist in the Rio Grande, and two NWS hydrologists in the Arkansas. While there is no counterpart NWS hydrologist in Alaska and the Yukon, the NRCS snow survey supervisor for Alaska participates in the development of those forecasts. The operational perspective, herein, is that of the NRCS hydrologist.

At the start of the water year, 1 October, the hydrologist is typically involved with developmental activities including but not limited to touring watersheds and snow measurement sites, interacting with user groups in the field, miscellaneous research activities and redevelopment of forecasting equations. The balance between off-season "research" vs forecast environment maintenance depends on the individual forecaster's personality. Under special circumstances, the hydrologist may be addressing an exceptional event with ad-hoc forecasts or analysis on demand, such as the effects of a recent fire on runoff efficiency or the expected impacts of El Niño on the coming season.

Forecasters have broad liberty to develop forecast equations as they see fit. Equation development is a balance between several factors. Variable selection is guided by a desire to have complete geographic coverage of the basin (across the land and with elevation), recognizing that not all sites have a long period of record. Data sites within the basin boundaries are favored over those outside the basin boundary. SNOTEL sites are favored over sites from other agencies (such as NWS Cooperative Observer data); because SNOTEL data are easily accessible in realtime during forecasting, whereas NWS data may arrive several days after the start of the month.

In general, the equation developer is encouraged to use sound hydrologic judgment, as subjectively interpreted by the developer. The hydrologist presents a list of candidate variables to the development software, RegComb (Garen, n.d.), which then returns a list of the 20 best variable combinations. The hydrologist then selects one or more equations that have a set of variables that are generally consistent from month to month. This prevents the discontinuities that would arise if site “A” is used to create a forecast in January and site “B” in February.

Each forecaster has different perspectives and preferences on the importance of, for example, record length compared to complete geographic coverage. This issue is important with respect to SNOTEL sites that have 25 or fewer years of data. Some hydrologists feel comfortable with an equation developed using as few as 12 years of data whereas others are more conservative. The NWS Colorado Basin River Forecast Center has a policy of developing equations using only serially complete observations from 1971-2000 to be consistent with the period used by the NWS to define climatological “normal”. This implies that no SNOTEL data would be used in equation development, but operationally, the SNOTEL data has been back estimated using snowcourse data.

One positive aspect of this system is that the forecasters have flexibility to innovate and develop a system that they feel works best for their region. The drawback is that there is less uniformity and consistency across forecast regions than there would be otherwise. One forecaster may have a complex set of equations, using fall precipitation, antecedent streamflow, climate indices, several snow sites and spring precipitation as variables. Another may have a limited set of equations only based on snow, with the expectation that the results will be manually adjusted at forecasting time based on qualitative information about those other factors. Although the general policy is complete redevelopment of all equations every three to five years, a forecaster may even inherit decades-old legacy equations from previous forecasters.

The snowpack accumulation season starts in October and November, depending on the region, and the hydrologist measures the season’s early progress on relatively infrequent intervals (i.e., twice a month). In December, the hydrologist can begin using the software environment (the FCST program) to produce forecasts. Procedures are available (the Loadswe program) to extrapolate December snowpack at any given SNOTEL site to what it might be on 1 January. These extrapolated data are then used in the 1 January forecast equations. The results may be shared internally or with other forecasting agencies and users on an informal basis. Some, but not all, NWS River

Forecast Centers have been known to put such “early bird” forecasts on their Internet webpages.

The NRCS water supply forecasting software produces ascii text tables of forecast equation coefficients, inputs, and outputs. The latter two are expressed in volumes and as percent of the long term average (Table 4.1). The forecast tables are the only uniform guidance available across NRCS forecasters. Individuals, on their own initiative, may draw in external information from various sources. In the throes of forecasting the hydrologist does not have easy, immediate access to historical NRCS data to analyze, for example, the historical probability of exceedence of a snow measurement, or a list of historical years with similar snow conditions. Some, but not all NWS River Forecast Centers have this information available interactively on their webpages. The forecaster may draw a map of basin snow conditions; GIS data visualization support is not officially provided, although some forecasters have taken the initiative to struggle with this technology. Within three years, the NRCS forecasting software environment will be redeveloped and the data visualization issues may be addressed (see also SCS 1988 for a more detailed operational forecast improvement plan).

Table 4.1. Abbreviated 1 June 2003 water supply forecast guidance form for the Pecos River near Pecos, NM. The forecast target is Mar-July volume in thousands of acre-feet (k-ac-ft). Explanation at bottom.

```
Pecos River nr Pecos                                                MAR-JUL VOLUME


T Mnth     SITE NAME         I.D.       COEF        VALUE      AVERAGE     %
----------------------------------------------------------------------------
P -9  GALLEGOS PEAK SNOTEL   S243       1.893        1.80         2.28     79
P 10           "             S243       1.893        3.00         2.03    148
P  4           "             S243       1.298         .90         2.55     35
G -9  PECOS R NR PECOS       378500      .287         .00          .00    107
G 10           "             378500      .287        1.50         3.20     47
G  3           "             378500     4.205        2.20         2.90     76
G  4           "             378500     1.281        5.90         7.10     83
G  5           "             378500      .588       22.30*       22.30    100
S  3  GALLEGOS PEAK SNOTEL   05N18S     2.368       10.00         9.50    105
S  5  WESNER SPGS SNOTEL     05P08S      .636        8.20        11.30     73
S  6           "             05P08S      .636         .00         2.70      0
C     --   INTERCEPT    --            -19.385        1.00          .00    100


                    FORECAST =    50.0   (  86.0%)         AVERAGE =    58.0
                     (  10%) =    58.0   ( 100  %)              SE =   6.469
                     (  30%) =    53.0   (  91  %)        MIN FLOW =    8.59
                     (  70%) =    47.0   (  81  %)
                     (  90%) =    42.0   (  72  %)


             JANUARY    FEBRUARY      MARCH      APRIL        MAY
          Volume   % Volume   % Volume   % Volume   % Volume   %
Equation    86.2 149   58.0 100   70.0 121   72.0 124   65.0 112
Published   61.5 106   55.0  95   62.0 107   58.0 100   54.0  93
```

Column 1: Data type (P = precipitation, S = snow, G = streamflow, C = constant).
Column 2: Measurement month (i.e., G10 = October streamflow)
Column 3,4: Station name and identifier code.
Column 5,6: Regression coefficient and this year's data value (*=missing, long term normal used as substitute). Units: P, S in inches, G in k-ac-ft.
Column 7: Long term average of this data value
Column 8: The current data value as percent of average

In the bottom half of the table are the various forecast probability of exceedences (as volumes and a percent of the long term normal), as well as ancillary statistics about the average flow, historical minimum, and equation standard error. The bottom section traces the evolution of the preliminary (top) and official (bottom) forecast values by issue month, as a volume and as percent of average. The official guidance was consistently lower than the regression equation throughout this season.

On the first working day after 1 January, the author has about four working hours to acquire preliminary data, create, analyze, and adjust forecasts for close to 100 points on the Upper Colorado Basin. The author is then contacted by the Colorado Basin River Forecast Center for an hour-long conference call to develop preliminary forecasts for 13 major downstream locations, such as the inflow to Lake Powell. This coordination involves a discussion of basin snowpack and soil moisture conditions, with mention of the seasonal climate forecasts, followed by the hydrologists presenting their desired forecast numbers. It is impossible to generalize, but on the Upper Colorado Basin, the disagreement is typically within 5% of the average flow. In the San Juan it can be as much as 10% different, and in Arizona the discrepancy can be much wider in part because the flows there are more variable.

Rectifying the forecasts and agreeing on a final number is a subjective process and varies by region and forecaster. The forecasters may state their reasoning for and confidence in their forecast, cite ancillary analysis they have done, offer historical analogues and anecdotes or probe for weaknesses in the other hydrologists technique (e.g., “do you account for the lack of low elevation snow?”). The NWS hydrologist may compare the output of his regression equations and the median of the ESP distribution. Sometimes, out of convenience or an inability to agree, the forecasts may simply be averaged. In the rare instance of contentiousness that precludes coordination, each group has an official site list of which agency has the final authority over the forecasts.

Typically, the NRCS is responsible for headwater locations and the NWS has control over downstream forecasts into major reservoirs.

In the next 1.5 working days, new data filter in from NRCS state snow survey personnel and the hydrologists' analyses continue. NRCS state personnel receive reports of manual snow course measurements and conduct their own manual ground-truthing of snow measurements at SNOTEL sites. Using human judgment and a limited set of simple quantitative tools, the state personnel also adjust or estimate data from problem sites. As necessary, a forecaster may also estimate, adjust or ignore any piece of datum, although these changes are not recorded in the official database archive. Each forecaster may deal with 200-300 unique snow measurements per month, depending on the complexity of his forecasting equations and region.

Throughout the analyses, the first question the hydrologist is typically trying to answer is "How much snow is in the basin and where?" with subtexts about data quality and spatial representativeness of SNOTEL sites. Next the hydrologist may ask, "How much will soil moisture deficits affect the runoff efficiency of the snow?", "What can be said about the future climate being wet or dry?" and finally, "What is not accounted for in the forecast equations, that could have a significant impact on runoff (e.g., the major storm in the weather forecast)?" In the midst of forecasting, there is little time to find anything more than a superficial answer to any of these questions for a limited number of locations. Therefore, some hydrologists try to begin this analysis before the start of the forecast season or during the middle of the month before the next forecast cycle.

The coordination discussion continues either by E-mail or telephone and a full set of coordinated forecasts is typically available for the Upper Colorado River Basin on the third working day of the month. A final set of forecasts for the other 80 forecasts on the Lower Colorado, Rio Grande and elsewhere may follow a day or two afterwards. Coordination on the Columbia basin may take up to a week.

The discussions are clearly focused on the 50% exceedance probability (or "most probable") forecast and the forecasts are framed in deterministic terms (often as percent of average). After this number is agreed upon, the other probability of exceedance forecasts are determined. This probability envelope is based on the root mean squared error (RMSE) of the forecast equation during jackknife calibration and assumes a normal distribution unless the forecast equation is non-linear (NRCS 2004). Although the statistically savvy may question the validity of the practice, the forecast error bound is centered on the coordinated 50% exceedance probability forecast. Often, out of convenience, one hydrologist may agree to accept the bounds from the other hydrologist for all locations. The bounds typically do not receive much attention unless the results are grossly physically unrealistic. Examples include when the lower bound is lower than the minimum flow on record, is negative or, in the case of a forecast to be issued in May or June for a forecast target starting in April, is less than what has already been observed (implying negative flows into the future). No standard procedure exists for determining the new bounds in these situations and the forecaster generally improvises. As of late 2004, an internal document about NRCS "best practices" in equation development, forecast adjustment and bounds determination was nearing completion (NRCS 2004).

The finished set of coordinated forecasts are sent to the NRCS state personnel who develop the text and supporting information for the State Basin Water Supply Outlooks. Within a day or two, this information is returned to the hydrologist, who assembles the text with the forecasts and posts the information to the Internet. State personnel also then send this information to print shops and eventually send the bulletins out in the post. Westwide publications can only begin when all regions have been finished, and as such are typically posted to the Internet around the 10th working day of the month. The agency is under constant pressure to reduce the time it takes to create and deliver the forecasts. Users have been known to begin contacting the NRCS as early as the second working day of the month, inquiring about the delay in the forecasts. The phrase "1 January" forecast can be misleading to users who expect the forecast to be issued and distributed on the first of the month; the date refers to the data used to develop the forecast.

The process begins again every two weeks, on the 1st and 15th of each month until May. The NWS continues to issue and the NRCS may coordinate on forecasts into mid July or later for select locations. This rapid cycling affords the hydrologist limited time to engage in other operational activities, such as fielding requests from users and questions from the media, visiting the field to inspect basins and give presentations to user groups, prepare special reports on the status of the season or do developmental activities. The forecasts issued on 1 April are generally thought of as the most critical water supply forecasts of the season, with those issued on 1 May a close second. Beginning in April, the hydrologist is inspecting the status of the snowpack, its melt rate and realtime streamflow data, on a daily to hourly basis. Aside from a procedure that seamlessly integrates climate information into the water supply forecast process, forecasters would be most receptive to climate information from an external source during the week before the 1st of the month.

In absence of an exceptional hydrologic event, in mid-June or after the streamflow peak has passed, the forecaster's attention returns to research and developmental activities, such as those being done at the beginning of the water year. A subset of users begin to crave early season forecasts in September and, if he chooses, the hydrologist may provide this guidance directly, or at water users meetings that typically occur in the fall.

4.4. Summary and conclusions

Water supply and climate forecasting have many aspects in common. Much like in other fields, such as economics, the forecasting process is a three-legged stool. The forecaster observes the system state (be it the status of snowpack or the global patterns of temperature in the ocean). Analytical tools and computer models process the data to provide objective guidance. Finally, the forecaster applies intuition and professional expertise to the results. Both fields are vulnerable to data quality issues and both must deal with uncertainty and the random elements of nature.

Climate forecasting is a fundamentally much more uncertain enterprise, however. Climate is a chaotic system, determined primarily by internal feedbacks such as the interaction between the ocean and atmosphere. Hydrology is a damped system, integrating processes over time (seasons) and space (watersheds). Both are sensitive to initial conditions but with different consequences. While hydrologists may disagree about whether a forecast should be for 63% of average or 68% of average, a climatologist may have three tools that say the future may be wet and five tools that indicate dry for the same location. Which tool(s) should the climatologist believe? The confidence in water supply forecasts is thus much greater than climate forecasts, as are their skill (Pagano and Garen 2005a; see also Chapter 6 and 7).

Perhaps out of necessity, climate forecast technologies are far more complex than those used for water supply forecasting. The NRCS hydrologist may have one or more regression equations per location. The NWS hydrologist may also have an ensemble forecast from a simulation model developed in the mid-1970s. In the United States, the climatologist has realtime operational guidance for precipitation from no less than ten different statistical techniques and twelve simulation models. The data stream of the climatologist is orders of magnitude larger than hydrologist.

This is not to say that the hydrologist is any less harried, as both forecasters' schedules are measured in hours to minutes. Climatologists have made better use of automatic data processing and data quality screening technologies. They also have vigorous support from the academic and research communities, routinely drawing in quantitative guidance from 17 external groups for ocean temperature predictions, and at least six groups for their precipitation forecasts (Tony Barnston, IRI, personal communication, 2003). Streamflow forecasting is a relatively isolated activity, in part because there is little incentive for an individual in Arizona to be concerned with the quality of predictions in Montana. In contrast, international forecasting groups that pool their resources to improve predictions of El Niño can all benefit.

All of the internal and external climate guidance is macro-scale in nature. With limited exception, the climatologist is analyzing continental to hemispheric maps compared to the hydrologist who is inspecting tables of data within individual basins. A small group of climatologists form a draft national forecast map and a broader audience of experts refines it. Hydrologists closely scrutinize conditions in their local region, and the final national forecast map is stitched together at the end with little cross-basin communication (or interest).

Hydrologists may benefit from the social science and usability research done by climatatologists. Indeed, aside from theoretical and economic studies (such as Held and Jacobs 1990; Kim and Palmer 1997) the seasonal water supply forecasting literature is practically devoid of any studies of its users and their issues. Climate forecasting grew out of an academic exercise and as such climatologists needed to search for and entice resource managers to use their forecasts (Hartmann et al. 2002b). In comparison, water supply forecasting grew out of its demand and that its links to its users are well understood, and its users satisfied (Lettenmaier 2004). Nonetheless, very long-range streamflow forecasts are likely to be much more uncertain than the traditional semi-deterministic forecasts and will require a format that more effectively communicates probabilistic information. In this regard, climatologists have more experience than hydrologists do.

Finally, there are parallels between the overall evolution of forecast tools in climatology and hydrology. Both fields have a tacit respect for the forecaster who understands physical processes and relies on intuition. The master climatologist is skilled at visualizing the evolution of features on a pressure height chart and the master hydrologist is intimately familiar with the character and features of a basin, and has an intuitive feel for, for example, the non-quantified effects of long-term soil moisture deficits. Both fields also recognize that intuition cannot be the only source of guidance and there is a need to quantify and objectify. Although statistical forecasting techniques are the mainstay of both operational environments, they are looked at as temporary measures on the road to the eventual goal of full simulation modeling. Simulation models are resource intensive, difficult to run and require significantly more care and feeding than statistical tools. One potential advantage of simulation models is their ability to manage situations outside historical experience (e.g., climate change, major landcover changes). Their adoption will not bring a quantum leap in operational skill in all situations although model improvement is a major thrust of the research community and will likely continue to be so in the foreseeable future. The formation of the IRI, a separate entity committed to seasonal forecasting using simulation models, represents an interesting development in the operational climate modeling gridlock. Hydrologists should track the activities of this agency and learn from its modeling experiences. One can only speculate if users would benefit from a similar public institution devoted to multi-model hydrologic simulation modeling.

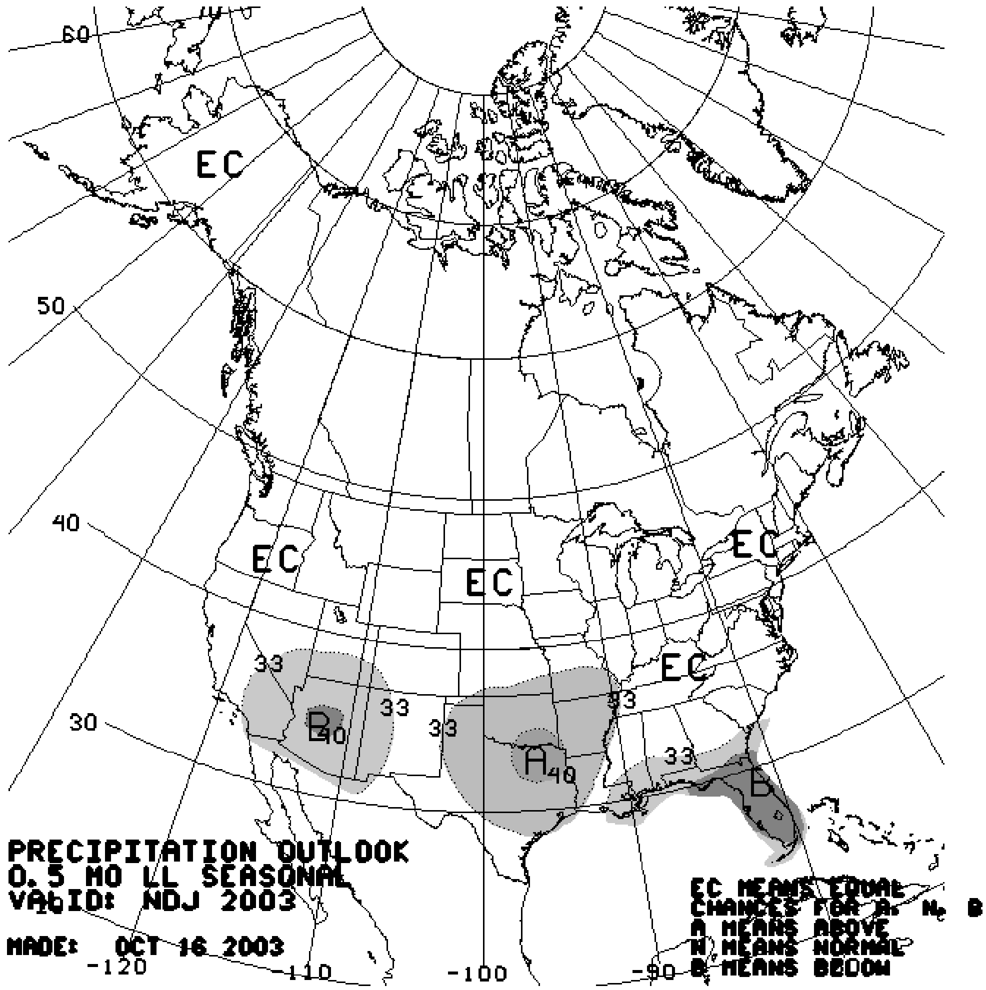


Figure 4.1. Example CPC seasonal climate outlook issued October 16th 2003 for November-December 2003 precipitation. "EC" indicates an "equal chances" forecast whereas the area within the "B 40" contour has a 40% chance of falling in the dry (below normal) category, a 33% chance of near-normal and a 27% chance of wet (above normal).

5. HISTORY OF LINKING OPERATIONAL CLIMATE FORECASTS AND WATER SUPPLY FORECASTS

5.1 Introduction

While there are significant contrasts between the operational climatology and hydrology environments, hydrologists have long recognized the potential gain in using climate information and forecasts. This chapter reviews the use of climate information in operational water supply forecasts at the Natural Resources Conservation Service's National Water and Climate Center. The use of climate information in other operational hydrologic forecasting agencies, such as the National Weather Service and the Salt River Project, are also highlighted. This review focuses on three main periods, from early attempts to link climate and water supply forecasts, through the period of widespread operational use of climate indices in NRCS forecasting equations, and finally the recent age of greatly expanded climate forecasting efforts. Chapters 4.2.2 and 4.3.2 traced the evolution of forecast tools in individual disciplines, whereas this chapter traces the evolution of linkages between the disciplines. The results from this chapter are used to focus and place in context the results of investigations in chapters 7-9.

5.2 Early history (1935-1983)

Taking advantage of seasonal climate forecast skill has been a long-standing moderate to high priority within water supply forecasting agencies. Church's (1935) seminal publication about snow surveying and water supply forecasting identifies precipitation variability during the runoff period (after forecast issuance date) as the largest source of forecast error. Schaake and Peck (1985) estimate that for the 1947-1984 forecasts for inflow to Lake Powell, almost 80% of the 1 January forecast error is due to unknown future precipitation; by 1 April, Shaake and Peck find that future precipitation still accounts for 50% of the forecast error.

There has also been a long history of attempting to incorporate seasonal climate forecasts into operational water supply forecasts. The NWS (then called the US Weather Bureau) started creating bi-monthly 30-day weather outlooks for internal use in 1943. In 1953, they began issuing these forecasts to the public. Shortly afterwards, the Columbia Basin Interagency Committee (of which the NRCS was a member) evaluated their usefulness in forecasting Columbia River streamflow (CBIAC 1955). This report concluded that the potential benefit was great but that the actual skill was too low for practical use. Specifically, the report was concerned about the increased chance of incurring a major forecast "bust" when using the climate forecasts versus existing practices (i.e., assuming near-normal future precipitation). A later report (CBIAC 1964) revisited the issue and found that forecast skill was improving, but it was still not satisfactory for operational considerations, particularly when applying the climate forecasts to geographically small basins (i.e., “downscaling”).

The 1970s were an active period in climate and water supply forecasting. Although produced internally since 1958, the first seasonal (i.e., 90-day) temperature forecast was released to the public in 1974, with the first seasonal precipitation outlook following in 1978. As part of the long-range streamflow forecasting “Project Hydrospect”, which began in 1971, William Arvola of the California Department of Water Resources (CDWR) reviewed all historical and ongoing research in seasonal climate forecasting (Arvola 1975; Peters 1984). In the fourth year of Project Hydrospect, CDWR began sponsoring research by the famed climatologist Jerome Namias, then at the Scripps Institution of Oceanography. The technical linkages between climate and streamflow forecasting in California grew in sophistication (Zettlemoyer 1982). These activities also inspired Jim Marron, an operational NRCS water supply forecaster, to use the Southern Oscillation Index to forecast streamflows around Lake Tahoe and in Nevada beginning in 1976. Marron soon abandoned the practice because of the Southern Oscillation Index’s lack of predictive skill in that region (James Marron, personal communication, May 2003).

Among the several other early attempts to use climate information in water supply forecasts, Schaake (1978), in northern Virginia, used the 30-day precipitation outlook in October 1977 to remove a series of “anti-analogues” from the available ESP input meteorological traces. Similarly, in the mid-1980s, Croley and Hartmann (1987) used climate outlooks subjectively to alter ESP traces in forecasting Great Lakes levels. This method has evolved into the objective procedures described by Croley (2000). In managing Lake Okeechobee, the South Florida Water Management District also employs this climate outlook-weighted ESP forecasting technique (Cadavid et al. 1999). The NWS currently has a variety of procedures for climate-weighting its ESP traces, ranging from a simple technique developed in 1995 by Larry Rundquist at their Alaska River Forecast Center (Werner et al. 2004) to the complex method of Perica et al. (2000). The Colorado Basin River Forecast Center is currently testing no less than five different methodologies for climate-weighting its ESP traces (David Brandon, personal communication, October 2003). Most recently, in a non-operational environment, Hamlet and Lettenmaier (2000) are routinely producing real-time climate-weighted ESP traces from an advanced spatially distributed hydrologic simulation model for basins in the western US. Similarly, the WaRSMP software package offers climate-based subsetting of its routine but non-operational ESP traces (Douglas Boyle, personal communication, April 2003). WaRSMP is a United State Bureau of Interior effort to link hydrologic simulation models with reservoir optimization models in an operational setting.

Just as the 1982-1983 El Niño was a focusing event for the climate community, the 1983 Colorado River flood was equally focusing for the water supply forecasting community (Rhodes et al. 1984). Until April 1983, snowpack on the Colorado River basin was near average, and the median forecasted inflow to Lake Powell was similarly near average (109%). An exceptionally cold and wet spring ensued, followed by a rapid warming. The observed April-July flow, at over 210% of average, overwhelmed the already full reservoir system. Glen Canyon Dam sustained severe damage to its spillway tunnels because of the high volume of water it was passing. The integrity of the dam was threatened, and plywood board extensions were added to the top of the spillway gates to hold back the flow. As subsequent analysis revealed (Pagano et al. 1999), this simultaneous occurrence of an exceptional El Niño and an exceptional flood remains imprinted in the institutional memory of water managers in the region.

5.3 Routine use of climate indices (1983-1995)

Interest in climate and streamflow grew throughout the 1980s, spurred on by research characterizing El Niño's global and regional impacts, such as Ropelewski and Halpert (1986, 1987). Cayan and Peterson (1989) investigated El Niño and Western US streamflow, which coincided with work being done by Redmond and Koch (1991) on the same topic. David Garen, one of Koch's students at the time, was also an operational water supply forecaster with the NRCS. Garen began using the Southern Oscillation Index (SOI) as a predictor variable in forecasting Columbia River Basin streamflow. Around the same time, in 1988, Tom Perkins, also an NRCS forecaster, began using SOI as a predictor in the Lower Colorado River and southern New Mexico (Tom Perkins, NRCS National Water and Climate Center, personal communication, May 2003). During this period, many other hydrologists, including counterparts in the NWS, were skeptical that factoring in El Niño information sufficiently increased water supply forecasting skill and did not adopt this practice until later. Robert Hartman, however, transferred to the NWS Colorado Basin River Forecast Center (CBRFC) in 1990 after being an NRCS forecaster during Perkins' and Garen's activities. At CBRFC, Hartman continued investigating the climate-streamflow connection and generally found discouraging calibration results in all areas except the Lower Colorado (Robert Hartman, California Nevada River Forecast Center, personal communication, May 2003).

Although myriad research publications about El Niño and streamflow appeared (e.g., Cayan and Webb 1992; Kahya and Dracup 1993; Piechota et al. 1997; among many others), operational hydrologic forecasting procedures generally remained unchanged for several years. The Salt River Project (SRP), a central Arizona water manager and a coordinator in the water supply forecasts, adopted the "Entropy Limited" precipitation model (Christensen and Eilbert 1985) in 1988. Although this model is proprietary, it appears to be conceptually similar to the statistical multiple discriminant analysis model of Young and Gall (1992). Young and Gall's model uses air temperature and precipitation data at many global sites to produce probabilistic estimates of central Arizona precipitation and runoff. SRP developed a post-processor to convert the probabilistic forecast into a deterministic forecast (Reigle 1998). SRP also conducts extensive statistical analysis of climate and winter streamflow (Skindlov et al. 2000), which are used to support water supply forecast activities. SRP hydrologists, like most operational water supply forecasters, consult the official Climate Prediction Center (CPC) seasonal outlooks and use them at least qualitatively.

5.4 A new age in climate (1995-2004)

In 1995, climate forecasts changed significantly, both in terms of their creation and methods of display. The format, which is still in use today, presents the forecasts as tercile probability anomalies for 13 overlapping 3-month forecast periods with lead times up to 1 year (see also chapter 4.2.1). While more information is presented using the current format, the methods for directly incorporating the climate forecasts have increased in complexity. In particular, disaggregating the overlapping 3-month periods into monthly values produces undesirable artificial "ringing" (e.g. filtering of the data produces an artificial periodicity in the data, particularly near sharp boundaries, Wilks 2000a). For example, a wetter than normal January-February-March forecast, followed by February-March-April and March-April-May climatology forecasts may counter-intuitively imply a drier than normal 1-month forecast for March. Although not statistically precise, Schneider and Garbrecht (2003) have developed an algebraic disaggregation that may be suitable for operational purposes. Briggs and Wilks (1996) addressed the issue of quantitatively translating precipitation probability anomalies into shifts in precipitation amounts. This methodology is conceptually similar to the underpinnings of the experimental "Probability of Exceedence" forecasts issued by CPC (Barnston et al. 2000, see also chapter 9.5). Garen (1998) and Modini (2000) attempted to ingest these "probability of exceedence"-style forecasts into the regression-based streamflow forecasting framework with mixed success. The procedure is complex, operationally intensive, and does not yield results significantly more accurate than simply using a climate index (e.g., SOI) directly as a predictor variable in a regression equation. Interestingly, the procedure involves developing equations with “future variables”, a practice that fell out of favor at the NRCS in the 1980s (see chapter 4.3.2).

Another climate index of importance in western water supply, recently developed by Mantua et al. (1997), is the Pacific Decadal Oscillation (PDO). This index describes decadal-scale sea surface temperature variations in the north Pacific (see also chapter 3.3). Originally related to fluctuations in the salmon fishery, it has subsequently been shown that it has a modulating effect on the El Niño / La Niña climate signal. The phase of the PDO (cool or warm) has a significant effect on the strength of the relationship of the SOI with winter and spring streamflow in western Washington and Oregon, the relationship being much stronger during the cool phase than during the warm phase of the PDO (Koch and Fisher 2000). By splitting the data record into cool-phase and warm-phase years, Koch and Fisher (2000) developed separate regression forecasting procedures to account for this effect. This method, however, has not yet found its way into operational forecasting, in part because of the difficulty in knowing the PDO phase in real-time (see also chapter 8.2). While PDO may excel at explaining long term variability in the historical record, the forecaster is left wondering which phase of the cycle is relevant to the impending streamflow forecast season (i.e., “Has the PDO shifted or not?”). For example, debate still exists about whether the PDO shifted some time in the late 1990’s.

When the very strong 1997-1998 El Niño occurred, attention was yet again refocused on climate and seasonal water supply issues (Pagano et al. 2000, 2001, and

2002). Comparisons between the 1997-1998 and 1982-1983 El Niño events alarmed water and emergency managers. Forecasters responded with comprehensive statistical analysis of the historical impacts of El Niño (such as NRCS 1997; Brandon 1998) and by adding climate indices to streamflow forecast equations where appropriate. Most significantly, the analyses revealed that there is not a reliable signal for El Niño in the Great Basin or the Upper Colorado River Basin above Lake Powell. While the 1983 event caused major flooding in this region, there are a greater number of counter-examples where El Niño did not bring wetter than average conditions. Research by the Bureau of Reclamation revealed an under-forecast bias for inflow to Lake Powell during El Niño years (Pagano et al. 1999). Perhaps this may be related to El Niño favoring cold April-June conditions in the Upper Colorado River basin (Pulwarty and Melis 2001). If true, then a streamflow forecast based on snowpack alone would underestimate the observed flows because runoff efficiency is increased during cold springs. During spring 1998, water managers responded to public and political pressure to prevent a repeat of the 1983 event by releasing more water from their reservoirs than what would have been called for by using only the water supply outlooks. In the end, the water supply forecasts did underestimate the observed flow into Lake Powell, but not by an exceptional amount compared to previous years. Elsewhere, the forecasts accurately anticipated a wet season in Arizona and New Mexico and dry conditions in the Pacific Northwest.

The most recent developments in the history of climate and western water supply forecasts are the 2001 La Niña and the ensuing Pacific Northwest drought. In the fall of 2000, a strong La Niña was underway, combined with the cool phase of the PDO. These phenomena together provided the strongest possible climate-based indication that the Pacific Northwest would be wetter than average in 2001. For example, at the time, the driest of the other nine La Niña/cool PDO years since 1936 on the North Fork Flathead River near Columbia Falls (Montana) had April-September streamflow almost exactly 100% of average; the wettest year on record, 1974, at over 160% of average, was a La Niña/cool PDO year. Although official climate forecasts do not consider PDO-precipitation relationships, a group at the University of Washington, among others, publicized a confident forecast for a wet winter. Official climate forecasts were more conservative but still indicated a higher chance of a wet winter/spring.

For a variety of subjective reasons, the NRCS did not issue any early season forecasts in the fall of 2000. In the end, 2001 tied or broke records for the driest year on record in the Pacific Northwest, contrary to the climate forecast guidance. The North Fork Flathead experienced its third driest year on record at close to 50% of average flow. In retrospect, water supply forecasters felt that they had “dodged a bullet” by ignoring the climate forecasts. Many streamflow forecasters have a “What about 2001?” anecdote readily available as a justification as to why they do not rely on climate forecasts more heavily (see also chapter 9.3). Although no one has documented the negative impacts of unofficial climate based streamflow outlooks during this episode, the event is imprinted on the institutional memory of Pacific Northwest forecast agencies. Despite this imprinting, surprisingly little, if any, retrospective analysis has been done on the failure of the climate forecasts during 2001.

5.5 Summary

There are many examples of previous attempts to link climate and water supply forecasts throughout the history of the NRCS. If climate forecasts were perfect, their value to water supply forecasters would be immense. However, climatologists are only able to explain a relatively small (although non-zero) amount of observed precipitation variability. Hydrologists have had perennial concerns about low climate forecast skill, as well as low climate forecast spatial and temporal resolution. Nonetheless, especially since the 1980s, hydrologists have used the available climate forecast skill to improve existing waters supply outlooks in some cases. Methodologies to link the forecasts range from subjective hedging of water supply forecasts based on qualitative climate information, to the quantitative use of climate indices and forecasts in hydrologic regression forecasting. New developments in climate-based weighting of ESP traces may appear in coming years, but other methodologies for linking climate and water supply forecasts are relatively well established and suitably stable for the operational environment.

6. HISTORICAL FORECAST ACCURACY

6.1 Introduction

Unknown accuracy hinders the use of climate forecasts by some users. Some users may only have subjective notions of how good various types of forecasts are at different locations. In finding the place for climate forecasts in the operational hydrology environment, it is useful to evaluate each kind of forecast. This chapter provides a brief description of the quality and accuracy of climate forecasts. It follows with a longer analysis of observed water supply forecast skill. These results are necessary to assess how well the objective water supply hindcasts developed in the next chapter match the behavior of the official forecasts. These hindcasts will be used to measure the relative merit of using climate information in water supply forecasts (chapter 7), and to detect and diagnose long-term climate-induced trends in water supply forecast skill (chapter 8).

6.2 Climate forecasts

Many authors (e.g., Livezey 1990; Preisendorfer and Mobley 1984; Wilks 2000b) have tracked the progress of the official climate outlooks. Therefore the reader in search of details is referred to their publications as well as the proceedings of the annual Climate Diagnostics Workshop, which includes a review of forecast skill every year.

Hartmann et al. (2002a) is a particularly relevant work to water supply forecasters; the unique aspect of that study is its user-oriented approach. For example, it selects only those forecasts and lead times that would be relevant to water supply forecasting in the Western US. It found that precipitation forecast skill is modest in those regions where El Niño exerts a strong influence (i.e., the Pacific Northwest and Arizona) and low to negative skill elsewhere, such as Colorado, Utah, and Wyoming. Lower Colorado River region climate forecasts display excellent discrimination in that when wet conditions occur, there is a prevalence of forecasts for higher chance of wet conditions and vice versus for dry conditions. Upper Colorado (and presumably Great Basin) forecasts have poor discrimination because the forecasters rarely venture far from climatology or “Equal Chances” in their forecasts. Arizona forecasts have excellent reliability, indicating, for example, when the forecast says there is a 60% chance of wet, wet occurs about 60% of the time. In contrast, of the times that CPC has forecast 60% chance of upper tercile (wet) conditions in the Upper Colorado basin, upper tercile conditions have happened 0% of the time.

In summary, climate forecasters are aware of the difficulty of predicting climate in the El Niño “dead zone” between 38-42 deg N latitude. As a result, they rarely issue any forecasts besides climatology in that region. The forecasts have fared poorly during those times CPC have issued non-climatology forecasts in this “dead zone”. Robert Livezey (personal communication, November 2002) has unpublished analysis showing that during El Niño and La Niña periods, the CPC forecasts have displayed skill. During non-Niño conditions, CPC’s skill is indistinguishable from that of a climatology forecast.

6.3 Water supply forecasts

The following chapters draw from Pagano et al. (2004b), the first comprehensive analysis of official water supply accuracy in the history of the NRCS. This chapter is also the first synthesis of the many NRCS “gray literature” publications concerning water supply forecast evaluation.

6.3.1 History of previous evaluation studies

Although the water supply forecasting community has long recognized the importance of forecast evaluation, it has also long struggled to find appropriate forecast evaluation measures. The challenge lies in normalizing the forecast errors in some fashion so as to allow fair comparison between large rivers and minor creeks. Additionally, one must find measures that are understandable and relevant to forecast users. The measures discussed and compared in this chapter are summarized in table 6.1.

The early history of the NRCS water supply forecasting program in the 1930s-1950s contains many evaluations of individual forecast locations and years (e.g., Paget 1940) similar to those put together for the informal water-year-end summary meetings for users that still occur today. Forecast bulletins as early as 1931 contain tables of the previous year’s forecasts and observations and include a text discussion of the performance of the forecasts. Very early evaluations aimed to establish the credibility of the water supply forecasting enterprise.

Name	Form	Source	Advantages	Disadvantages
Year-end summary, percent error	Tables of F and O. $\lvert F-O\rvert/\bar{O}$	Paget, 1940; Church; 1935	Direct relevance to user interested in one location. Simple.	Difficult to know abnormality of forecast error. Cannot be compared across regions
Forecast Error	$\lvert F-O\rvert/O$	Work, 1940; Work and Beaumont, 1958; Shafer and Huddleston 1984	Same as percent error. More sensitive to errors when observation is low, the focus of agriculture.	Same as percent error. If O = 0, error is infinite.
Graphical evaluation analysis	$\lvert F-O\rvert$ normalized by the range of observations or long term flow, plotted as an exceedence. Area under curve is a "single number" measure	Kohler, 1959	Allows comparison of different techniques. Is a measure of skill. Provides detail about distribution of errors	More complicated. It is not necessarily clear what is the best way to compare forecasts across locations.
Skill Coefficient	$\mathrm{Sum}(\lvert\bar{O} - O\rvert)/\mathrm{Sum}(\lvert F - O\rvert)$	Shafer and Huddleston 1984	Allows comparison of different locations and techniques. Is a measure of skill.	More complicated. Positive skill is unbounded, i.e. skill can equal positive infinity if F – O = 0
Nash Sutcliffe (NS), Coeff. of Pred.	$1 - (\mathrm{Sum}((F-O)^2)/\mathrm{Sum}((\bar{O} - O)^2)$	Nash and Sutcliffe, 1970; Lettenmaier, 1984	Allows comparison of different locations and techniques. Is a measure of skill.	More complicated. Is more sensitive to extreme forecast errors.
Error Variance	1 - NS	Schaake and Peck, 1985	Same as NS	Same as NS

Table 6.1 Comparison of forecast evaluation techniques and measures presented herein. F indicates forecast, O indicates observed, Ō indicates the long-term mean of the observations.

In the earliest such evaluation, Church (1935) computed the absolute difference between forecast and observation as a percent of long-term average runoff for six basins in Nevada and California. The exceptional result that two thirds of the forecasts had an error of less than 10% should be tempered by the fact that these forecasts were issued on 15 May, the midpoint of the spring melt period. Additionally, the standard deviation of the observations in this region is typically one third of the average, indicating that a "no-skill" forecast every year equal to the long-term average might produce errors of less than 10% one fourth of the time (Appendix A).

In 1944, the NWS and the NRCS began publishing forecasts independently for many of the same locations. Pressure was put on the agencies to coordinate their forecasting programs to prevent the duplication of effort and to head off the problems natural resource managers would face when confronted with conflicting forecasts (e.g., Medford Mail Tribune 1959). The agencies could not agree on the best forecasting method, with the NRCS favoring the use of snow survey data, and the NWS favoring low-elevation accumulated precipitation data. While efforts to institutionalize coordination failed in 1956, regional pockets of coordination continued informally. Most forecast evaluations between 1945-1960 were motivated by a desire to show the superiority of the forecasts of one agency over another.

Work and Beaumont (1958) performed a westwide evaluation of NRCS and NWS forecasts (1944-1952), including three tables of analysis arranged by state, basin, and year. Forecast error was defined as the forecast flow divided by the actual flow, expressed as an absolute difference from 100% (after Work 1940). The authors also averaged together the historical forecasts by the various agencies to determine which agency would have benefited by coordination. Finally, the authors presented time series of which agency had the majority of "best" forecasts by year and a map of which agency performed best overall at each location. In aggregate, NRCS forecasts were "better" than NWS forecasts 11 out of 13 years of the evaluation. The NRCS forecasts were "better" at 55% of the locations, although there was no obvious spatial pattern to the performance of the agencies. The authors concluded that snow survey data produce superior streamflow forecasts, as snow is the source of most of the water (see also chapters 2.2-2.3).

Kohler (1959), chief research hydrologist of the NWS, rebutted this study using a different, graphical evaluation technique that was in use by Soviet hydrologists at the time (figure 6.1). First, the absolute difference between each forecast and observation was expressed as a percentage of the range of the observations. These differences were ranked and plotted as a probability of non-exceedence. A second curve was displayed based on the differences between each observation and the long-term mean, again as a percentage of the range of the observations. If desired, several curves based on the forecasts from different agencies or lead-times could have been overlain and the performance compared (as was done in CBIAC 1961, 1964). The area under each error curve was a measure of forecast performance (small area being good).

Although not necessarily intuitive to a user, this technique had many appealing aspects in that one could visualize the distribution of errors as well as compare the skill of the forecasts relative to a baseline (such as climatology or always guessing average). While the formulation was not exactly the same, the area between the line for the performance of the forecasts and the line for the performance of the long-term mean approached the spirit of the Nash-Sutcliffe Coefficient of Efficiency (NS, Nash and Sutcliffe 1970). Using this new forecast evaluation measure, Kohler concluded that the NWS's early season forecasts were far superior to those issued by the NRCS, yet conceded that the differences were slight later in the season.

A lull in forecast evaluation activities followed until the 1980s. Internal NRCS records indicate a significant update of forecast archives and evaluation tables in 1968, although there is no documentation, beyond standard bookkeeping, of research on forecast errors supporting that update. Forecast evaluations at individual locations occurred in the research literature, oftentimes to compare the historical forecasts against new techniques being developed. It was not until the work of Shafer and Huddleston (1984) that a westwide look at water supply forecast evaluation was revisited. Shafer and Huddleston analyzed a database of close to 50 000 seasonal streamflow forecast errors, representing the complete history of NRCS forecasts except those from Alaska.

It is difficult to identify the exact motivation of Shafer and Huddleston's work. One objective is to measure changes in accuracy associated with the adoption of new technologies (e.g., the adoption of computers in the mid-1960s) and institutional changes (e.g., the collection of state forecasting responsibilities into a national center). Also, the introduction of electronic databases to the NRCS enabled the digitization of historical forecasts and a phasing out of the hard-copy tables that NRCS personnel had been maintaining for decades; when the digitization was complete, an evaluation seemed in order (Jon Lea and Ken Jones, Natural Resources Conservation Service, personal communication, April 2003). In the early 1980s, the NRCS and NWS agreed to coordinate forecasts at locations where both agencies have forecast procedures, although it is unlikely that this inspired the forecast evaluation study. The large forecast error incident that occurred in 1983 at Lake Powell as a result of a very unusual late spring snow storm (Rhodes et al. 1984, see also chapter 5.2) did not motivate the evaluation, although it may have caused greater interest in and wider distribution of the results.

Following Church's definition, as opposed to Beaumont and Work's, Shafer and Huddleston calculated the average forecast error for 345 forecast locations and aggregated the results by state and lead-time. As expected, the forecast error decreased as the lead-time decreased. They also found an exceptional relationship (R^2 = 0.966) between statewide average forecast error and the statewide mean coefficient of variation (the ratio of the standard deviation of the observed flow to the mean), as had Lettenmaier and Garen (1979) in their analysis of streamflow hindcasts several years earlier. In other words, it was easy to incur a 100% forecast error on, for example, the San Francisco River, Arizona, where observations varied between 17% of average to over 750% of average. It was more difficult to do so on a river such as the Stehekin River, Washington, where the streamflow ranged only between 60% and 150% of average.

Shafer and Huddleston also employed a unique "Skill Coefficient" score, the sum of the absolute differences between the long-term average and the observation in each year, divided by the sum of the absolute errors between the forecasts and observations. A score of 1.0 indicated no skill, and a score of 2.0 indicated that the forecasts were twice as skillful as climatology. Like Kohler's work, this score was attractive because it enables a normalized comparison across states. The analysis revealed that 1 April forecasts for the period of record until 1980 were most skillful in Arizona and Washington, fair in Nevada, Idaho, and Wyoming, poor in Colorado, New Mexico, and Utah, and least skillful in Oregon and Montana.

Shafer and Huddleston qualitatively attempted to detect a westwide long-term change in skill, but none was apparent. Individual sites were becoming more skillful, others less skillful. The authors stated that the observations displayed a trend towards increasing variability, and when one subtracted out this effect (based on the analysis in the paragraph before last), average forecast error decreased “virtually” by 2.2% in 1966-1980 compared to 1951-1965 but decreased “actually” only 0.2%. In other words, had the streamflow variability not increased recently, the forecast error would have decreased by 2% more than it actually did. Instead, the forecasters were challenged with a more variable (hence tougher to forecast) sequence of flows than what occurred in earlier years and forecast skill suffered. The authors inferred “a 10 percent relative improvement in forecasting skill in recent years compared to a long-term average”, although the source of that value is not evident in their analyses

In the context of discussions of changes in forecast skill, Lettenmaier (1984) deflates hopes of measuring the expected 6% increase in forecast accuracy (and multimillion dollar annual benefit to the economy) associated with satellite snow cover information. Lettenmaier shows that it would take more than half a century to accumulate enough forecasts to detect such a small accuracy trend with confidence. While

Lettenmaier evaluates synthetic forecasts and not historical forecasts, he uses the NS score, calling it the Coefficient of Prediction.

Schaake and Peck (1985) used a similar score, called the error variance (1 – NS), in an analysis of forecasts during 1947-1984 for the inflow to Lake Powell, on the Colorado River in Utah. In an attempt to determine the most lucrative avenue for improving streamflow forecasts, the authors decomposed the errors into climate, data, and model based error (Lettenmaier and Garen [1979] explored this issue as well). Climate based errors could be addressed by having accurate seasonal forecasts of precipitation and temperature. Data based errors are rooted in the density of the data monitoring network, location of sites, and the quality of the data. Improving forecast tools and techniques could reduce model-based errors. Schaake and Peck concluded that almost 80% of the 1 January forecast error was due to unknown future climate; by 1 April, future climate still accounted for more than 50% of the forecast error. Model and data errors were approximately equal and were steady throughout the season.

While water supply forecast evaluation ceased after the mid-1980s, the climate and weather forecasting communities reached new heights of complexity in forecast evaluation. Long-lead climate forecasts were originally issued categorically, e.g., “Above normal”, and the community has no less than 19 categorical evaluation measures at their disposal, with the Heidke Skill score (Heidke 1926) as the most popular in operational circles. With the transition to probabilistic climate forecasts in the 1980s, probabilistic forecast evaluation scores such as the Brier Score (Brier 1950) and Ranked Probability Score (Epstein 1969) gained popularity. The most sophisticated evaluations involve distribution-oriented approaches (reliability and discrimination diagrams, Wilks 1995) and measures of the value of the forecasts to a theoretical optimal decision-maker. Regrettably, the robustness and scientific rigor of forecast evaluation techniques are inversely proportional to their accessibility and understandability by the lay forecast user. The current challenge is in linking the forecast evaluation to the user in a meaningful and relevant manner (Hartmann et al. 2002a).

6.3.2 Selected evaluation methodology

Although past NRCS forecast evaluations focus on the average percent error, and this measure is the most easily understandable by users, this study did not use it. It primarily measures the local variability of the observations and not the value added by the forecaster. Instead, the forecasts were judged by the NS score:

$$NS = 1 - \frac{\sum_{i=1}^{N} (f_i - o_i)^2}{\sum_{i=1}^{N} (\bar{o} - o_i)^2}$$

where f_i and o_i are the forecast and observations in year i for a collection of N years, and $\bar{o}$ is the mean of the observations of N years. An NS of 1 is perfect, 0 indicates no skill over always guessing average, and values less than zero indicate negative skill. In essence this score is one minus the mean squared error of the forecasts divided by the variance of the observations. It is important to note that this skill score already accounts for changes in forecast skill associated with changes in variability of observations (this issue will become important in chapter 8, especially chapters 8.4-8.6). During periods of high variability, the potential for greater forecast error is offset by increased error in the "no skill" baseline forecast of guessing average.

Although not necessary, it is useful to avoid situations of heteroskedastic error, such as where the forecast error is typically greater during high flows than low flows. Among those locations where the seasonal flow volumes have skewness greater than 1.0, the natural logarithm is applied to the forecast and observed seasonal totals before analysis (see Table 2.4). Otherwise, individual large floods would dominate the analysis, resulting in an evaluation reflecting the behavior of the streamflow in a few years rather than the quality of the forecaster on the whole. For example, the 1 January 1993 squared forecast error for the San Francisco River, Arizona was almost 110 times the median squared forecast error over the period of record even though it was, by far, the wettest forecast ever issued for this location. Instead, analysis of log-transformed flows provides more information about the performance of forecasts across a range of streamflow conditions. In this study, the transformation increased the skill scores (e.g. San Francisco January NS was 0.64 if the data were transformed and 0.54 if it was not). This transformation shifted the forecast evaluation emphasis to drought, which is the primary concern of NRCS agricultural customers in semi-arid regions (as opposed to NWS customers concerned with the protection of lives and property from floods). Further, droughts (FEMA 1995) cause approximately 3-4 times the annual economic damages of floods (Myers 1997) and therefore any impacts-oriented forecast evaluation may prefer to give more emphasis to performance during dry years. Although this study did not attempt to link forecast accuracy to user benefits, the log transformation of skewed flows increased the user-relevance of the evaluation.

6.3.3 Historical water supply forecast evaluations

The evaluation of the historical forecasts as a function of location and lead-time is presented below. A limited number of factors that shape forecast performance in a general sense are identified and examined, with examples from individual years. Forecast accuracy is explored further in chapters 7 and 8.

Figures 6.2 contains maps of the NS score for the forecasts issued in the most recent 20 years, 1983-2002. The top and middle maps indicate the performance of the forecasts issued 1 January and 1 April, respectively. The size of the circle reflects the skill of the forecasts, large being preferable over small. The outermost circle is a reference to perfect skill (NS=1), and an empty circle indicates no skill (NS=0). The hollow circle over the Sandy, Oregon in the top panel indicates that 1 January forecasts had slightly negative skill. The bottom panel reflects the change in forecast performance (NS) between 1 January and 1 April. Large inner circles indicate great forecast improvement, and the hollow circle in Arizona shows a decline in skill for the Verde, Arizona 1 April forecasts compared to those issued 1 January. The outer circle is a reference for a change in NS equal to 1.0.

During 1983-2002, the most skillful 1 April forecasts were issued for the Salt, AZ, West Walker, CA, and Little Colorado, AZ, whereas the least skillful forecasts were for the Umatilla, OR, White, CO, and Sandy, OR. The most improvement in skill between January and April occurred for the West Walker, CA, Carson, NV, and Martin Ck, NV, and the least improvement occurred for the Verde, AZ, Animas, CO, and White, CO. The westwide average NSs in January through April from 1983-2002 were +0.36, +0.53, +0.59, and +0.65, respectively.

Figure 6.2 shows the westwide average forecast skill versus issue month for 1983-2002. Skill was lowest but generally positive in January and steadily improved throughout the season. This result is intuitive in that in January the character of the seasonal precipitation has yet to reveal itself. For many locations in the Western US, snowpack is at its peak on or around 1 April, and there are fewer opportunities for dramatic changes in the amount of available water in the basin.

This is not to say that significant changes cannot occur after 1 April. A notable example of this occurred in the Colorado River basin in 1983 (as mentioned in chapter 5.2). Until April 1983, snowpack was near average, and the median forecasted inflow to Lake Powell was similarly near average (109%). An exceptionally cold and wet spring ensued, followed by a rapid warming. The observed April-July flow, at over 210% of average, overwhelmed the already full reservoir system. Perhaps a less well-known example, the Animas basin had extremely low snowpack on 1 April 1999 after an exceptionally warm March, and the median streamflow forecast was approximately 40% of average. Near-record rain fell in April-May, and the snowmelt pulse volume was near to above average. A monsoon of unprecedented strength, however, produced summer floods. April-September streamflow totals were close to 140% of average. Although not one of this study's basins, Ponil Ck nr Cimmarron, NM (USGS id 07207500) to the southeast had a forecast of 20% of average in the same year and the eventual March-June flow was over 370% of average. The five largest 1 April errors in the database, defined as the squared difference between the forecast and observed, divided by the long-term observed variance, were (beginning with the largest first): Animas, CO, 1999; Verde, AZ, 1988; Sandy, OR, 1981; Bruneau, ID, 1963; East, CO, 1957. These were all due to low snowpack conditions followed by exceptional storms.

While these examples represented large underforecasts, causes for overforecasts tended to be more complex. The two largest forecast overestimates (Yellowstone, MT, 1949; Weber, UT, 1936) were due to moderate snowpack being followed by hot, dry, windy weather (i.e. intense sublimation). The Lamoille, NV, 1943 error, however, is more difficult to interpret. In this case, forecasters overestimated the effects of a high water table, and this very small high-elevation basin seemed to be in a dry microclimate surrounded by heavy snows. The streamflow data were also questionable because a series of floods had recently occurred, possibly affecting the instruments and rating table (Church and Boardman 1944).

Forecast error in any given year is strongly related to the character of precipitation that falls subsequent to the forecast issue date. Similarly, one might expect the average increase in forecast skill between January and April to be proportional to the percentage of precipitation that typically falls in January, February, and March. The following analysis determined if the operational forecasts displayed this characteristic.

The streamgage coordinates and a 1-km digital elevation model (HYDRO1k; http://edcdaac.usgs.gov/gtopo30/hydro/namerica.html) were used within a geographic information system to delineate the 29 basins of this study. The long-term 1961-1990 climatological average PRISM precipitation for each month was calculated within each basin's boundaries. The long-term normal January-March precipitation was then divided by the average seasonal total precipitation, beginning in January. For example, if the water supply forecast target season was April-September, "seasonal precipitation" meant January-September (or January-July if the target season is April-July). A high value indicated that a large portion of the streamflow-relevant precipitation typically fell in January-March, and the spring was a climatologically dry period. A moderate value (~40%) indicated a relatively flat seasonal cycle to precipitation, and a very low value indicated that most precipitation tended to fall in the spring and summer. Arizona rivers were withheld from this analysis, as their target period shrank as the season progressed, and a smaller target was not necessarily easier to hit. This chapter investigates the relationship between the typical improvement of forecast skill versus leadtime and the typical amount of precipitation falling in winter. Section 8.7 will relate forecast error in individual years with unusual springtime events.

The strength of the correlation between average forecast improvement and the climatological cycle of precipitation in Figure 6.3, as expected, was relatively strong (R = 0.64). The river with the greatest improvement in skill between January and April was the

West Walker, CA, because 46% of the annual precipitation fell in January-March, compared to only 17% that fell in April-July. In comparison, the Tongue, WY showed little improvement in skill by April, in part because April-June in the Plains states was usually the wettest time of year. The Animas, CO showed the least improvement of any basin outside Arizona due to the aforementioned 1999 event and because the basin was under the influence of the summer monsoon; the recent operational switch of forecast target periods from April-September to April-July may address some of this problem. The outlier (filled triangle) in the upper left corner of the diagram was the Bruneau, ID; as the target season began in March, observed March flow for this location was known in real-time. On 1 April, the forecaster was given an artificial advantage because part of the target season was in the past and therefore known with complete confidence. This explanation may not entirely suffice because the Pecos River near Pecos, NM similarly has the advantage of having part of its verification period in the past, yet does not have the same inflation in skill. Perhaps this advantage may be offset by the skill degradation associated with the Pecos, like the Animas, being affected by the summer monsoon.

The other exceptions to this rule, in the lower right corner of the diagram as hollow circles, were the four rivers in Oregon that did not improve as much as expected versus lead-time. Most snowmelt-dominated basins around the Western US have a strong seasonality in streamflow, with low baseflow from September to March, a rise in late spring, a peak in summer, and recession in the fall. For the East River, CO, only 5% of the January-September streamflow typically occurred in January-March. In contrast, Oregon basins experienced a mix of rain and snow and have displayed “peaky” hydrograph behavior (i.e. transient short-duration rainfall-runoff events) during the winter. On average, 47% of the January-September flow on the Sandy River typically occurred in January-March (43%, 42%, and 37% for the Umatilla, Malheur, and Rogue, respectively). It is possible for a large snowpack in February to be wasted away by March rains and run off before the April-September forecast target period begins. For example, on 1 February 1996, the Sandy watershed had near-average snowpack and near-average streamflow forecasts. Warm temperatures and heavy rains caused major flooding and resulted in February's streamflow being 260% of average. By 1 April, the snowpack was 50%-60% of average, and the eventual April-September flow was among the driest third of record. The special challenge of seasonal streamflow forecasting in Oregon is evident.

Given this information, a measure of "expected forecast skill" could be derived from climatological parameters and compared with the jackknife calibration error of the forecasting equation to detect possible over-fitting. Spring precipitation in the context of the seasonal cycle would be the primary driver of forecast improvement versus leadtime. Secondary information may include an index of winter runoff relative to annual runoff; this ratio may be related to temperature. Future research may indicate what other factors are relevant to expected forecast skill.

6.4 Summary

As evaluated by others, climate forecasts are least skillful in the Upper Colorado River Basin and Great Basin, during summer, and for years not affected by ENSO events.

They perform best in the Southwest US and Pacific Northwest, during the cold seasons and for El Niño or La Niña years. Temperature forecasts perform better than precipitation forecasts.

The water supply forecast evaluations presented herein reconfirm past analyses and the intuitive notion that forecast skill generally increases as lead-time decreases. The increase in skill between January and April is directly related to the proportion of seasonal precipitation that typically falls in January-March. Therefore, regions such as California, with relatively compressed precipitation seasons, see dramatic increases in forecast skill between January and April. The exceptions to this rule are mixed snow-rain basins in the Pacific Northwest. Even if January-March accounts for a large percentage of the seasonal precipitation, where winter streamflow comprises a significant portion of the annual streamflow, 1 April forecast skill could be low.

FIG 3 COMPARISON OF APRIL 1st WATER - SUPPLY FORECASTS PUBLISHED BY THE WEATHER BUREAU AND THE SOIL CONSERVATION SERVICE (1947-57)

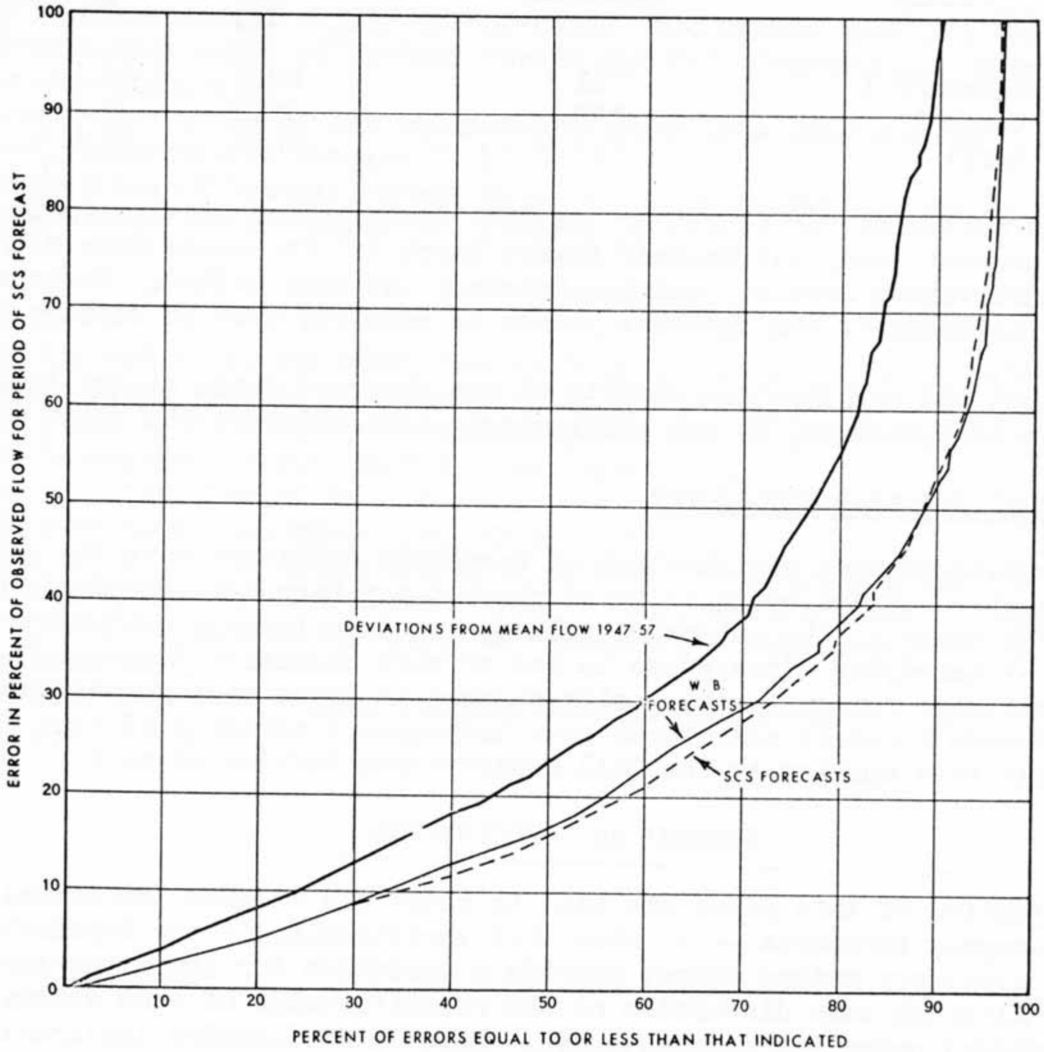


Figure 6.1 Graphical forecast evaluation technique used by Kohler (1959). The forecast error (absolute deviation of forecast from observed) as percent of the long term average flow is presented as exceedence curve. Shown are the SCS (NRCS) and WB (NWS) forecasts issued 1 April for all locations from 1947-1957. The long term normal (naïve baseline) curve is also shown. That the agency forecast curves are below the naïve baseline curve indicates they are skillful. SCS also displays a tendacy for having smaller low-magnitude errors as well as having larger high magnitude errors. The WB, in contrast, has a less variable error distribution.

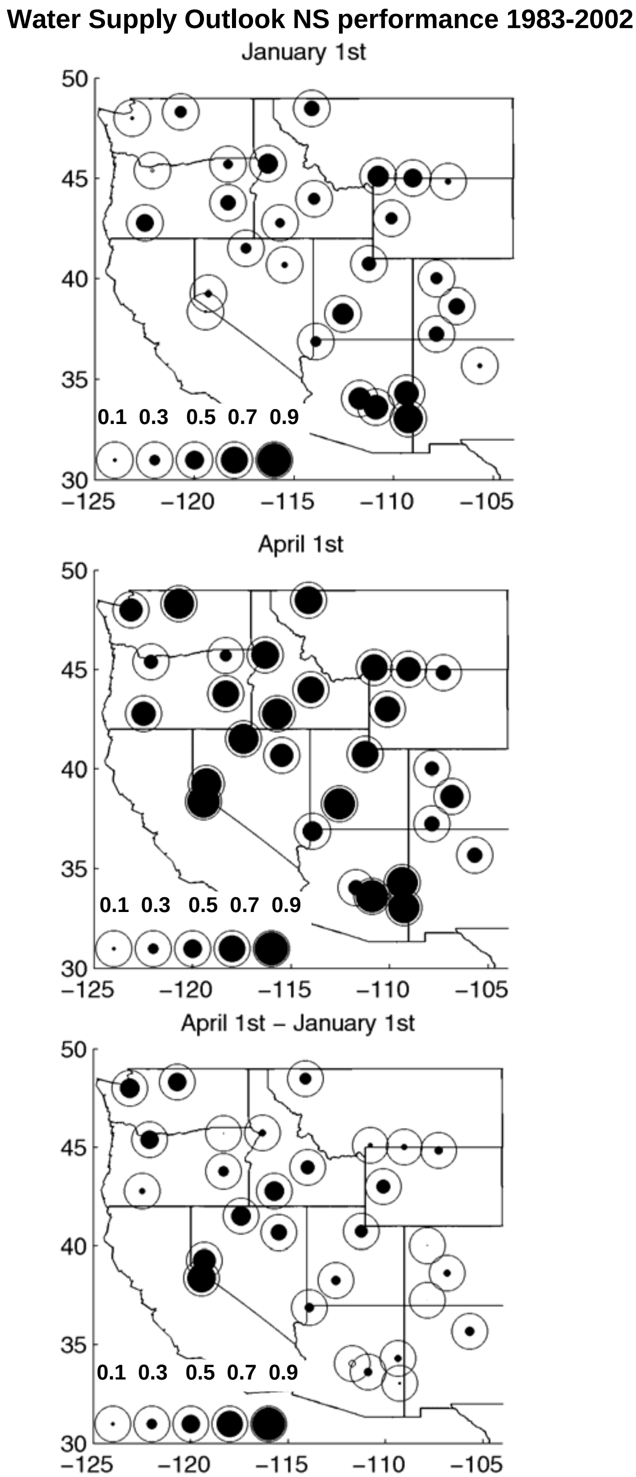


Figure 6.2 Skill of 1983-2002 water supply forecasts issued 1 January (top) and 1 April (center). Forecast improvement between January and April is shown in the bottom panel. Large filled circles indicate high skill or great improvement.

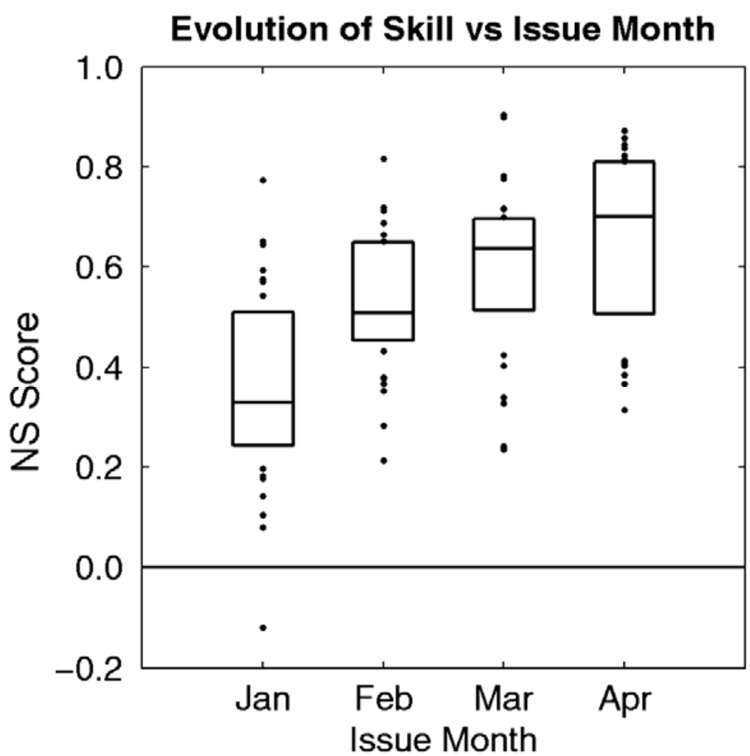


Figure 6.3. Box diagram of forecast skill versus issue month for the 29 basins during 1983-2002. Skill is measured by the NS score, with NS=1.0 indicating perfect skill and NS<0 indicating negative skill. The box has lines at the lower quartile, median and upper quartile values. Individual sites outside this range are shown as dots.

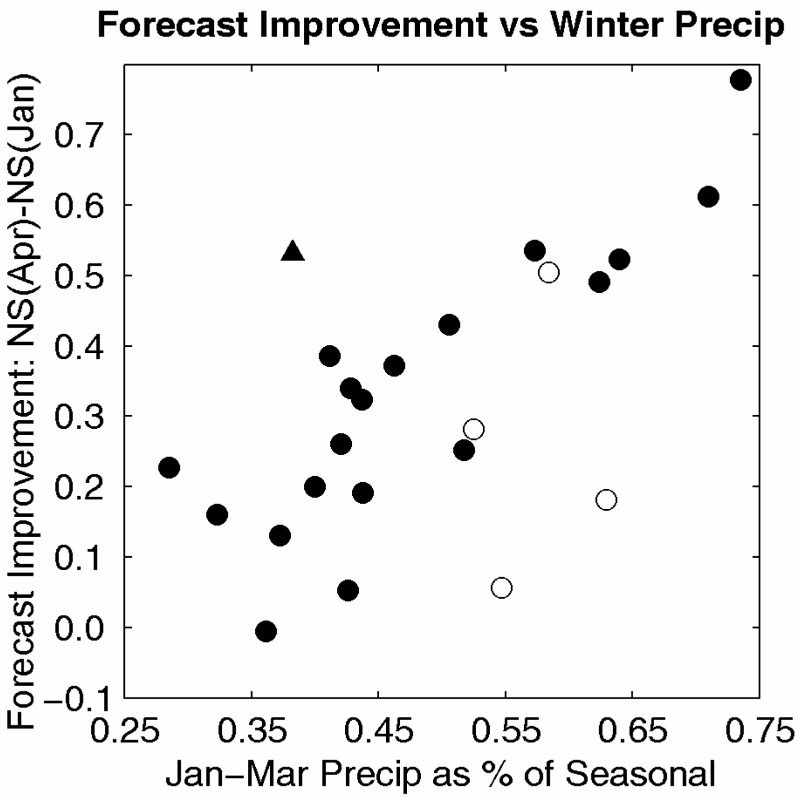


Figure 6.4 1983-2002 Forecast skill improvement between January and April, as a function of the climatological average January-March precipitation, relative to the full seasonal (January through the end of the forecast target period) precipitation. Large values on the X-axis indicate most precipitation typically falls in the winter and small values indicate a relatively flat seasonal cycle to precipitation. Hollow dots are basins within Oregon and the triangle is Bruneau, ID (see text for discussion).

7. RELATIVE MERIT OF USING CLIMATE INFORMATION IN WATER SUPPLY FORECASTS

7.1 Introduction

Previous chapters have reviewed past studies of climate and Western US streamflow. A sound scientific basis exists for the linkage between climate and water supply forecasts. The question remains, what are the specific quantitative benefits of using climate information in NRCS hydrologic outlooks? The relative benefit of climate information can be assessed by conducting sensitivity tests on an objective water supply forecasting system that mimics the behavior of the official forecasts. The change in skill of the "synthetic hindcasts" that include climate information, versus those that do not, is a rough indication of how operational forecasts might improve with the broad adoption of climate information. This system can also be used to explore the potential skill of forecasts with longer leadtimes than those currently issued. The following chapters develop, evaluate and apply such a synthetic hindcasting system. This system will also be used in chapter 8 to detect and diagnose long-term variability in water supply forecast skill.

Sections 7.2-7.6 outline the construction of a synthetic hindcasting system for the 29 forecast points described in chapter 2.7, and chapter 7.7 discusses its strengths and weaknesses compared to the official forecasts evaluated in chapter 6.3. Section 7.8 describes the results of the sensitivity tests including versus excluding climate information. As a matter of terminology, the words simulated or synthetic "hindcasts" and "forecasts" are used interchangeably to describe the objective guidance provided by the system described herein.

Figure 7.1 is provided as a visual guide to the hindcast development process.

7.2 Synthetic hindcast development

Each basin is delineated in a Geographic Information System (GIS) framework using the USGS station location and the USGS "Hydro1k" 1-km resolution digital elevation model. To be eligible for analysis, a meteorological data site must be within a basin or within 35 km of its boundary (figure 7.2). The forecaster desires a mix of high quality sites with long-term records that capture the behavior of the snow at a variety of elevations. At the same time, he tries to avoid far-away sites that in an individual year may be affected by a micro-scale climate feature and not reflect actual snow conditions within the basin. Although no firm guidance exists on how far a site can be from a basin and still be considered in operational forecasting, 35 km is a fair approximation for what is done in practice. Of the over 180,000 stations in the COOP network, 387 have 20 or more years of data and are within 35 km of the study basins. Of the nearly 3000 snowcourse and SNOTEL measurement sites in the Western US, 674 sites are with 35 km of the study basins and have more than 20 years of data.

Separate streamflow hindcast configurations were developed using monthly snow data and accumulated precipitation data. The first system most resembles the NRCS forecast of the 1950s whereas the second resembles a system in use at the NWS during the same period (Work and Beaumont 1958). A third system that used persisted (i.e., September-November) streamflow as a proxy for soil moisture conditions had such poor results that they are not shown here. Although persisted streamflow correlations are very low ($r \sim 0.3$-0.5), they behave in such a way that they can lead to highly negative skill forecasts. For example, in almost all years, fall streamflow data **can** reflect baseflow conditions, which in turn reflect long term soil moisture deficits, but they can also reflect anomalous rare transient fall storms (e.g. a hurricane) that may or may not contribute to the following season's water supply. In this case, a regression equation would overreact to the fall streamflow that was an order of magnitude or larger than any other year on record, yielding highly negative forecast skill. A fourth streamflow hindcast system using climate information is described in chapter 7.5. Lastly, hybrid snow-precipitation and climate hindcasts are developed in chapter 7.6.

This experiment attempts to reproduce the characteristics and constraints of the existing operational environment. As is routine at the NRCS, the hindcast equations are developed using a "jackknife" technique involving the calibration of an equation on all but one historical year of data and then using the equation to "hindcast" the single year that was removed. This process is repeated leaving out each historical year in turn until a full set of hindcasts is obtained. A more realistic approach involves calibrating equations using data entirely in the past, namely that someone developing a forecast in 1957 would only have data available before 1957 and would not know about the behavior of the basin in, say, 1960 or 1983. Mason and Mimmack (2002) describe and use this "strict" jackknife calibration approach; Unger (1996a,b) uses a slightly relaxed approach termed "bi-directional retroactive real-time validation" which involves the normal strict approach as well as a strict approach with the flow of time reversed.

The uni-directional strict system was tested, but the simulated forecasts early in the period of record had a low correlation with the official forecasts in part because of the current inability to access data from "legacy" snow courses (mentioned in chapter 2.2). The hindcasts produced in the early part of the record were very unstable (i.e. based on such small calibration sample sizes that the forecast relationships changed dramatically with the addition of each year). Visual inspection of the observed and forecast time series revealed that the period of record skill (or lack thereof) was being dominated by these early forecasts. Given that the signal of climate in streamflow is subtle, the behavior of these few unstable forecasts may have dominated the sensitivity analysis, obscuring otherwise useful results. Although a very limited number of operational forecasts points suffer from high error due to small calibration sample sizes, this situation is not representative of the broader spectrum of forecast situations. Therefore, the strict jackknife approach was abandoned in favor of the weak jackknife approach described above.

Finally, Garen's (1992) principal components based-regression technique finds wide operational use, in part because it is very effective at dealing with the extremely high levels of inter-correlation among input variables. For example, in the upper Rio Grande headwaters, the correlation between 1 March SWE measurements for the Upper San Juan (06M03S) and Wolf Creek Summit sites (06M17S) is 0.95. Under normal multiple linear regression, it is likely that highly correlated variables would compensate for one another. For example, it is possible that although both Upper San Juan and Wolf Creek Summit are strongly positively correlated with streamflow, one of these variables may receive a negative coefficient under multiple linear regression. In effect, small (likely random) differences between the two time series are such that giving one time series a negative coefficient may yield a marginally better calibration fit to the historical data. Having a correlation coefficient and regression coefficient of opposite signs is generally frowned upon because of the unpredictable outcomes that may result when one or another site is missing data (NRCS 2004). The principal components technique ensures that all sites for a variable within a basin that are highly inter-correlated have regression coefficients of the same sign (which are almost never negative). The technique adopted here reproduces this key feature of the operational technique, although it does not use the standard principal components approach primarily because of the restrictions it puts on the available period of record (discussed in the next section).

7.3 Snow hindcast system

This system uses the first of the month snow measurement as an indicator of future streamflows. The first necessary step is to aggregate individual snow measurements into a single basin-wide index. In the operational forecasting environment, the NRCS hydrologist would use a principal components analysis as outlined by Garen (1992). The operational technique involves a search algorithm that does an exhaustive search of site and variable combinations. This technique has many benefits, such as its ability to deal with intercorrelation among predictor variables and its ability to find the optimal combination of predictors. However, it also has its drawbacks. Most notably, the period of record of the calibration data under the standard technique is the intersection of sites (i.e. all predictors must be serially complete). If one site of many is missing in a particular year, all sites for that year must be excluded from the analysis, despite the very strong correlation of that predictor with other predictors. For example, if snow site A has data from 1984-2003 and sites B-F have data from 1934-2003, the NRCS hydrologist's calibration period must be 1984-2003 (20 years), as if no information were available about 1934-1983 (50 years). In this example, less than one-third of the available data are being used. In practice, the hydrologist would probably discard the site with the short period of record.

Operationally, the hydrologist must balance making sure that all processes that affect the basin are represented, having complete geographic coverage of sites within the basin, having a long period of record for calibration and having an interesting mix of years, including years with large unavoidable forecast errors (such as extreme spring precipitation events). This last issue is an acute vulnerability of the operational search algorithm because it favors combinations of variables with missing data in extreme years because the summary skill score would be lower if those extreme years were included.

This study requires a modified version of the operational approach to satisfy its objectives. To investigate decadal variability, a very long period of record of forecasts is necessary. Therefore, the approach must use the union of the available data, rather than the operational approach involving the intersection of data. This approach must also take a collection of stations, emphasize those whose data are good predictors of streamflow and produce a representative basin-wide time series of that variable. Finally, this approach must be almost entirely objective, based on a limited set of fixed rules, so that close to 100 000 retrospective forecasts can be generated and analyzed with minimal human involvement. In comparison, the operational environment is rich with subjectivity incorporating a broad range of “gray” information into the forecasts.

In order to overcome the operational requirement of the union of datasets and to counter the operational search algorithm’s tendency to exclude sites with data during extreme years, this study develops a modified site weighting scheme as well as a primitive principal components analysis to develop forecast equations.

Aside from the data from the year being hindcast, if at least 20 years of coincident streamflow and 1 April snow measurements are available, the period of record correlation coefficient between each snow site and streamflow is calculated. If the correlation is greater than 0.1, the site is considered eligible for analysis. Although a snow site may be located near a basin, it may not capture the streamflow-relevant snow conditions within the basin, or it may suffer from data quality issues. During forecast equation development, a hydrologist would exclude such sites. Additionally a hydrologist would eliminate sites with negative correlation with streamflow; it is difficult to imagine a scenario where wetter basin conditions would reliably indicate less streamflow.

The Z-scores of the remaining sites are computed relative to their individual period of record:

$$Z_{sijk} = (S_{ijk} - \text{mean}(S_{ik}))/\text{std}(S_{ik})$$

where S represents the snow measurement (SWE in inches) and subscripts "i", "j", and "k" are the measurement month, year and location, respectively. The Z-score transformation is done to standardize the values so the relative variations in station conditions are considered equally in regression, rather than the analysis being dominated by the stations with the greatest absolute variability. This normalization is also the first step of a principal components analysis.

Next, the Z-scores are then combined to form a single basin-wide composite snow index time series. Later, this time series will be regressed against streamflow to form the streamflow forecast equation. Simply averaging the Z-scores gives them equal weighting. In contrast, multiple regression techniques would emphasize the individual snow sites that are well correlated with streamflow and give less weight to unimportant sites. The hybrid technique adopted here develops a weighted average of the snow sites, based on their correlation with streamflow

$$SWE_{ij} = \Sigma(R_k^2 * Z_{sijk})/(\Sigma(R_k^2))$$

where "k" depends on the number of sites reporting. R_k is the correlation coefficient between 1 April snowpack and the streamflow for the target period corresponding to forecasts issued in January (i.e., April-September volume). This selection of months is likely to work well in most locations except Arizona basins where snowpack peaks and melts early and 1 April snow can but might not reflect the overall character of the runoff season (i.e. 1 April snow is being compared to January-May runoff).

The R_k weighting is used for all snow indexes developed, regardless of the leadtime of the hindcast being developed (e.g., the correlation between 1 April snowpack and seasonal streamflow is used as the weighting factor while developing a hindcast equation between 1 January snowpack and seasonal streamflow). The date of 1 April is selected because it is generally the apex of the snow accumulation season when snow has its highest correlation with streamflow. Of any month of the year, 1 April also almost always has the longest period of record of snow measurements (see chapter 2.2). Choosing one set of R_k weightings to use in all months allows the longest period of record to be used in the estimation of weights. Also, while the absolute correlation of snow and streamflow may change with leadtime, there is little hydrologic reason to believe that a site's **relative** importance compared to its peers would change throughout the season, except for ephemeral sites whose snow typically peaks well before 1 April and has all but disappeared later in the season.

At this stage, this combination technique is similar but not identical to principal components analysis. One key difference is that in principal components analysis, the weighting factor is not based on the correlation with an external variable (in this case, streamflow) but rather is based on intra-variable correlation so as to capture the primary modes of variability within the dataset. This study assumes that the primary mode of variability is related to streamflow volume and that both techniques would yield similar results. The other key difference between the techniques is that principal components analysis requires serially complete data whereas the selected technique uses however many or few sites are available and can compensate for missing data.

The disadvantage of the selected technique is that early in the period of record, when only one or two sites are available per basin (especially if those sites are ephemeral), the Z-score time series tends to exhibit more variability whereas later in the period of record when more sites are available, the composite Z-score can exhibit less variability. The selected technique also artificially benefits from always knowing the long-term mean and standard deviation of the measurement; the first few years after a site is installed, stable estimates of such moments are unavailable. If the sites have different periods of record, their Z-scores are not strictly comparable if there are strong trends in the mean or variability over time, something that might degrade the simulated forecast skill. In comparison, principal components analysis requires that all sites have identical periods of record so mixing stations with different baseline periods is not an issue.

As mentioned earlier, if only one site is available in a given year, the index is equal to the Z-score for that site. If more than one site is reporting, the index is the weighted average of the Z-scores of the sites (weighted by the correlation between data values at that site and seasonal streamflow). If an individual snow site is well correlated with streamflow, it is given relatively more importance than a site not well correlated with streamflow. Again, this technique suffers early in the period of record when only 1-2 snow sites are available; if only one low correlation site is available in a given year, this site is given full weighting for that year in the composite snow index. The overall technique is described visually in figure 7.3. This technique differs from the operational search technique in that in the selected technique, all sites are considered although some with a greater weight than others. In the operational technique, sites can be entirely excluded from the analysis if they are not in the optimal mix of sites. Again, the current technique may be causing too much spatial smoothing of the data, but given the size of the basins, this effect should be minor compared to the benefit of having a longer period of record in the calibration.

A linear regression equation then relates the composite snow index to streamflow and is used to "forecast" the censored year's flow. If the "forecast" is less than zero and the streamflow has not been log transformed (see Table 2.4), zero is used in its place. Operationally, regression equations can produce negative forecasts. In these cases, the hydrologist does additional analysis to arrive at a physically realistic forecast, such as the previous minimum flow observed on record. The replacement with zero is used here, instead, out of the desire for a parsimonious set of rules for developing forecasts. Negative forecasts most often occur when a linear fit is applied to data whose fundamental relationship is non-linear. It is reasonable to assume that log-transformation of the skewed flows mentioned earlier eliminated most of the possibilities of negative forecasts. Finally, if no snow sites are available, the hindcast is considered missing.

7.4 Precipitation hindcast system

Accumulated precipitation data at relatively low elevations can serve as an indicator of how much moisture has been accumulated in the snowpack at higher elevations, the origin of most basins' summer streamflow. Additionally, spring rainfall can contribute directly to runoff and fall rainfall can contribute to soil moisture. However, at some times of year, higher evaporation rates can reduce the net availability of precipitation for streamflow. Likewise, some fall rainfall leaves the basin as runoff before the land surface is sealed over with snow. Therefore, if precipitation data are to be used to forecast streamflow, the adopted technique must balance the relative importance of precipitation during some months over others.

In the 1950s-1970s, the NWS advocated such a forecasting technique using accumulated precipitation, as described in CBIAC (1961, 1964). This system is especially attractive for use in this study because it matches the operational practices of NWS water supply forecasters during that period. Originally, the NWS method used precipitation from August through June to derive an index used in forecasting water year runoff. Forecasts issued in January would use observed accumulated precipitation data from August-December and an estimate of future precipitation after December. Each month's precipitation was given a semi-subjective weight to recognize evaporative losses in the early fall and late spring. This study uses a modified version of the weights used in CBIAC (1961), shown in Table 7.1. The original CBIAC weights were developed for a northern basin in Montana and the weights selected below better reflect broader conditions throughout the Western US; in practice, the weightings assigned to each month would depend on the basin being studied.

Table 7.1. Monthly precipitation weighting coefficients. August-1 indicates August from the previous water year.

August-1	0.03	December	0.11	April	0.11	August	0.03
September-1	0.06	January	0.11	May	0.09	September	0.01
October	0.09	February	0.11	June	0.07		
November	0.11	March	0.11	July	0.05		

Therefore, a forecast issued in January would use a precipitation index of the form

0.03 x (August-1)

0.06 x (September-1)

0.09 x (October)

0.11 x (November)

\+ 0.11 x (December)

Net August – December Precipitation Supply Accumulation

This index is a proxy for the net precipitation supply, the amount of water available for streamflow, accounting for the lesser importance of rain in the fall and the increase of evaporation in the spring. Additionally, precipitation-elevation gradients during winter frontal storms are much steeper than during spatially heterogeneous summer convective events. Therefore, low elevation NWS stations during winter would have much less precipitation than the high-altitude snow-producing parts of the basin, whereas during summer sometimes an equal amount can fall in both places. In winter, the NWS stations would be less than the basin wide average precipitation and therefore would need higher weights.

The precipitation amount is accumulated to date for each site, and then converted into a Z-score as with the snow hindcast system. A basin-wide precipitation index is developed in the same way, using an average of sites, weighted by their correlation of net August-March precipitation supply with seasonal streamflow. As with the snow-based system, a single set of weights is used for all months because it is assumed the relative importance of sites does not change from month to month. Again, if a forecast is negative, it is reset to zero and if no data are available to make a forecast, it is considered missing.

7.5 Climate-based hindcast system

Similar to the snow and precipitation systems, the 8 climate indices were correlated with streamflow and combined into a “composite” climate index for the basin. In the previous experiment, snow and precipitation could only be positively correlated with streamflow in order to be considered. The climate indices with negative correlations with streamflow were considered, although their sign was reversed before entering into the composite index. For example, if streamflow is positively correlated with Niño3.4 ($r = 0.5$) and negatively correlated with SOI ($r = -0.5$), the composite index would be

$(0.5^2 * \text{Niño3.4} - (-0.5)^2 * \text{SOI}) / (0.5^2 + (-0.5)^2)$. The end result is that all of the climate indices are given a positive orientation before aggregation into a "basin wide" climate index. The climate R_k weightings are proportional to the correlation between seasonal streamflow and the September-November values of the climate indices. Again, the weightings are fixed and are used for every month. At the time of forecast, the previous 3-month average of the climate index is used. For example, a forecast issued in January might use Niño3.4 averaged from October-December whereas a forecast issued in April would use the average Niño3.4 value from January-March.

7.6 Hybrid hindcast systems

Since the 1980s, the NRCS and NWS are required to coordinate streamflow forecasts. Both agencies have relatively separate approaches and each has its strengths and weaknesses. Since neither agency's approach can be said to be conceptually or quantitatively better in all respects (see chapter 6.3.1), there can be benefit in combining them. However, the coordination system, in practice, is simplistic, ad hoc and subjective, rather than based on quantitative analysis and objective techniques (despite the rich research literature on methods of combining forecasts [Clemen 1989 provides a comprehensive review]). Despite the flaws of the methodology, a "unified voice" in the forecasts eliminates confusion among users about which guidance they should follow (as described in chapter 6.3.1). The current operational merging of NRCS and NWS forecasts is non-quantitative and does not follow a fixed methodology. Sometimes the forecasts are simply averaged. Other times, a forecaster may exert more influence over the other through a convincing argument or force of personality. The research literature suggests that the forecast should be an average of individual forecasts, weighted by the historical skill of each individual forecast system. When such historical skill measures are not available, the research literature cautions that individual forecaster **self-confidence** is not necessarily a good basis for weighting forecasts when combining them (especially if an individual is prone to overconfidence).

In the experiment in this study, the combined forecast is the weighted average of the forecasts from the individual snow, precipitation and/or climate-based systems. The weights are determined by a measure of historical skill, the strength of the calibration R^2 of the individual forecast equations. For example, if the calibration R^2 of the snow, precipitation and climate hindcasts are 0.8, 0.6 and 0.2, respectively, the combined hindcast is

$$Fspc = (0.8 * Fs + 0.6 * Fp + 0.2 Fc) / (0.8+0.6+0.2)$$

Where Fs is the forecast produced by the snow system, Fp is for the accumulated precipitation-based forecast, Fc is based on climate, and Fspc is the combined forecast. As with the other systems, the hybrid hindcasts are developed in jackknife mode. When hindcasts are combined and one of the hindcasts is not available in an individual year, its weighting is set to zero. For example, if snow and precipitation hindcasts are being combined in 1953, and no snow hindcast is available for that year, the combined hindcast is identical to the precipitation hindcast.

In later sections, hindcasts developed by joining the hindcasts of two or more of the snow, precipitation and climate systems are referred to interchangeably as "hybrid" or "combined" hincasts. Also, the terminology "snow+precipitation" hybrid hindcast (e.g., in the next section) refers to the joining of a snow hindcast (chapter 7.3) and accumulated precipitation hindcast (chapter 7.4).

7.7 Performance of hindcast systems

Figure 7.4 shows the skill of the snow (top left) and precipitation hindcast system (top right) versus lead time. The lower left panel shows the performance of the combined snow+precipitation system and the lower right shows the performance of the official forecasts. As no snow measurements are available in November and December, the forecasts from the combined snow+precipitation system during those months are identical to those based on accumulated precipitation (see chapter 7.6). These jackknife hindcasts are evaluated over the period of record using the NS, as described in chapter 6.3.1. Each box summarizes the performance of the 29 study basins, as was done in figure 6.3. As is expected, the skill of the forecasts improves as the lead time decreases. The skill in January to April of the combined snow-precipitation system is competitive with the official forecasts. While the synthetic hindcasts show that skillful non-climate based forecasts are possible in December, the skill in November is almost indistinguishable from zero.

The skill of the forecasts for the individual snow and precipitation systems is comparable although the spatial patterns of skill slightly differ (figures 7.5-7.6). These figures follow the same convention as figure 6.2. Snow is generally a better predictor of streamflow for most locations, especially in the interior Western US and Cascades. Accumulated precipitation seems to fare better in eastern Arizona and southern locations (consistent with Lettenmaier and Garen 1979).

The overall resemblance of skill for the official forecasts and those from the synthetic snow-precipitation system is very good. In one sense this is somewhat surprising given that this system does not use any information about soil moisture/long term drought, the approach is objective and uniform for a broad range of very different climates and landscapes, does not incorporate any “gray” information and is not adjusted using human expertise. In another sense however, this system uses all of the snow and precipitation data that the official forecasters would have used, and uses a regression technique that generally resembles the operational technique.

As with the official forecasts, Columbia basin and Montana synthetic forecasts, as well as Arizona hindcasts are relatively skillful in January, with little skill in most other regions. In April, both the official and synthetic forecasts perform very well in California, Arizona, good in the Columbia and Colorado basins, and poor on the Tongue WY, Lamoille NV, Sandy OR, and Umatilla OR.

Figures 7.7-7.8 show the correspondence (as measured by the NS) between the synthetic hindcasts and the official forecasts for January and April, respectively. Large filled circles indicate high correspondence (i.e., the synthetic hindcast values match those of the official hindcasts in any given year), small circles indicate little relation and hollow circles indicate a worse relation than using the long term average forecast. The outer reference circle indicates a perfect match between synthetic and official forecasts. In January, there is excellent agreement in Idaho, Montana, Utah and California and poor agreement in the Southwest, Olympic Peninsula and Tongue WY. The poor correspondence may be due to the January official forecasts being more affected by climate and/or long term soil moisture deficit information. Also, relatively few official forecasts issued on 1 January are available and the results may be an artifact of the small sample size.

The agreement in April is much greater over almost all of the Western US. The official forecasts also bear a better resemblance to the snow-based hindcasts than the precipitation based hindcasts. This result is consistent with the fact that the official forecasts used in this analysis are the primarily snow-based NRCS forecasts as opposed to the precipitation-based NWS forecasts. The low 1 April forecast correspondence in Lamoille NV is mostly due to one extreme year (1943), the same year with a large forecast error described in Church and Boardman (1944). As mentioned earlier, the forecasts were (deleteriously) based on the expected effects of a high water table. If the forecasters had based their forecasts solely on observed snow and precipitation, they might have fared much better according to this analysis.

Figure 7.9 shows scatter diagrams of the 1 April official and synthetic (snow +precipitation combination) forecasts for four locations around the Western US. The diagonal line for perfect correspondence is included. In general, there is exceptional correspondence between the official forecasts and the synthetically generated ones. The most disagreement occurs at the tails of the distribution, where the official forecasts display more variability ("bullishness") than the synthetic hindcasts. Without a detailed evaluation of the quality of the forecasts at the tails, it is impossible to say if the official forecasters conduct special analysis during unusual conditions, or simply overreact in the face of extreme objective guidance.

7.8 Sensitivity tests to measure relative merit of climate information

By itself, climate information is a relatively weak predictor of streamflow. For example, figure 7.10 shows a plot of the skill of forecasts issued at various lead times for snow, precipitation, climate and combined for a location in the southwest (top) and northwest (bottom). Late in the season, combined forecast skill is almost entirely dominated by snow and precipitation information, which each have about the same skill. As lead time increases, snow and precipitation hindcast skill rapidly erodes to where there is little to no skill before January. Climate-based forecast skill, in comparison, is low but constant versus leadtime. Early in the season, it is the only predictor available with which one can make a skillful forecast.

Figure 7.11 shows a map of the skill of the climate based forecasts, following the same convention of figure 6.2. The skill, outside of calibration, is very low and in some locations (i.e., California, Nevada) is negative. The skill is best in regions heavily influenced by El Niño and the Pacific Decadal Oscillation (the Pacific Northwest and the Southwest). In southeast Oregon and Idaho, the skill is not as strong as one might otherwise expect due to the large forecast error associated with the 2000-2001 La Niña drought in the Pacific Northwest (see also chapter 5.4).

Figure 7.12 is a box diagram of the relative importance of various climate indices in forecasting streamflow. Each box summarizes the absolute value of the calibration correlation coefficient between the individual climate index and streamflow for the 29 study basins. High values indicate strong correlation during the calibration of a 1 January forecast equation. The NAO index is a very poor predictor of western water supplies. The WP index is also a relatively poor predictor, whereas PNA, SOI, Niño3.4 and PDO have about the same relative importance, in part because of the intercorrelation of the various indices.

Overall, both El Niño and PDO yield low to moderate skill streamflow forecasts. In April, climate contributes an almost imperceptible amount of skill to a system that also considers snowpack and accumulated precipitation. In November and December, however, most of the available skill is due to climate information. Considering the invariance of climate forecast skill with lead time, it is possible that a forecast in September may also have about the same skill as a forecast issued in December.

7.9 Summary

This chapter developed and tested a system to use observed snow, accumulated precipitation and climate indices to objectively hindcast seasonal streamflows. A combined system of using snow and precipitation data produces hindcasts that bear a remarkable resemblance to the official historical NRCS water supply forecasts. These hindcasts also have the same accuracy properties of the official forecasts, in both space and with leadtime. As with the official forecasts, 1 April hindcasts perform very well in most regions except Oregon and the Great Plains (northeast Wyoming). Hindcasts issued 1 January perform well in the Columbia basin, Montana and Arizona. These hindcasts suggest that purely accumulated precipitation-based water supply outlooks have very modest skill on 1 December but almost no skill on or before 1 November. If climate information is included, there is very little change in the overall skill of the 1 April hindcasts. Hindcasts issued 1 January benefit somewhat from climate information, and almost all of the available, albeit modest, skill before December is due to climate information. This information is most effective in the regions impacted by El Nino (i.e., the Pacific Northwest and Southwest US but almost no locations in the Great Basin or Upper Colorado basins).

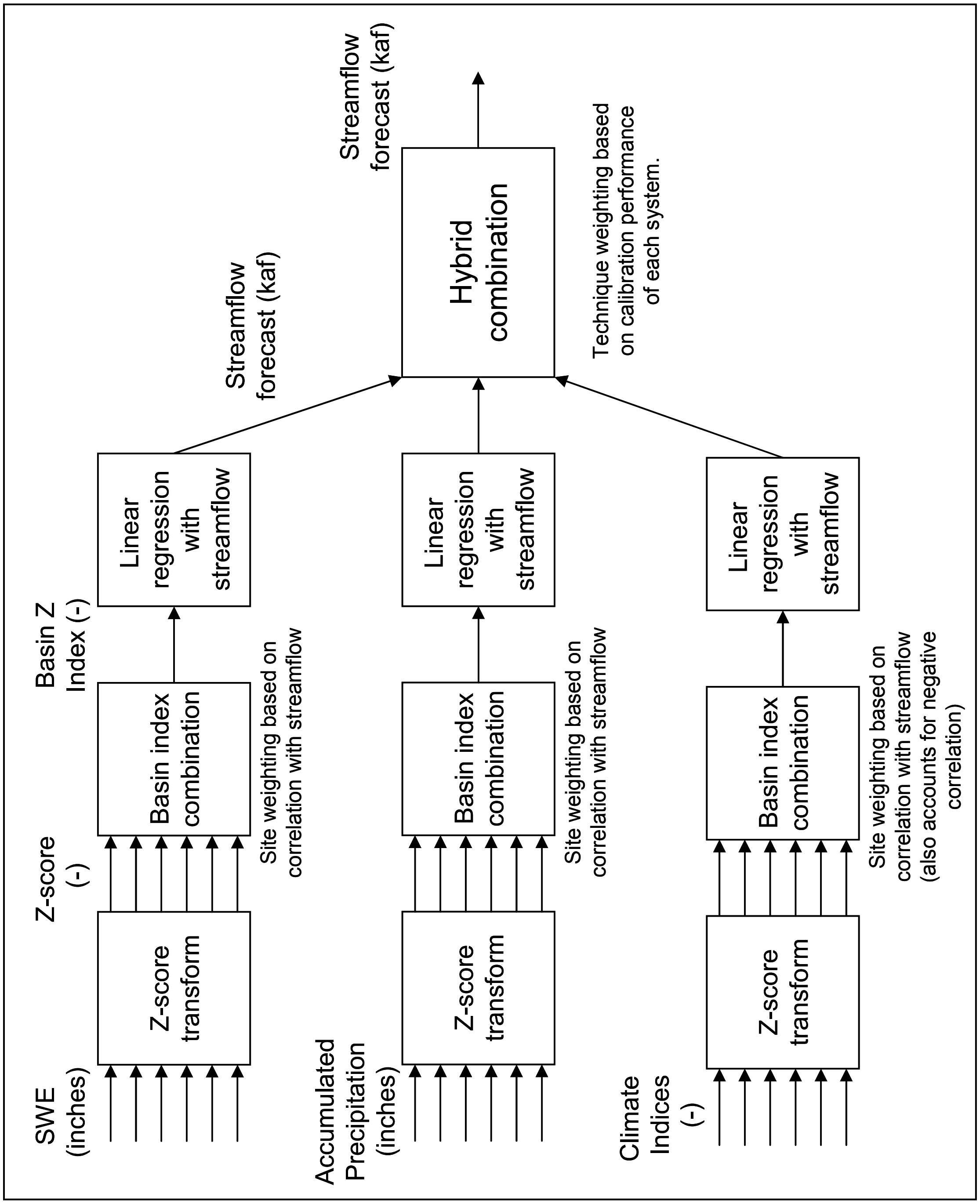


Figure 7.1 Schematic of the forecast creation and combination system described in chapter 7.

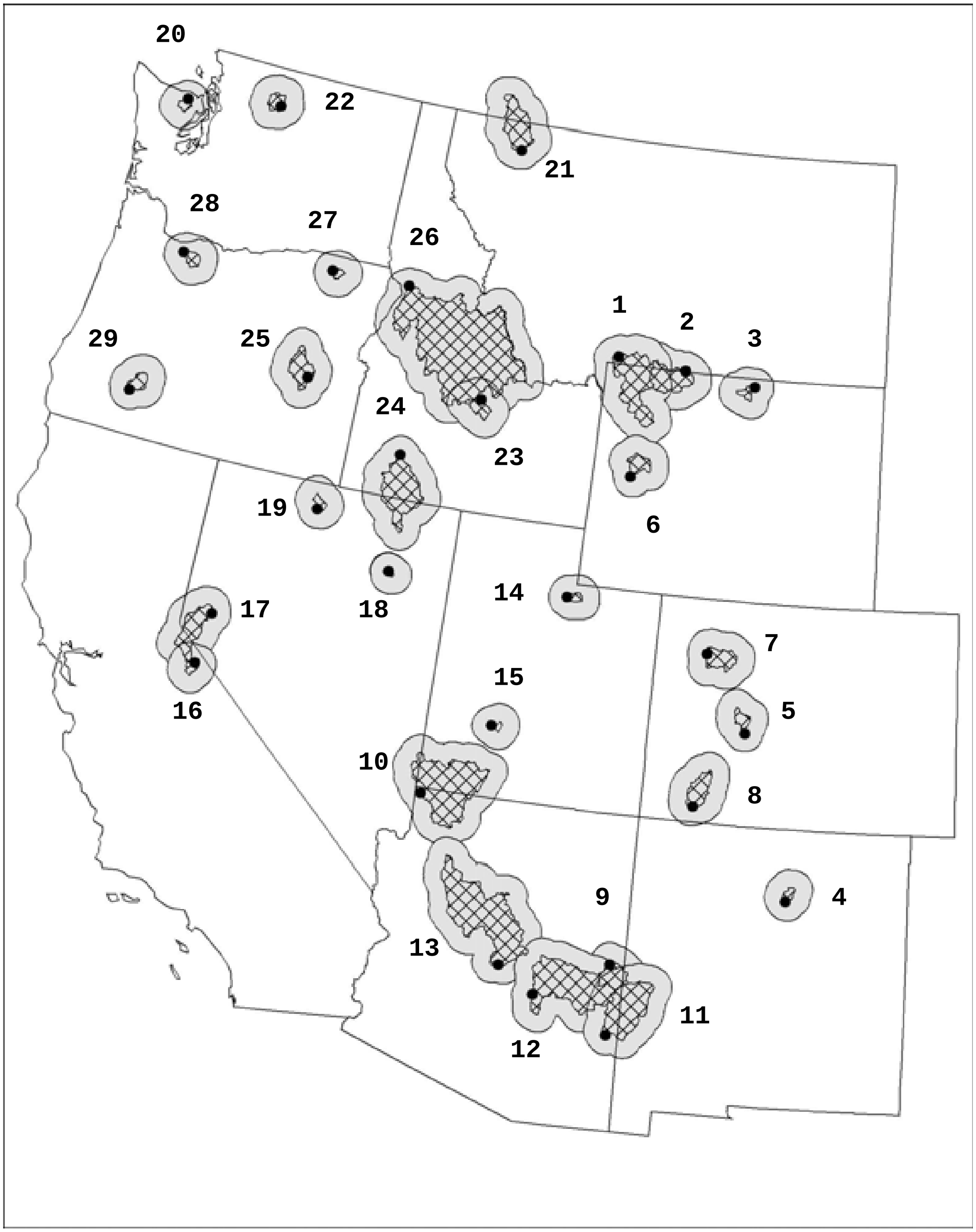
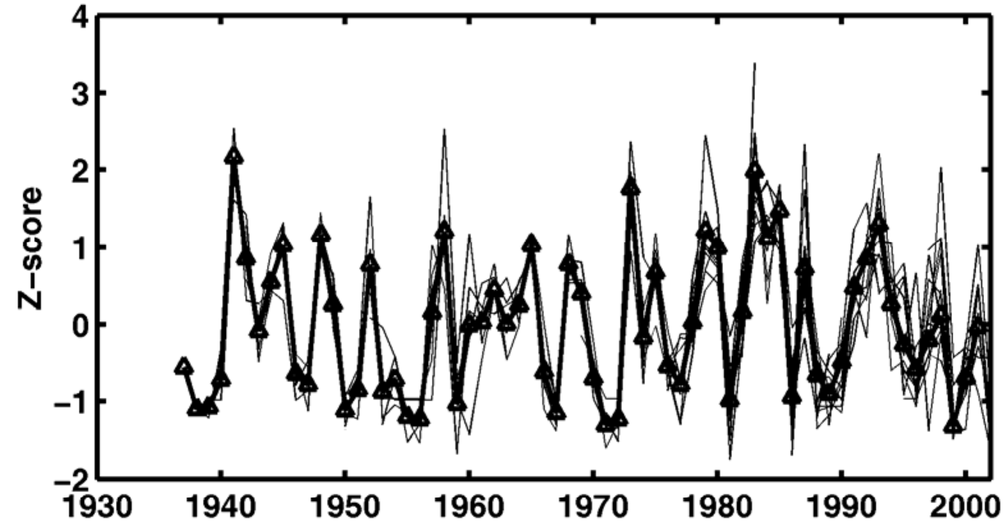


Figure 7.2. Map of study basin streamgage locations (solid dots), basin boundaries (cross hatching) and 35 km buffers (solid gray). Numbers refer to entries in table 2.4.



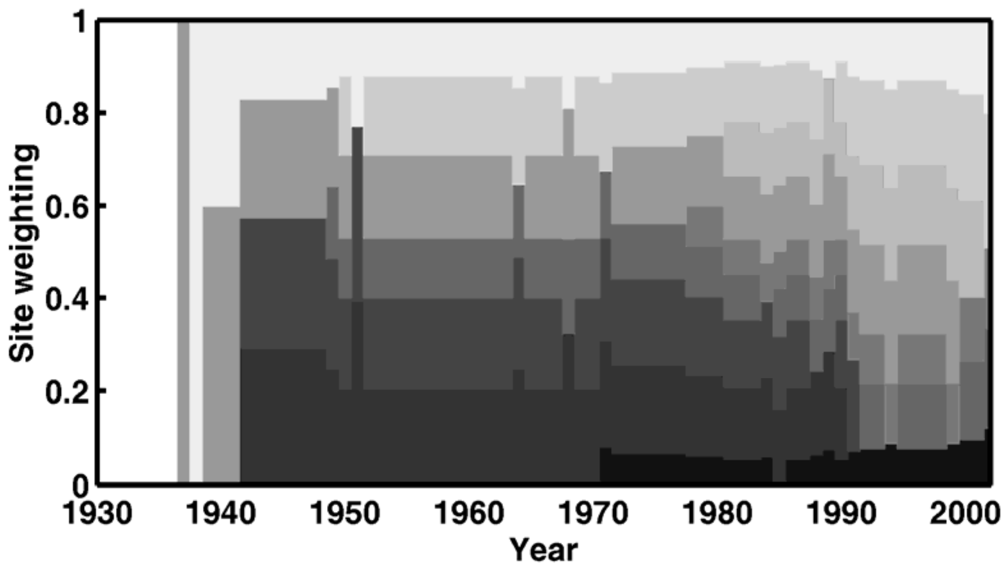


Figure 7.3. Visual depiction of the development of a basin-wide snow index for the Pecos River near Pecos. Top: light lines indicate 1 April snow measurement Z-scores for nine individual sites. Heavy line with triangles represents the basin-wide composite. Bottom: cumulative site weightings used in developing the basin-wide snow index. Each color represents the relative contribution of each site versus time. Note that a few sites at the beginning of the record receive most of the weighting while in the 1990's, the weighting is split among a diversity of sites.

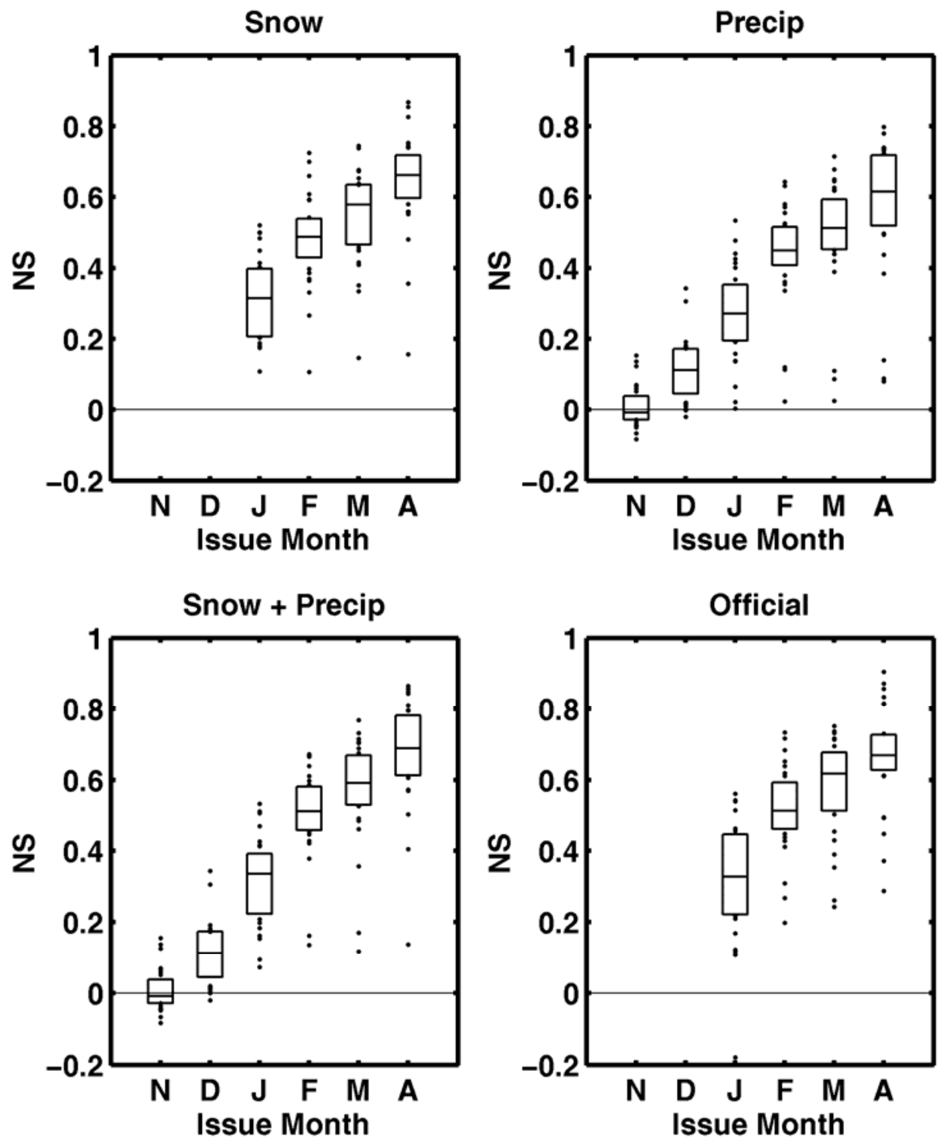


Figure 7.4. Evolution of skill (NS score) versus forecast issue month for the 29 study basins for the entire period of record. Top left, right) synthetic snow and precipitation hindcasts, respectively. Lower left) synthetic snow/precipitation consensus forecast Lower right) Skill of the official NRCS historical forecasts. Boxes represent quartiles and median, with outliers represented as dots.

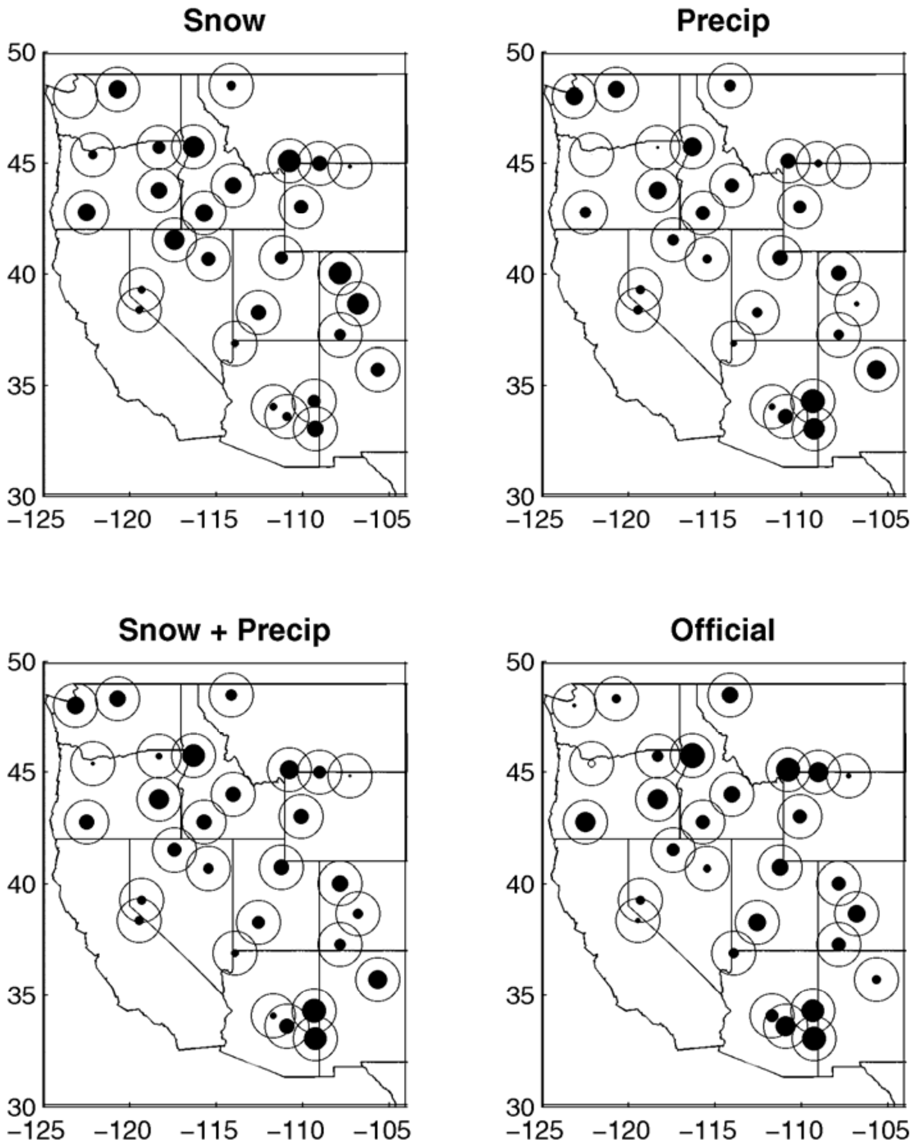


Figure 7.5 Maps of 1 January forecast skill (NS score) for 29 study basins for the entire period of record. Top left, right) synthetic snow and precipitation hindcasts, respectively. Lower left) synthetic snow/precipitation consensus forecast Lower right) Skill of the official NRCS historical forecasts. Inner symbol diameter is linearly proportional to forecast skill, with the outer circle representing “perfect” forecasts and the smallest circle indicating no skill. Hollow circles (i.e. Sandy near Marmot Official) indicate negative skill.

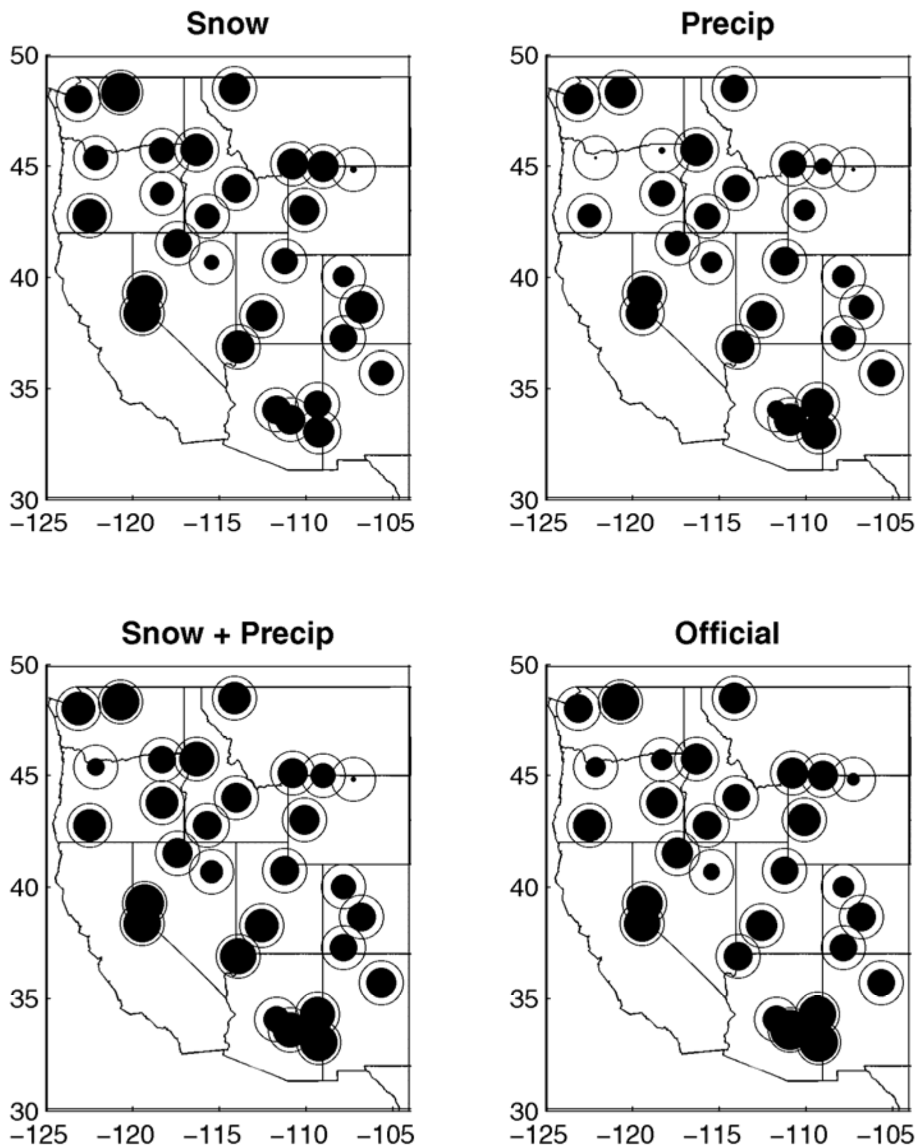


Figure 7.6. Same as figure 7.5 except for 1 April forecasts.

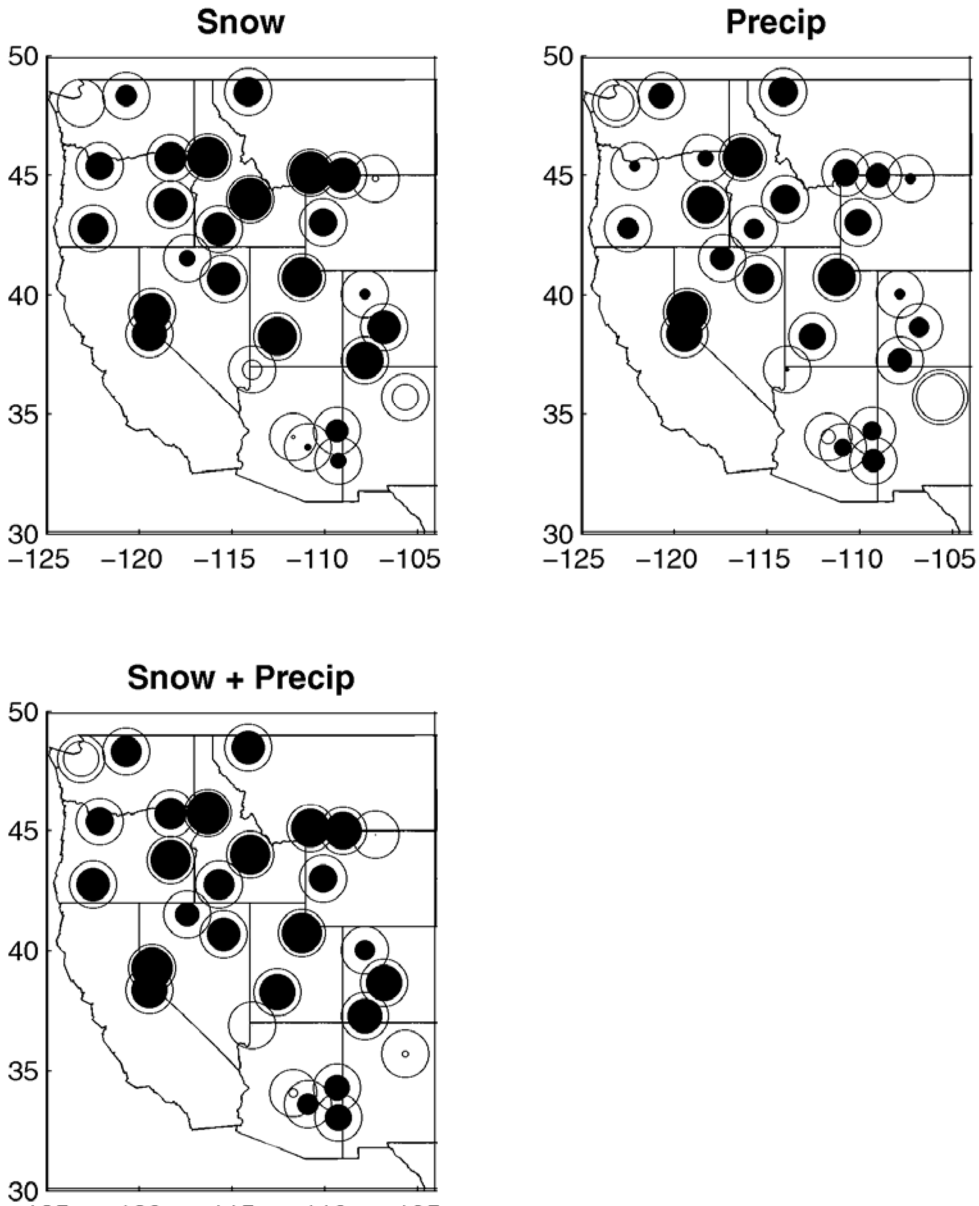


Figure 7.7. Correspondence (NS) between synthetic hindcasts and official forecasts for 1 January. Map convention is the same as figure 7.5 Hollow circles mean negative NS (e.g. Pecos, NM). Poor correspondence in the Southwest and Pacific Northwest are in part due to the short time series of 1 January forecasts. Also, official forecasts consider soil moisture deficits and climate predictions whereas the synthetic hindcasts do not.

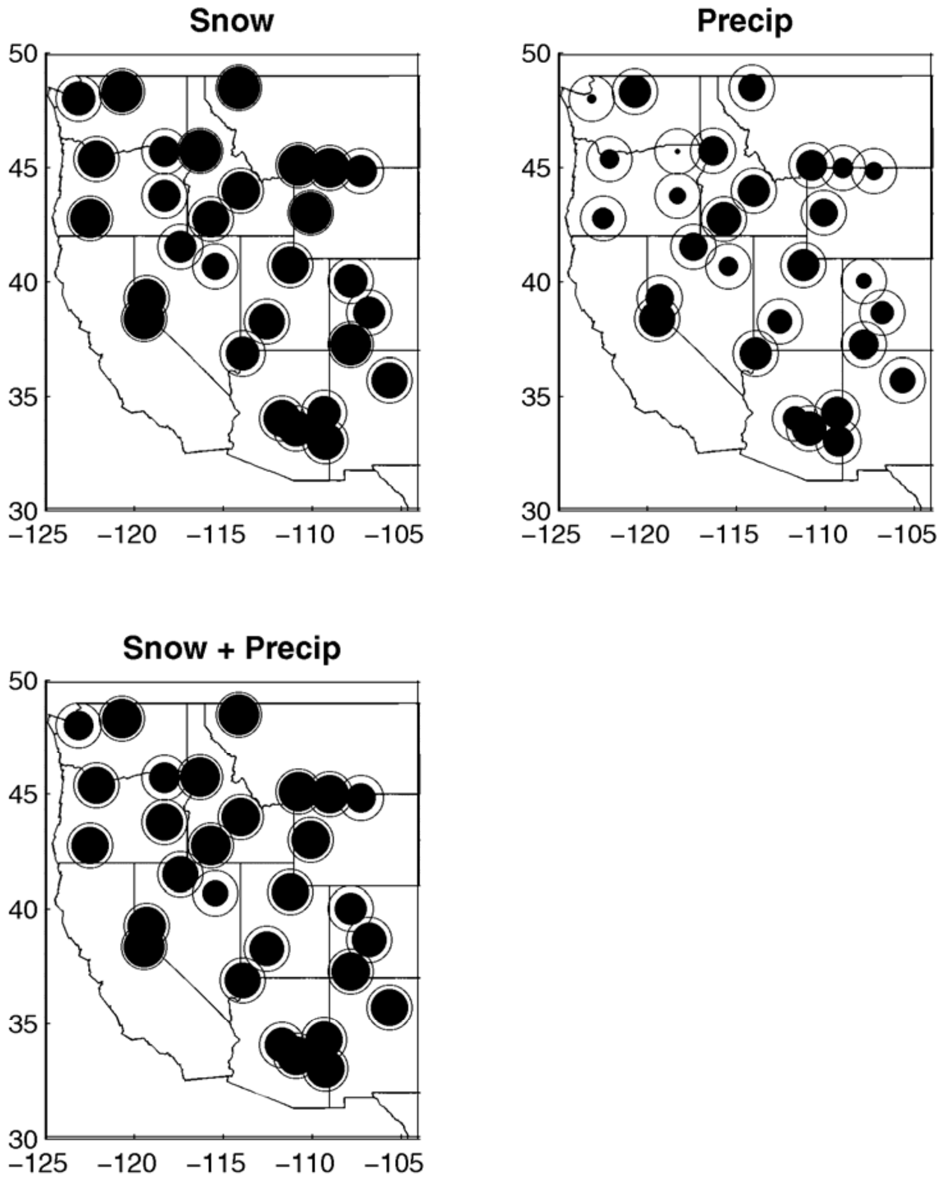


Figure 7.8. Same as figure 7.7 except for 1 April forecasts.

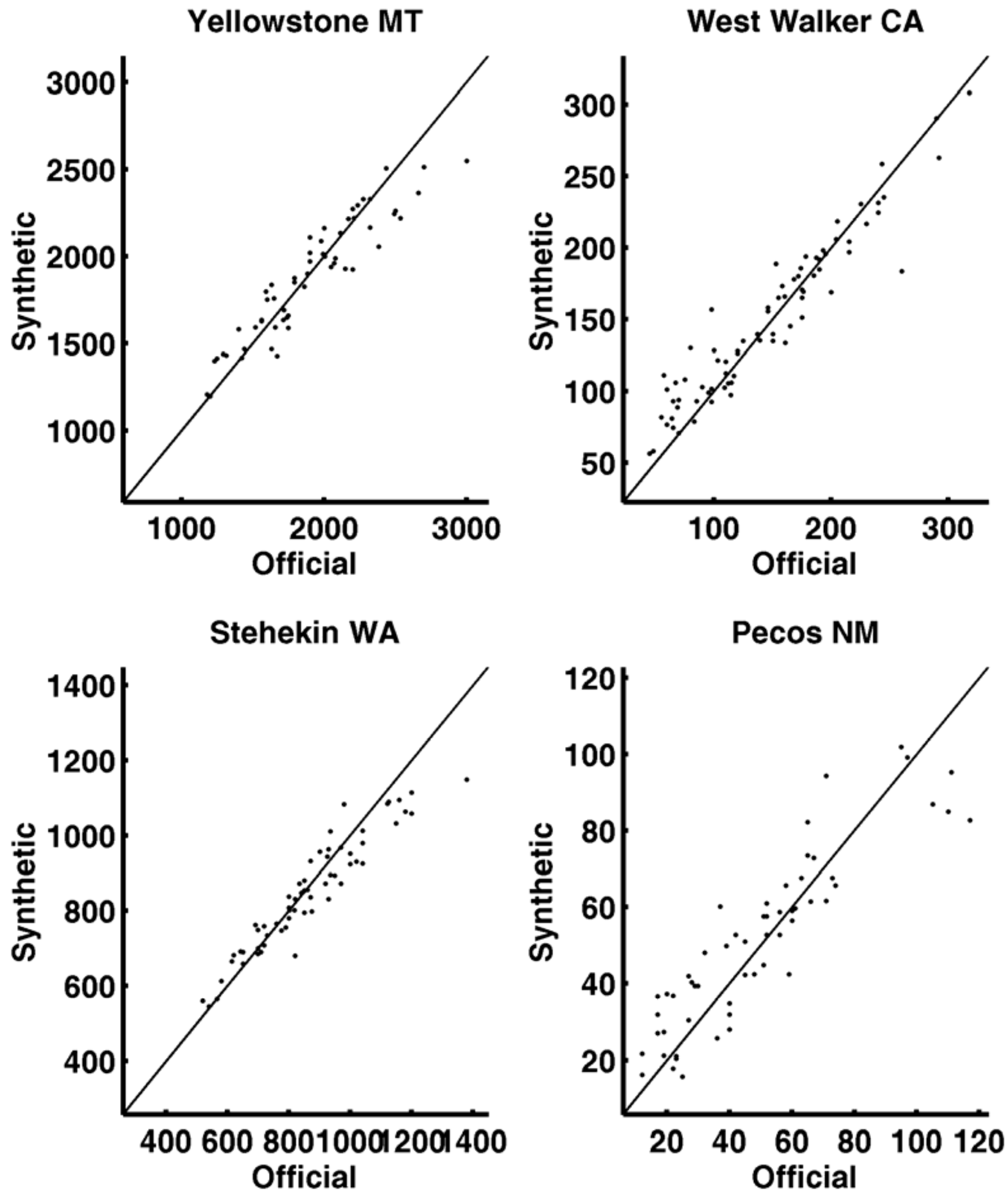


Figure 7.9. Representative sampling of 1 April official water supply outlooks (X-axis) versus the synthetic snow+precipitation hindcasts (y-axis) for four locations around the Western US. All measurements are in 1000's of acre-feet. Perfect correspondence is shown by the diagonal line.

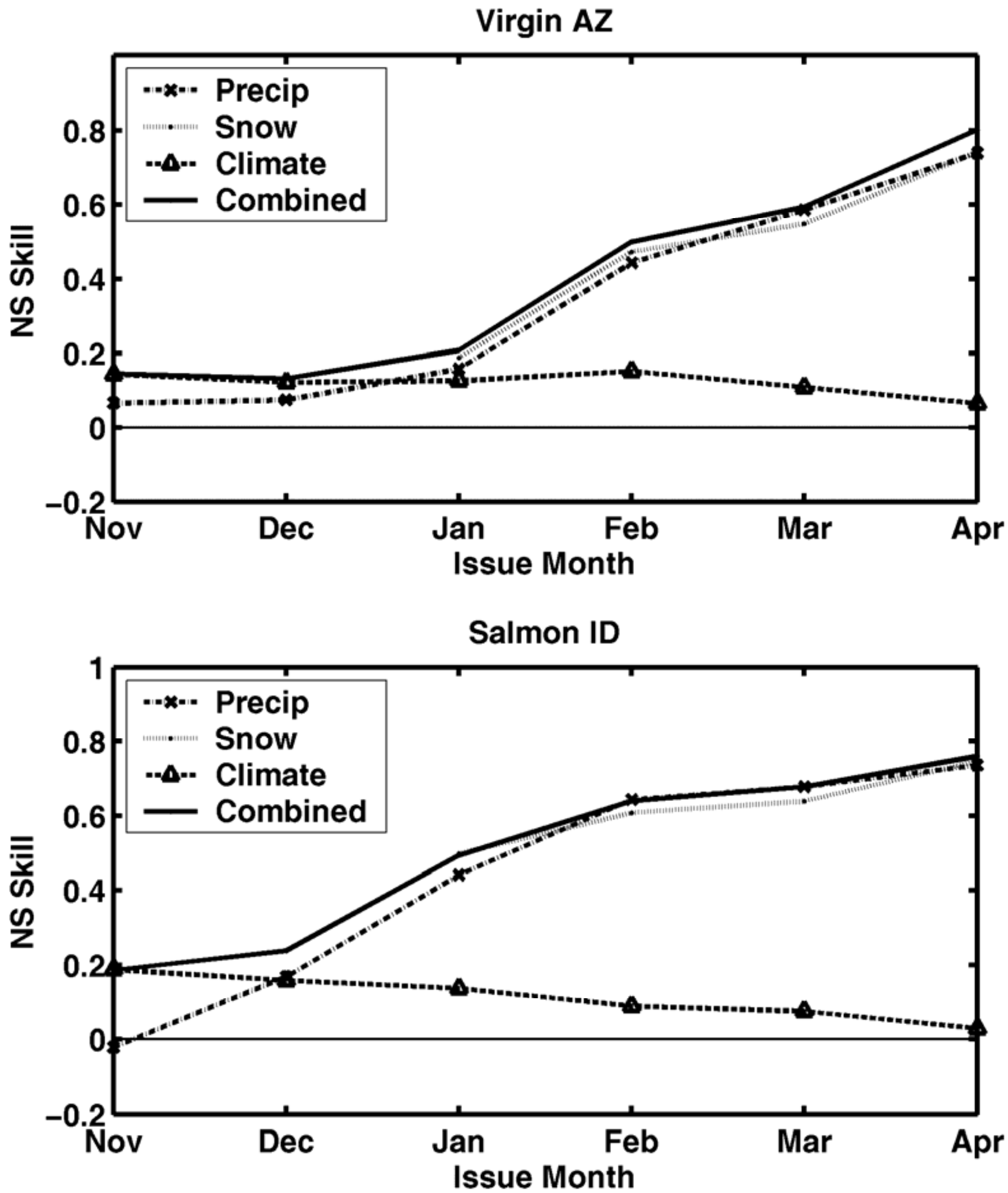


Figure 7.10. Time series of forecast skill components versus issue month for the Virgin AZ (top) and Salmon ID (bottom) rivers. Skill components are shown using different lines. The solid line represents the skill of the combined snow+precipitation+climate synthetic hindcasts. Skill is measured over the entire period of record.

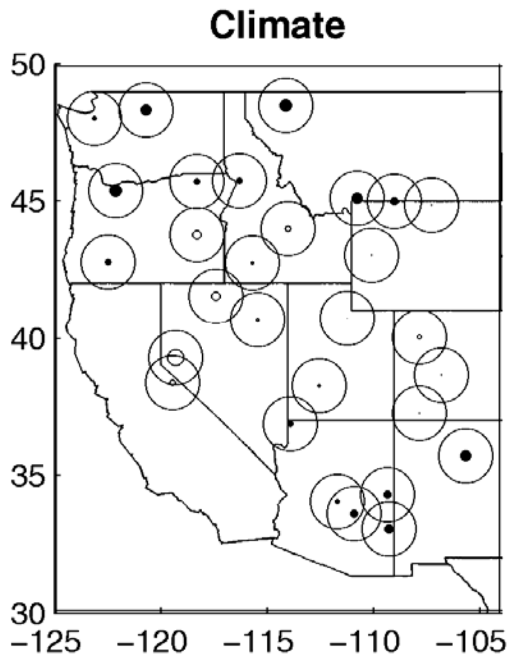


Figure 7.11. A map of the period of record skill (NS) of 1 January climate-based synthetic streamflow hindcasts. Figure convention is the same as figure 7.4. Hollow inner circles (e.g. Nevada) indicate negative skill. Large outer circles are provided for reference for perfect forecasts.

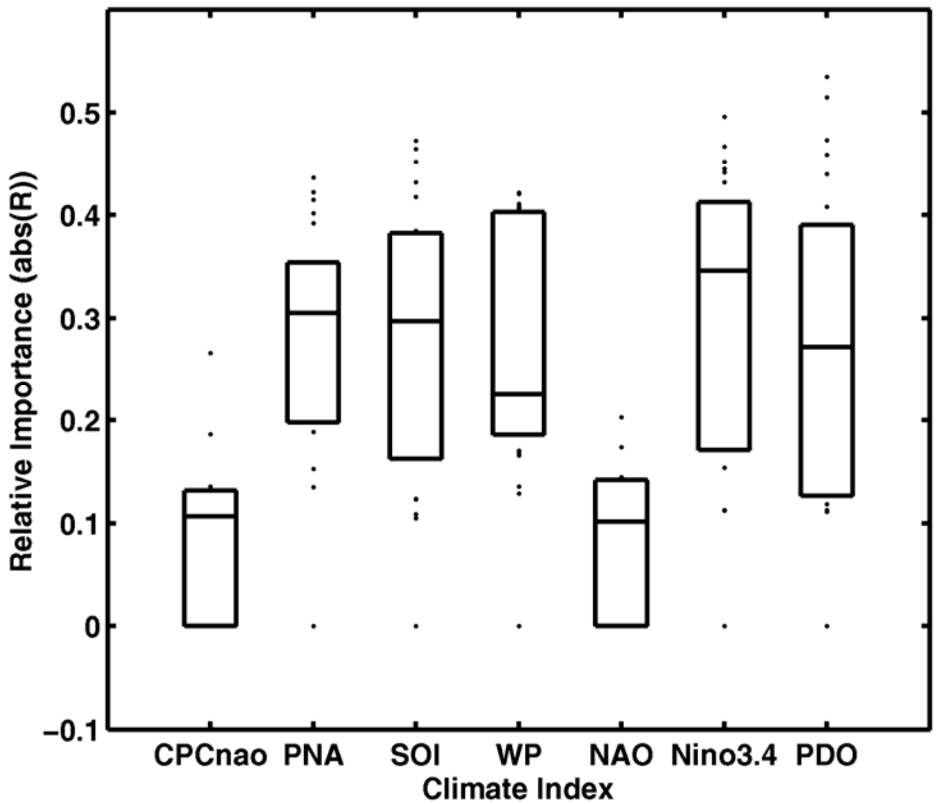


Figure 7.12. Relative importance (absolute correlation coefficient during calibration of 1 January forecasts) of various climate indices in predicting streamflow. Box plot notation same as figure 7.4. The NAO is a relatively poor predictor of western water supplies.

8. DECADAL VARIABILITY AND WATER SUPPLY FORECASTS

8.1 Introduction

Hydrologists have long recognized the relationship between forecast error and unknown future precipitation. Reducing uncertainty about future precipitation has always been looked on as a promising avenue to improving water supply forecasts. This and other studies have shown that, during the heart of the water supply forecast season, the existing state of the art in climate prediction can still only explain a small part of interannual precipitation variability. This modest predictive capability is all that is available, however, when it comes to issuing water supply forecasts before any snowpack has accumulated. Previous sections and chapter 9 discuss how operational water supply forecasters can best take advantage of existing climate forecast skill (i.e. in what parts of the country skill can be expected, at what times of year, and so on).

While the science of climatology may only be able to explain or predict a relatively small part of the overall seasonal-to-interannual variability in climate, the currently unexplainable aspects of climate variability still have an enormous impact on water supply forecast activities. Among other things, hydrologists are keenly interested in the causes for and implications of long-term climate variability and trends. Will the next 20-30 years be of a significantly different character than the recent 20 years? Will current water supply forecasting techniques fare well in this changed climate, or are different forecasting tools necessary? Statistical water supply forecasting techniques assume that climate is stationary and random; is this a valid assumption? If it is not, what is a better assumption?

This chapter begins with a discussion of the role of the Pacific Decadal Oscillation (PDO) in water supply forecasting. This feature is responsible for low frequency variability in Western US precipitation and thus seasonal streamflow. This chapter will test and evaluate a method for water supply forecasting using PDO information as suggested by several researchers.

This chapter will also investigate any observed long-term variability in water supply forecast skill. It will show that there is a recent downward trend in forecast skill caused by a resurgence of extreme springtime precipitation events. If the cause and expected future behavior of these extreme springtime events are discovered, hydrologists may have a great opportunity to control, or at least anticipate, decadal variability in water supply forecast skill.

8.2 Is subsetting by PDO state a good forecast strategy?

Koch and Fisher (2000) and others have found that the PDO index, by itself, is not a very good predictor of streamflow in any single year because is a decadal index but it also exhibits high frequency variability typical of mid-latitude phenomena. This high frequency variability is of questionable usefulness in predicting streamflow. Rather, the relevance of the PDO could be in its modulating influence on ENSO activity. Koch and Fisher show that the correlation between Cascadia streamflow and El Niño is about 0.3-

0.5 for all years. The correlation increases to 0.6 during low PDO years and decreases to 0.2 during high PDO years. Gutzler et al. (2002) showed a similar pattern in Southwestern US precipitation. Before the 1977 PDO shift, La Niña is strongly related to dry conditions in the Southwestern US yet El Niño is not necessarily related to wet conditions. After the 1977 shift, El Niño is a strong indicator of wet conditions and La Niña predictability breaks down.

Some have advocated the development of separate streamflow forecast equations for high and low PDO years (Alan Hamlet, University of Washington, personal communication, 13 January 2003). For example, one equation relating snow and streamflow would be calibrated using only high PDO years and another equation would be developed using only low PDO years. In realtime forecasting, the hydrologist would select the equation whose calibration PDO state matches the current PDO state. There are some potential challenges associated with such a system. First, not all basins are endowed with a long time series of continuous data. For example, some operational forecast equations are developed on fewer than 15 years of data. Subsetting that data would yield even fewer years for calibration. The gain in PDO-related skill might not offset the loss in skill due to the shorter calibration period. To counter this challenge, one might develop regional relationships between PDO and streamflow for a select group of streamgages with very long periods of record. Forecasts for the subset of streamgages could be extrapolated to locations with shorter periods of record using standard operational routing procedures (such as those used to translate forecasts at headwater locations into forecasts for mainstem points on major rivers). Such an approach would be worthy of future research efforts.

After deciding on the basis for splitting up the forecast equations (e.g. high PDO vs low PDO), one must determine how to categorize the historical years, as well as the current year. “Before 1977” and “After 1977” is a popular date for such subsetting. Hamlet and Lettemaier (1999, 2000) indicates the PDO regime shift dates are 1925, 1947 and 1977. Currently, what is the most recent PDO state? Hamlet has implied that a PDO shift could have occurred in 1996 or 1998. McFarlane et al. (2000) believe it was 1997 as does FOCI (2002). Schwing and Peterson (2003) suspect the regime change happened in 1998 when the PDO index changed from highly positive to highly negative. Others believe it might take many years to retrospectively determine the year of the regime change, if one has happened at all (Michael Dettinger, US Geologic Survey, personal communication, 21 July 2004). It does not help that the PDO index changed from highly negative in 2002 to near record positive in 2003. It is unclear which is the transient feature, the negative values in 1998-2002 or the positive values in 2003?

On a related issue, one hydrologist has gone as far as to say that the water supply forecasters should develop separate forecasting equations, one that includes climate information and one that does not (Phil Pasteris, NRCS Water and Climate Center, personal communication, 4 August 2004). Given that sometimes climate “works” and sometimes it doesn’t, the forecaster should use the non-climate equation in those years that the climate-streamflow relationship breaks down. The 2001 La Niña and the Pacific Northwest drought (see chapter 5.4) is given as an example of a year in which climate did not “work”. The flaw in this strategy is that one does not know in advance whether climate is going to “work” this year or not; one only knows after the fact. Additionally, a season that begins with precipitation contrary to the expected climate signal will not necessarily continue that way throughout the remainder of the season. In other words, if the climate-based guidance suggests a wet fall, winter and spring and the October-January precipitation has been unusually low, the forecaster may be tempted to assume that climate is not “working” this year and abandon the climate-based guidance. The logic in doing so would be just as flawed as the opposite extreme of assuming that because a dry fall occurred, the climate was “storing up” its wet signal and would definitely “unleash” it in the spring (more so than the objective climate forecast might indicate). A forecaster may end up randomly selecting equations, throwing away some opportunities and falling into traps in other years.

This study uses four configurations to test the relative merit of the proposed PDO-based subsetting scheme. First, a control case uses snow, precipitation and climate data, excluding the PDO index in the hybrid system described in chapter 7.6. The second case is the same as the first, except that the PDO index is included in the list of candidate variables. The first and second cases use all years from the period of record in their calibration (except the jack-knifed year being hindcast). The difference in skill between these two cases is the relative merit of using PDO information as an index over not using PDO information at all.

The third and fourth cases are the same as the first case (excluding the PDO index among candidate variables) except that the calibration years are subsetted by PDO state.

The sign (positive or negative) of the PDO index average for the 3 months prior to the issue month is used to determine the PDO state of that year. In the third case, the equation corresponding to the PDO state of the current year is selected. This scenario is the equivalent to the approach used by Koch and Fisher (2000). In other words, if this year's PDO value is positive, the hydrologist uses the forecast equation whose calibration years are also PDO positive. The fourth case selects the equation calibrated on the PDO state **opposite** that of the current year. This final experiment is akin to the hydrologist who tries to guess what the current PDO state is, but always selects incorrectly and uses the "wrong" PDO equation. The difference in skill between these two cases provides an estimate of the importance of accurately knowing PDO state when selecting the forecast equation. Both experiment 3 and 4 relax the minimum number of historical years needed to calibrate a regression equation from 20 now to 15 (see chapter 7.3). As mentioned previously, subsetting by PDO state can halve the available data for calibration of a forecasting equation. A limited set of study sites (particularly the Dungeness near Sequim, WA) have less than 40 but more than 30 years of snow measurements. Without the relaxation of the period of record limit, no analysis would have been possible for these sites.

Experiment 4 also accounts for the changes in forecast skill solely due to the subsetting reducing of the length of the calibration data set; experiment two may have an advantage over experiments 3 and 4 because it is able to select from a larger pool of calibration years. If subsetting by PDO is a good strategy, experiment 3 must yield more skillful forecasts than experiment 4. If it is the ideal strategy, then experiment 3 must outperform experiment 2. As with elsewhere in this dissertation, the forecasts are developed using a jackknife technique and are evaluated by the NS score.

Figure 8.1 contains four scatter diagrams of the skill of the various experiments previously described for hindcasts developed on 1 November. Each diagram contains the period of record NS score of 1 November hindcasts where each point is one of the 29 study basins. In all figures a 1-to-1 line is provided for reference. The top left figure compares experiments 1 (no PDO) and 2 (PDO as an index). For the 9 data points above the 1-to-1 line, the inclusion of PDO information adds a small amount of skill to forecasts made in November. All but one of the nine basins that improve with the inclusion of PDO as an index lay north of 42 degrees latitude (the California/Oregon and Utah/Idaho border).

Next, the top right figure compares experiments 1 (no PDO) and 3 (matched PDO subsetting). Although the PDO subsetting scheme does improve the forecasts for a select number of locations above the line (in Washington, Northern Montana and the Virgin AZ), it degrades the forecast skill for a majority of points including the Southwest US.

Third, the bottom left figure compares experiments 2 (PDO as an index) and 3 (matched PDO subsetting). Only two basins above the line improve by PDO subsetting (Malheur OR and Virgin AZ), and for the other 27 sites, using the PDO index as an input variable in the regression equations is a better strategy.

Finally, the bottom right diagram shows the performance of experiment 3 (matched PDO subsetting) versus experiment 4 (mis-matched PDO subsetting). Interestingly, consistently choosing the wrong PDO state improves the forecasts in half the basins (in particular AZ, CA/NV, UT, CO, southern Idaho and eastern Oregon) and degrades the skill in the other half of basins. For some locations, however, choosing the wrong PDO state yields significantly worse forecasts than choosing the correct PDO state (especially points in Washington and Northern Montana).

Figure 8.2 is the same as figure 8.1 except for hindcasts issued in January. The overall picture is the same. PDO as an index improves the forecasts over not having any PDO information at all. Subsetting by PDO state is a worse strategy than using PDO as an index. Always choosing the wrong PDO state is somewhat but not always just as good a strategy as choosing the correct PDO state.

8.3 Decadal variability in water supply forecast skill

Although PDO subsetting does not yield significant improvements in jack-knife forecast skill, operational agencies are still interested in whether or not water supply forecast skill exhibits decadal and long term variability. Water supply forecast errors might not be completely random; errors may cluster together during particular epochs, contrary to the assumptions of climate stationarity. The technological determinist would expect that forecast skill is monotonically increasing as technology and scientific understanding improve. Investments in science yield reductions in forecast uncertainty and narrowing of uncertainty bounds. An extreme pessimist might note the dramatic organizational changes within the NRCS (i.e. the downsizing of forecast personnel, the gathering of hydrologists in a national center removed from the basins being forecasted) and expect an erosion of forecast skill. This section measures if there have been changes in forecast skill and subsequent sections search for causes responsible for these changes.

There is no shortage of human factors that may potentially influence trends in forecast skill. The early 1980s saw a major restructuring of forecast facilities within the NRCS, from being state-based to being centrally located (Barton 1983). The NRCS forecasting staff in 2003 was only one third of its size in 1980. Garen (1992) developed a significantly different statistical forecasting technique that found wide use after the 1990s. Some NWS offices adopted seasonal simulation modeling of streamflow and the ESP system in the late 1970s.

The automation of snow courses was phased in over the 1980s with the advent of the SNOw TELemetry (SNOTEL) network, which was an improvement but also a discontinuity in data collection technology. Changes in land use, small water impoundments, and undocumented diversions could have affected the future representativeness of historical flow. It is unknown whether the current snow-based forecasting equations are representative under a climate that is warming and oscillating on decadal timescales. The forecasts were objectively based initially, but the published values were sometimes adjusted using non-quantifiable and non-reproducible human professional judgment. Both the statistical procedures and the human operators changed over the history of forecasting, as will they change in the future (although it would be misguided to ascribe a significant rise or fall in skill to an individual person, given the many parties involved with creating a forecast). Some, all, or none of these factors may have shaped how forecast skill evolved recently.

Figure 8.3 (top) documents the observed trend in westwide average forecast skill for 20-year moving window periods. Data must be serially complete in 20-year window to be computed for a basin. At least 8 basins must have a valid NS for a value to be shown. Plotted is the westwide average of all of the available NS scores. Given the short period of record of January forecasts, it is difficult to measure any kind of trend in such long-lead time forecasts. February forecast skill appears to be steadily increasing. Most interestingly, 1 April forecasts were least skillful in the 1940s-1960s, reached a relative maximum in skill from the 1960s-1980s and rapidly declined afterwards. There is some evidence of an upswing in 1 April forecast skill in the past few years although it is too soon to say if that improvement will persist.

Figure 8.3 (bottom) shows the same trends in forecast skill, except for the objective synthetic snow+precipitation hybrid hindcast system developed earlier in this study (chapter 7.6). It is highly intriguing that this system almost exactly reproduces the decadal trends in forecast skill. Figure 8.4 shows this coincident variability more clearly. Here, the westwide skill time series is presented as an anomaly relative to the forecast skill for each leadtime from 1960-2000. Again, official forecasts (dashed) were poor in the beginning of the period of record, peaked in skill in the mid-1960s-1980s and sharply declined afterwards. The synthetic hindcast performance is shown as the solid line, almost exactly matching the decadal variability in official forecast skill.

The interpretation and implications of figures 8.3-8.4 are that, for better or for worse, many of the human factors described above are practically irrelevant to long term trends in water supply forecast skill. Considering that the synthetic hindcast system is entirely objective, the downward trend in skill is in no part due to irresponsibility or fallibility among the current generation of human forecasters. Simply, the forecasters followed their guidance and it led them astray in recent years. Figure 8.4 also implies that the peak in skill the 1960s-1980s is unusual relative to the entire period of record, is not evidence of a higher “caliber” of human forecasters during those years. Figures 8.3-8.4 suggest that the recent “slump” may be more of a return to normal, and that far worse forecasting periods existed (i.e., the 1940s-1960s).

Figure 8.5 contains maps of the relative skill of the 1 April official forecasts over four 20-year periods. Shown is the anomaly in forecast skill, relative to the period of record skill for that location. Large filled circles indicate a 20-year period where the skill of the forecasts was considerably less than other periods (“a slump”). Hollow circles indicate skill better than other periods (“a streak”). Small circles show skill near to the long term average skill. Data must be serially complete within the 20-year window for a symbol to be shown.

This figure shows that in 1951-1970 (upper left), the interior west had relatively low forecast skill for almost every forecast point. 1961-1980 (upper right) was an especially poor period for Lamoille, NV and the Great Basin, although forecasts performed well in the Upper and Lower Colorado basins and parts of the Pacific Northwest. The “Golden Age” of forecasting in 1971-1990 (lower left) had no location in the west with very poor skill, and had many locations with skill much above the long term mean (e.g., Montana and the Pacific Northwest). Afterwards, skill collapsed in

Pecos, NM, the Colorado Basin and in Oregon in 1981-2000, as shown by the large filled circles in the lower right panel.

Figure 8.6 is the same as the previous figure except for the synthetic snow/precipitation hindcasts. The synthetic hindcasts had poor skill in the early period of record in the great basin and the Columbia (upper left and upper right). In 1971-1990 (lower left), every location in the west experienced a “streak” in forecast skill, only to have skill fail in the southwest and Oregon in the years that followed (lower right). Once again, because this feature is reproduced in the objective hindcast system, it is entirely misguided to lay blame on the Oregon or Southwestern US hydrologist for the decline in skill scores in those regions.

What, therefore, is responsible for the recent downward trend in skill? The previous analysis proves that the objective guidance is not as good a predictor of streamflow as it once was. This statement can be decomposed into two parts. First, has the relationship between snow data and actual basin snow conditions changed? This is a question of data quality and it is possible but highly unlikely that the changes in snow measurement technology (i.e., snow courses to SNOTEL) cause snow data to be now less accurate than they used to be. One way to test the data-quality hypothesis is to see if the trends in skill are reproduced in an independent data set. For example, if poor data quality in the NRCS were worsening the water supply forecasts, one would not expect to see similar erosion in skill in forecasts based on accumulated COOP precipitation.

Figure 8.7 shows the long-term trends in 1 April forecasts based solely on snow (heavy dashed) and precipitation (solid) data, as well as the combined snow+precipitation hybrid system (light dashed). This figure shows that throughout most of the period of record, snow data has performed better than precipitation data at forecasting streamflow. Only in the most recent period has precipitation matched and overtaken snow as the most skillful predictor. In the early period, snow data accounted for almost all of the skill in the combined forecast system but after the 1980s, the skill of the snow and snow/precipitation hybrid systems diverge. In the virtual world of this experiment, it seems like in the early part of the record, the NRCS (using the snow hindcast system) could have done without guidance from the NWS (using the accumulated precipitation hindcast system). However, after the 1980s, both "agencies" have equally skillful guidance, and they both benefit by "coordination".

Regretably, one cannot draw firm conclusions about the potential erosion of forecast skill due to NRCS data quality issues. The decline in skill of the snow-based hindcast system could be due to a change in snow-rainfall precipitation partitioning (Serreze et al. 1999). It is also possible that the normally more accurate, and thus more "bullish" snow hindcast system performed much worse than a normally less accurate and thus more "conservative" precipitation forecast system only in a recent set of highly unusual years.

The second question posed earlier in this chapter asks whether the relationship between actual basin snow conditions (as a natural indicator) and seasonal streamflow changed? In figure 8.7, the quality of the snow measurement may be the same, but a given state of snowpack no longer precedes a certain character of seasonal streamflow.

There are several reasons why this might be the case, and the issue is explored further in the next several sections.

8.3 Are Western US streamflows becoming more erratic?

When natural indicators change in ways that are unusual in the modern instrumental record, the antropogenic release of "Greenhouse Gases" is often suspected as a possible forcing mechanism for this change. As mentioned in chapter 1 and section 3.6, humans are currently engaged in what has been described as a grand irreversible and "uncontrolled experiment whose [environmental] consequences could be second only to global nuclear war" (Fraser 1999). Since World War II, the emission of carbon dioxide to the atmosphere has increased almost sevenfold, due to the burning of fossil fuels. This rate could increase dramatically during the impending modernization and industrialization of developing nations. Carbon dioxide in the atmosphere contributes to the "Greenhouse Effect" and could raise the Earth's temperature at an unprecedented rate. The resulting changes on the rest of the climate system are, as of yet, uncertain.

Water managers have been increasingly concerned about the expected impacts of climate variability and anthropogenically induced climate change on the hydrology of the Western US (Dracup 1977). As suggested by historical data analyses and model simulations, these variations and changes include shifts in the balance between snow and rainfall, resulting in earlier snowmelt and reduced late summer streamflows (e.g., Hamlet and Lettenmaier 1999; Leung and Wigmosta 1999; McCabe and Wolock 1999; Mote

2003a,b, 2005), and the possibility of an intensification of extreme hydrologic events (i.e., droughts and floods; Hamlet and Lettenmaier 1999; NAST 2000). Such climate-induced changes could add another layer of complexity to the management of natural resources in an already challenging environment of changing demographics and competing interests. Climate change may interfere with what were once reliable natural indicators of future streamflow, leading to degradation in the streamflow forecasts.

It is not the place nor ability of this study to make statements about the origins, causes, impacts or expected future behavior of climate change in the Western US. Nonetheless, climate stationarity is a fundamental assumption of the statistical forecasting techniques used by hydrologists. A stationary process is a random process where all of the statistical properties (e.g., the mean and variance) do not vary over time. In other words, the future is entirely random but it bears the same overall characteristics of the past. If climate is not stationary, that is, it exhibits low frequency shifts or is experiencing monotonic change, the implications for statistical water supply forecasting are serious.

Several studies have focused on detecting trends in mean streamflow and in the magnitude of extreme events. Lettenmaier et al. (1994) found upward trends in monthly and annual streamflow volumes across most of the US during 1948-1988. Mauget (2003) identified an increase in annual streamflow volumes after the 1970s, primarily in the Southeast, New England, and the Corn Belt. Several other studies (Lins and Slack 1999; Douglas et al. 2000) found that these increases were due to increasing low and moderate flows, not high flows. McCabe and Wolock (2002) reinforced these studies, documenting a dramatic national increase in median and minimum flows after the mid-1970s. These studies together suggested that the hydrology of the US was becoming more benign, with low flows becoming higher and high flows staying the same, despite the skyrocketing costs of flood damages (Pielke and Downton 1999). Groisman et al. (2001) asserted, contrary to other studies, that heavy precipitation events did increase and the increases in high streamflows were detectable when one regionalized the data, as opposed to doing a site-by-site analysis. Specifically in the Western US, however, Groisman et al. (2001) asserted that there were no trends in streamflow volumes because less extensive snow cover was offsetting heavier precipitation. None of the above authors detected widespread significant trends in Western US streamflow; significant trends have mostly been observed in the eastern two thirds of the US.

No previous study, however, has investigated the trends in streamflow variability and persistence. Long term changes in the mean may have only subtle societal and environmental impacts, but changes in the magnitude and sequencing of extreme events could have direct impacts on ecosystems and natural resource managers (e.g., Voortman 1998). The hydrologic community has addressed streamflow variability and persistence, but mostly in the context of developing statistical forecasting models and defining hydrologically homogeneous regions (Vogel et al. 1998). These studies assumed that streamflow persistence was caused by carryover storage of water in lakes and below the land surface and that precipitation, in general, is random, stationary, and lacks persistence.

Conventional wisdom among the operational community is that forecast skill decreases as streamflow variability increases; Shafer and Huddleston (1984) indicated that streamflows were becoming more variable, and this masked improvements in average forecast error. A “wild” target is harder to “hit”. If streamflow is, indeed, becoming more variable, this might be responsible for the decline in skill of the official forecasts. In addition, statistical forecast procedures work best under typical basin conditions; when many consecutive dry years create unprecedented basin moisture deficits, forecast skill may suffer. Therefore, the next section tests whether streamflow variability and persistence are changing. If it is, then following sections will address whether water supply forecast trends identified in the previous section are related to changes in streamflow variability and persistence.

8.5 Observed trends in streamflow variability and persistence

As mentioned in chapter 2.6 Slack and Landwehr (1992) identified a subset of “Hydro-Climatic Data Network” (HCDN) streamgages as being free of significant human influences and therefore appropriate for climate studies. In the continental Western US, there are 475 such points west of 104.5 west longitude, excluding Alaska and Hawaii. Of these HCDN locations, a subset of 141 still-active gages with 50 or more years of data was chosen. HCDN sites with “constant” yet significant irrigation withdrawals or regulation, as indicated by the HCDN metadata, were removed from this analysis.

Monthly streamflow data were aggregated into April-September flow volumes. This period corresponds to the snowmelt and irrigation season across most of the interior Western US. For many locations this is also the target season of the NRCS water supply outlooks. Summer flows correlate very highly with annual flows in the interior Western US. For example, almost 90% of the annual flow on the East River at Almont, Colorado occurs during April-September. However, this period does not correspond to the primary snowmelt period in Arizona and southern New Mexico (December-March) and instead reflects baseflow conditions and monsoon-driven variability. Similarly, the Cascade Mountains (e.g. Oregon) can experience rain on snow and mid-winter melt before April. Analysis of annual flows in those regions may yield different results than those shown here.

At each streamgage, the variance, lag-1 year autocorrelation, coefficient of variation (CV, the standard deviation divided by the mean), mean, and skewness of the April-September flow volumes were computed for a 20-year moving window over the period 1901-2002. Data had to be serially complete within the 20-year moving window for these statistics to be computed; after 1940, at least 100 of the 141 available gages had complete data during any given 20-year period. There was a sharp rise in data availability beginning in the 1930s. Although pre-1930 data were included in the computation of the various long-term moments (i.e. mean, variance, and skewness) mentioned below, the results during this early part of the record are highly sensitive to the clustering of sites in Idaho and Montana and do not reflect the behavior of the Western US as a whole. Therefore, this early period is not focused on in this study. Future researchers may fill the gaps in coverage and extend the period of record by investigating additional indicators that reflect hydrologic variability, such as tree rings.

The 20-year time frame allowed a large enough sample size to develop reliable inter-period estimates of the variance and persistence (although a longer period would be more suitable for reliably estimating higher order moments such as skewness). It also created enough moving-window periods to observe any decadal variability or trends. This paper analyzes annual values over 20-year periods and therefore cannot provide any information about inter- or intra-seasonal variability and persistence. Future results could be generalized across all timescales using wavelet analysis, which investigates the changes in the power spectrum of data versus time (Torrence and Compo 1998). For example, Cahill (2002) used wavelets to describe the long term increase in short term (< 2 week) streamflow variability across the US.

The variance for each 20-year window was expressed as a ratio, relative to the variance of the period of record. If this ratio was greater than one, the period's streamflow was more variable than usual, and if the ratio was less than one, the period's streamflow was less variable than usual. The 20-year moving window lag-1 year autocorrelation was used to identify trends in persistence. Negative autocorrelation means that wet years tend to be followed by dry years and vice versa ("anti-persistent") whereas positive autocorrelation indicates a tendency for consecutive dry and wet years ("persistent"). An increase in persistence implies a shift from high frequency to low frequency variability.

To support this analysis, the CV, mean and skewness for each 20-year moving window were also computed. The CV is of interest in that it accounts for changes in variability associated with changes in the mean (i.e., drier periods tend to have lower variability). Each station's skewness was expressed as an anomaly relative to the period of record skewness. Most basins in the Western US have positive skewness, indicating a tendency for lower streamflow years to outnumber higher streamflow years and for a few unusually large events to have a major influence on the shape of the distribution.

One objective of this study was to determine if the variability during any given period was different from the variability of the period of record. The null hypothesis could be evaluated using an F-test. This test, however, assumes that the data are independent and follow a Gaussian distribution, the second being a poor assumption for skewed flows in the semi-arid Western US.

Instead, the significance of the change in variability was evaluated empirically. For each site, 20 (not necessarily consecutive) years of available data from the period of record were selected at random without replacement. The ratio of the 20 random years' variance to the variance of the period of record was computed. This resampling was repeated 10 000 times (200 000 random selections of years per site) to obtain an empirical distribution of the variance ratio for each location. This size resampling pool was chosen because it allows a reliable estimate of the 10% significance level without being computationally prohibitively expensive. The observed variance ratio for each 20-year period was compared to the variance ratios of the 10 000 synthetic periods to determine the probability of obtaining the observed ratio by chance. This same jackknife procedure was repeated for the CV, mean, and skewness.

This sampling technique, which drew randomly from the entire period of record, may have underestimated the statistical significance of the result compared to a technique that only drew from years outside the 20-year window (e.g., an analysis of 1950-1969 flows that only randomly selected years before 1950 and after 1969). This more appropriate technique, however, would have involved almost two orders of magnitude more analysis than the selected technique and thus was computationally impractical. The selected technique violates the assumption of independent samples in that it allows data to be compared to itself, although the effect on the results is likely to be small.

Figure 8.8a shows a time series of the fraction of available stations reporting periods of significantly increased variability (solid) and decreased variability (dashed). Significance was determined by comparing each 20-year moving window variance ratio to the empirically derived distribution and identifying those values that were greater than the 90th percentile (increased variability) or less than the 10th percentile (decreased variability). The period 1945-1964 was the most geographically widespread period of low variability in modern history, with 48% of sites reporting significantly below-average variability. During this period, none of the 104 reporting sites in the Western US had significantly above-average variability. The variability decrease was most pronounced in Idaho and Montana, with decreases also in the Cascades, central California, the Great Basin, and the Southwest (see also Figure 8.9, upper left). Increased variability marked the period after the mid-1960s, when 29% and 27% of sites reported significantly above-average variability in 1982-2001 and 1971-1990, respectively. This variability increase was focused primarily in California, the Great Basin, and northwestern Colorado (Figure 8.9, lower left). From 1976-1995, there was also a small rise in the number of sites that had less variability, primarily confined to the Cascade Mountains of Washington and Oregon.

Figure 8.8b shows a time series of the fraction of available stations whose lag-1 autocorrelation was greater than (solid) or less than (dashed) +/- 0.30 (approximately 10% significance for a one-tailed test) in a 20-year moving window. This time series shows that in 1936-1955, 26% of sites had high year-to-year persistence. By 1959-1978, 33% of sites had highly negative autocorrelation, a tendency for wet years to be followed by dry years and vice versa. In the most recent twenty years, 28% of the sites had high year-to-year persistence, the most widespread of any period of history (Figure 8.9, lower right).

One might expect that the increase in variability in the recent period was associated with increases in mean flows (e.g., the 1950s drought was a period of low variance). Especially where flows are skewed, one or several exceptionally wet years can significantly raise both the 20-year mean and variance. As mentioned in the introduction, few authors have been able to detect trends in mean flow in the Western US. Additionally, in the analysis performed here, statistically significant changes in the CV were very well correlated with changes in the variance, suggesting that the changes in the mean had a secondary influence on these results, if at all. Figure 8.8c shows a time series of the fraction of available stations reporting periods of significantly increased mean flows (solid) and decreased mean flows (dashed). Generally, periods of high flow (1940s-1970s) and low flow (1930s-1950s and 1980s-present) do not necessarily match periods of high or low variance. Additionally, the spatial patterns of mean and variance trends do not match (compare figure 8.9 left and figure 8.10 left).

Finally, figures 8.8d and 8.10 (right), show the behavior of streamflow skewness over time. The late 1940s to early 1970s was a period of relatively high streamflow skewness, followed by lower skewness from the 1960s-1980s and a return to higher skewness in recent times, especially in northern states. The most recent 20-year period had the highest percentage of stations reporting a positive skew anomaly (62%) of any period in history while the 1978-1997 period had the most number of sites with statistically significant skewness increases (22%). Therefore, there has been an increasing tendency towards having many low flow years, punctuated by a limited number of relatively high flow years. It is very difficult to develop reliable estimates of higher order moments (e.g., skewness) with a limited set of data (e.g., 20 years) so these findings should be interpreted with caution.

The results shown in Figure 8.8 were somewhat influenced by the spatial distribution of sites used in this analysis, with a high density of sites in Idaho, Montana, and the Pacific Northwest and a sparse network of sites in Nevada and the Southwest. There was also a very strong cross-correlation of flows across sites; although the spatial correlation varies regionally, in the interior west, sites 650 kilometers apart at the same latitude generally have interannual flow correlations of 0.7. This correlation is less for sites in the same longitude band, consistent with Cayan’s (1996) investigation of snowpack. Future research should investigate the field significance of these results, perhaps by using cluster or principal components analysis.

It is also possible that autocorrelation, variance, skewness and the mean are interrelated and that some configurations (e.g., high autocorrelation and low variance) may be naturally easier to achieve than others (e.g., high autocorrelation and high variance). Each location's time series of 20-year moving window autocorrelation, variance, skewness, and mean were correlated and the interannual relationships between the four characteristics investigated. Although individual sites had large magnitude correlations among the different characteristics (e.g., autocorrelation and variance at a location may have tended to vary together), there was no obvious spatial pattern or regional coherence to these correlations. At best, there was a weak negative correlation between mean and skewness (57% of the sites had a correlation between mean and skewness less than -0.3) and a weak positive correlation between mean and variance (48% had a correlation greater than 0.3). The westwide median correlation between autocorrelation and variance was 0.055, far from being statistically significant. This site-based analysis suggests that it is likely that streamflow variance and persistence change independently of one another. It is unknown whether a more sophisticated analysis involving regionalization of the data (e.g., Groisman et al. 2001) would reveal clusters of sites whose variance and persistence vary together as a region.

As a representative example of a shift from low variability and anti-persistence to high persistence and high variability, Figure 8.11 displays a time series of flows in northwestern Colorado. Although individual years may vary, the general pattern of flows is typical of what was experienced throughout Colorado, Utah, and Nevada. The period 1935-1975 was relatively benign with many years of consistently near-normal flows especially from the mid 1960s-1970s. The series of four consecutive very wet years (1983-1986) was unprecedented, as was the extended deficit that followed (1987-1994). Again, a wet period ensued in 1995-1998 followed by near-record drought conditions in 2002. The period after 1980 displays very high variability and year-to-year persistence whereas prior decades were less variable and less persistent.

8.6 The implications of increased streamflow variability

Decadal timescale changes in streamflow variability and autocorrelation have been observed in the streamflow records of the Western US. The 1930s-1950s can be described as a period of low variability and high persistence, the 1950s-1970s as a period of low variability and anti-persistence, and the period after 1980 as high variability and high persistence.

These various streamflow characteristics were not necessarily varying on the same time scales or coincidentally; increased variability preceded increased autocorrelation by approximately 5-10 years, which in turn preceded increased skewness by another five years. Nonetheless, the various phenomena became “in phase”, making the most recent 20 years the only part of the record that was highly variable, highly persistent, and highly skewed.

A period of high persistence, variance, and skewness is perhaps the most challenging scenario for water managers. For example, a series of consecutive wet years (e.g. figure 8.11, 1983-1986) may overwhelm reservoirs and inflate stakeholder expectations about the amount of water available. An extended stretch of dry years then exhausts storage reservoirs and does not give them a chance to recover (e.g. 1987-1994). Smaller reservoirs that do not have multiple year storage capacity would be especially vulnerable. In comparison, individual dry years interspersed among wet years are much more tolerable.

These decadal oscillations also have implications for water supply forecasting. Statistical streamflow forecasting techniques that use persisted spring and summer streamflow as a predictive variable for next year's flows will lead the forecaster astray when the climate regime switches between positive and negative autocorrelation. It is unknown at this time whether procedures that use antecedent autumn streamflow (e.g., September-November) as a predictive variable to index the effects of soil moisture and groundwater variability are also vulnerable to this effect.

The causes of the recent sequence of flows are unknown. The HCDN stations are free of significant human influence, and the signals detected in this study were geographically widespread, ruling out changes in basin characteristics, soil properties, or local management practices as the causes of changes in variability. It is likely that the changes in persistence and variability have climatic origins.

If the driving factor in this recent variability increase is discovered, it may be a source for long-range decadal climate forecasts. Recently, McCabe et al. (2004)

explained almost three-quarters of the spatial and temporal low-frequency variability of drought across the US using the PDO, Atlantic Multidecadal Oscillation (AMO), and a unidirectional trend possibly due to climate warming. Those authors examined the frequency of average annual precipitation falling below a particular threshold and did not investigate variability or persistence. Nonetheless, future research is necessary to determine if the phenomena are related and if streamflow variability and persistence will continue to increase.

The current region of high streamflow variability (California, Nevada, Utah, and Colorado) is also known for its low seasonal precipitation predictability (Hartmann et al. 2002a). In contrast, the El Niño/Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO) typically affect Oregon, Washington, and Idaho as well as Arizona and New Mexico (Redmond and Koch 1991; Mantua et al. 1997). The mechanism driving increased variability could be masking otherwise useful ENSO and PDO-related climate forecast skill in this highly unpredictable region.

Interestingly, although streamflow variability is on the rise and forecast skill is on the decline across the Western US, the spatial patterns of these changes do not match. Comparing figures 8.9 (bottom left) and 8.5-8.6 (lower right), streamflow variability is on the rise throughout the Great Basin, southern Columbia and Upper Colorado River Basins. Variability is decreasing for rivers in Oregon. However, forecast skill is declining in Oregon, remaining steady in California and plunging in the eastern Colorado River Basin and Arizona/New Mexico (this last region not displaying any significant trends in streamflow variability). Increased variability is not sufficient to explain the downward trend in forecast skill, especially when one considers that the NS score is already normalized by the variability of the observations (as mentioned in chapter 6.3.1).

8.7 Is spring precipitation becoming more extreme?

Tracing the changes in streamflow variability back to changes in precipitation and temperature may be difficult because of the temporal and spatial integrative behavior of watersheds. The climate signal may be spread across seasons such that, for example, winter (December-March) precipitation variability in California may be rising while spring (April-June) precipitation variability is rising in Colorado, resulting in an April-September streamflow variability increase in both locations. Coincidentally, the seasonality of the increased precipitation variability has serious implications for water supply forecast accuracy. Given that water supply forecasts are primarily based on existing initial conditions (e.g., current snowpack), if extreme (high or low) precipitation occurs before the forecast issue date (e.g., 1 April), the forecast is more accurate than if the extreme event occurs after the forecast issue date.

As shown in chapter 6.3.3, specifically figure 6.4, the improvements in forecast skill versus leadtime are closely tied to the seasonality of precipitation. If, climatologically, the springtime is a dry period, the 1 April water supply outlooks typically perform better than those in parts of the country where springtime is, climatologically, the relatively wet period. Church (1935) also identified precipitation after the forecast issue date as the primary source of forecast error in any given year. It stands to reason that a decline in 1 April forecast skill could be related to a change in the behavior of precipitation that occurs after 1 April. Extreme wet or dry springs can cause large disagreements between a regression-based forecast which assumes “near normal” conditions for the remainder of the season.

Using the weightings described in Table 7.1, and the methodology described in chapter 7.2, basin wide Z-score indices of spring/summer precipitation are developed. Spring/summer precipitation is defined as the precipitation that occurs between 1 April and the end of the streamflow target season. For example, in Arizona where the 1 April forecast target season is April-May, precipitation is averaged over April-May. If the target is April-September, an index is developed for April-September precipitation. If the target is March-July, the index remains April-July. See Table 2.4 for a listing of forecast target seasons. When aggregating the individual Z-scores into a basin-wide index, no site weightings are used (such as those mentioned in chapter 7.2), i.e., all sites are considered equally important.

Figure 8.12 shows a scatter diagram of the Weber UT spring/summer precipitation index and 1 April forecast error. Wet spring conditions, as indicated by a large positive basin-wide spring precipitation index, tend to bring more streamflow than the forecasts expected. Dry spring conditions (large negative spring precipitation index values) cause the forecasts to overestimate the amount of future streamflow. Spring precipitation variability explains more than 60% of the error in the 1 April forecasts at this location. The remaining 40% of the error variance could be due to a number of factors including soil moisture deficits, sublimation, unaccounted-for streamflow impoundments and diversions, and so on.

Figure 8.13 shows the map of the correlation coefficient between the combined snow+precipitation hindcast 1 April error and a basinwide Z-score index of spring/summer precipitation. Error is defined as observed minus forecast, as opposed to elsewhere in this dissertation where error meant forecast minus observed. This was done to give a positive orientation between forecast error and spring precipitation. This figure shows that throughout almost all of the Western US except the northern Cascades, the relationship between spring precipitation and 1 April forecast error is exceptionally strong. Most correlations are in the range of 0.5-0.65, consistent with Schaake and Peck's (1985) finding that spring precipitation variability accounted for 50% of the error of the historical 1 April forecasts for inflow to Lake Powell.

It is now established that streamflow forecast skill is on the decline (chapter 8.3), and that extreme spring precipitation events have a negative effect on streamflow forecasting skill. The question can now be asked: Are there observed trends in the extremity of spring precipitation, responsible for the decline in forecast skill?

To test this, the Z-scores of individual sites are combined into a basin-wide average of spring precipitation "irregularity" as follows. Each site has a spring precipitation Z-score for each year, as calculated earlier. This value is then squared and all of the squared Z-scores for all sites within a basin are averaged together to form a single time series for each of the 29 study basins. The squared Z-score is a measure of the variance relative to a period of record mean as opposed to local variance that is relative to a local mean. The distinction is important because if every year within a 20-year moving window was identical but very wet, the variance would be zero while the squared Z-score would better capture the extremity of the period. Basin-wide spring precipitation departures from near normal (i.e., very high or very low) will give this index a high positive value. If spring precipitation is near normal, this index will be close to zero. As the original Z-score index has a standard deviation equal to 1.0, the squared index has an expected value of 1.0. In other words, a spring precipitation irregularity index value of 1.0 indicates "typical" (that is, neither calm nor extreme) precipitation irregularity. In following sections, this squared Z-score index is referred to interchangeably as spring precipitation "extremeness" and "irregularity".

The average of the spring precipitation irregularity index is then taken over a 20-year moving window for the period of record. It would have been possible to calculate the spring precipitation variance directly rather than the average of the squared Z-scores. However, this second-order moment would not be able to detect changes in the first order moment. In other words, if every year within a 20-year period had extremely high precipitation (high, positive Z-score), the variance of this time series would be low, as would a 20-year period with consistently near-normal spring precipitation. However, one would expect the water supply forecasts to perform poorly under the first scenario compared to the second. The squaring of the Z-score is therefore a better measure of the "abnormality" or "irregularity" of spring precipitation and will be more relevant to the skill of the water supply forecasts.

The Pecos River near Pecos has experienced a very sharp drop in forecast skill in recent years. Figure 8.14 (upper left) shows a time series of the 20-year moving window precipitation "irregularity" index just described for the Pecos basin. Note the monotonic trend towards increasingly "near-normal" spring conditions from 1930 to the early 1980s, after which there has been a dramatic rise in the frequency of extreme spring events. Spring precipitation on the White CO (upper right) is similarly the most extreme it has ever been in the modern record. Other locations, such as the Bruneau, ID (lower left) show decadal variability in spring precipitation irregularity, almost out of phase with that of nearby Umatilla, OR (lower right).

Figure 8.14 displays the westwide average of the 20-year moving window spring precipitation irregularity indices of all of the 29 study basins. Similar to the Pecos, spring precipitation had, on balance across the Western US, become increasingly calm in the early part of the record, reaching its most placid and benign state in 1969-1988. In less than a generation, westwide spring precipitation irregularity has skyrocketed and is now more extreme than any other time in modern history.

Figure 8.15 shows a map of the spring precipitation irregularity index anomaly (deviations from 1.0) for four 20-year periods. Filled inner circles mean that the average spring irregularity index for this 20-year period is greater than 1.0. Hollow circles indicate an index less than 1.0. Large circles indicate either highly irregular or exceptionally calm, depending on whether the circle is filled or hollow, respectively. In 1961-1980 (lower left), the entire Western US except Southern Idaho and Northeastern NV had very reliably near-normal spring precipitation, particularly in the Colorado basin, the Southwest US and Cascades. After 1981, almost exactly mimicking the pattern in the decline of official forecast skill (figure 8.5-8.6 lower right), spring precipitation has been highly irregular in the upper and lower Colorado basins and in the Pacific Northwest.

8.8 Summary

The overall picture that emerges from the findings of Chapter 8 is as follows: seasonal streamflow variability is on the rise in the central Western US, from California through Nevada, Southern Idaho, Utah, Colorado and northern New Mexico. Forecast skill is declining in the Colorado and Rio Grande Basins, although it is rising in California and Nevada. Therefore the increase in streamflow variability cannot entirely explain the trends in forecast skill otherwise forecast skill would be declining in all of the regions just mentioned. Additionally, skill would not be declining, as it is, in regions where streamflow variability is not on the rise (i.e., Arizona and New Mexico).

Although this study did not test it, it is likely that Californian fall and/or winter precipitation extremity is now higher than it was before whereas spring precipitation extremity is diminishing. By 1 April, the Californian hydrologist was aware of the anomalous snowpack condition and as a result issued extreme forecasts that performed well compared to a baseline forecast of near-normal.

In contrast, the forecasters for the Colorado and Rio Grande basins believed they had an accurate assessment of conditions by 1 April, but extreme precipitation events commonly occurred **after** the forecasts were issued. The resulting turnabouts (or intensifications) damaged the hydrologists' credibility, because forecasts increasingly failed to match the observations.

This study cannot explain causes for the rise in extreme spring precipitation situations but can only recognize the existence of a trend. Water supply forecasters would be keenly interested in knowing if the variability in streamflow and spring precipitation will subside (indicating an oscillation), remain high (a secular change) or push even higher (due to greenhouse warming).

The non-stationarity of climate throws into question the validity of forecast skill as a metric for the performance of the agencies or individuals involved with water supply forecasting. For example, the NRCS has 5-year quantitative goals to decrease the error of its water supply forecasts, as well as goals to increase in the numbers of forecasts produced per year. While forecast technology can make minor improvements, it has been shown here that overall water supply forecast skill is almost completely dominated by interannual climate variability, especially variability in precipitation that occurs after the forecast issue date. Perhaps the agencies can develop climate-adjusted skill scores (such as the improvement over the simple objective hindcast system used in this paper). Users and water managers, however, only care about the magnitude of forecast errors, not their underlying causes, and may not be as forgiving.

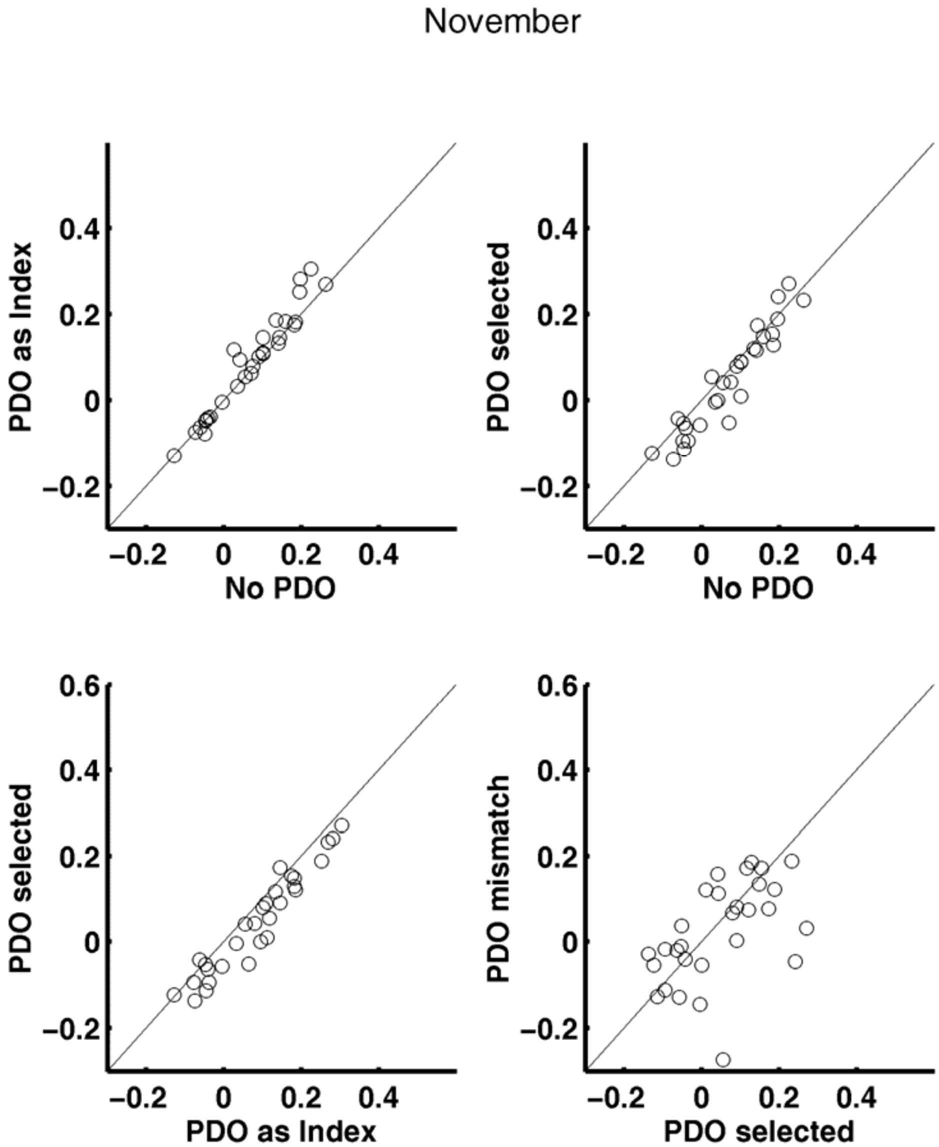


Figure 8.1 Scatter diagrams of period of record NS scores for hindcasts issued 1 November using a variety of configurations. All hindcasts are developed using the synthetic snow+precipitation+climate system described in section 7.6. Configurations 1-4 are as follows: 1) Excluding the PDO index, 2) Including the PDO index, 3) Proper subsetting by PDO state, 4) Mismatched subsetting by PDO state. Subpanels above compare configurations 1,2 (top left), 1,3 (top right), 2,3 (lower left), and 3,4 (lower right). The diagonal line indicates equal skill. See text for discussion.

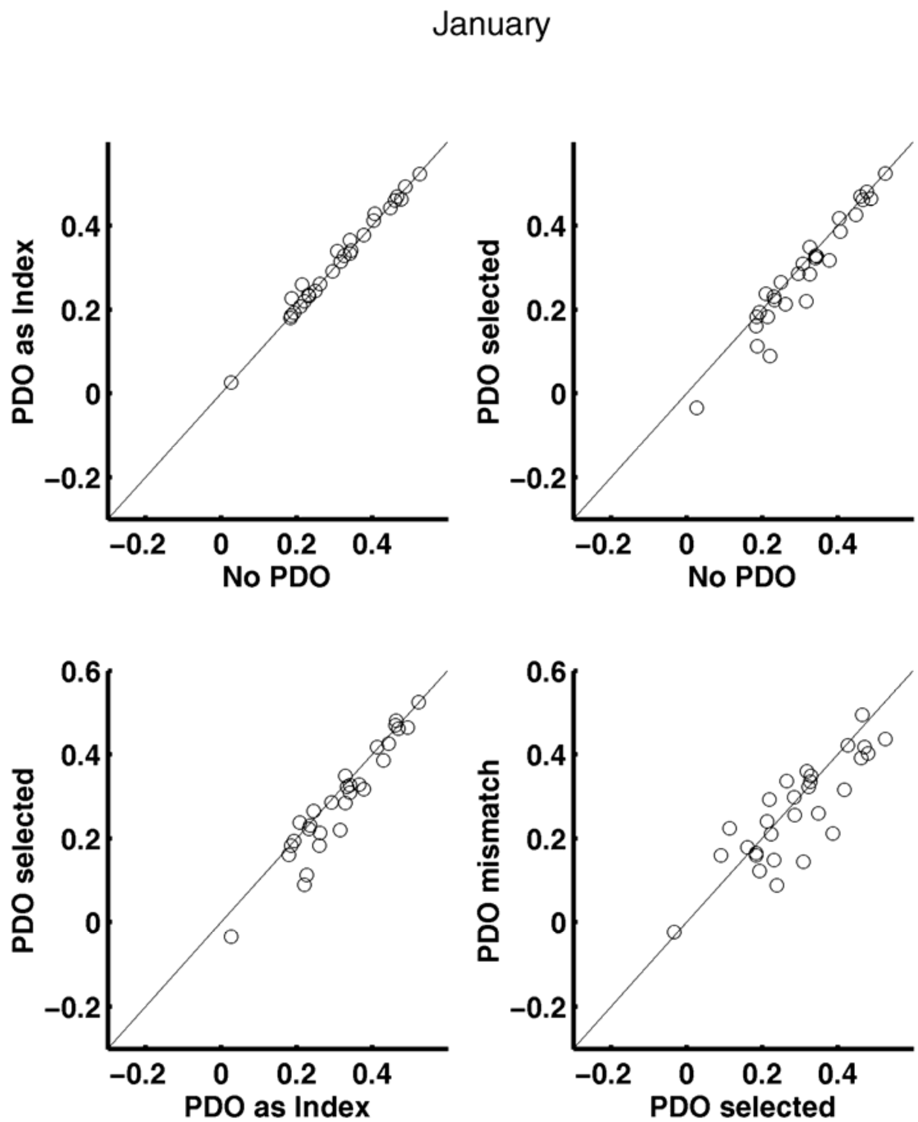


Figure 8.2. Same as figure 8.1 except for hindcasts issued 1 January.

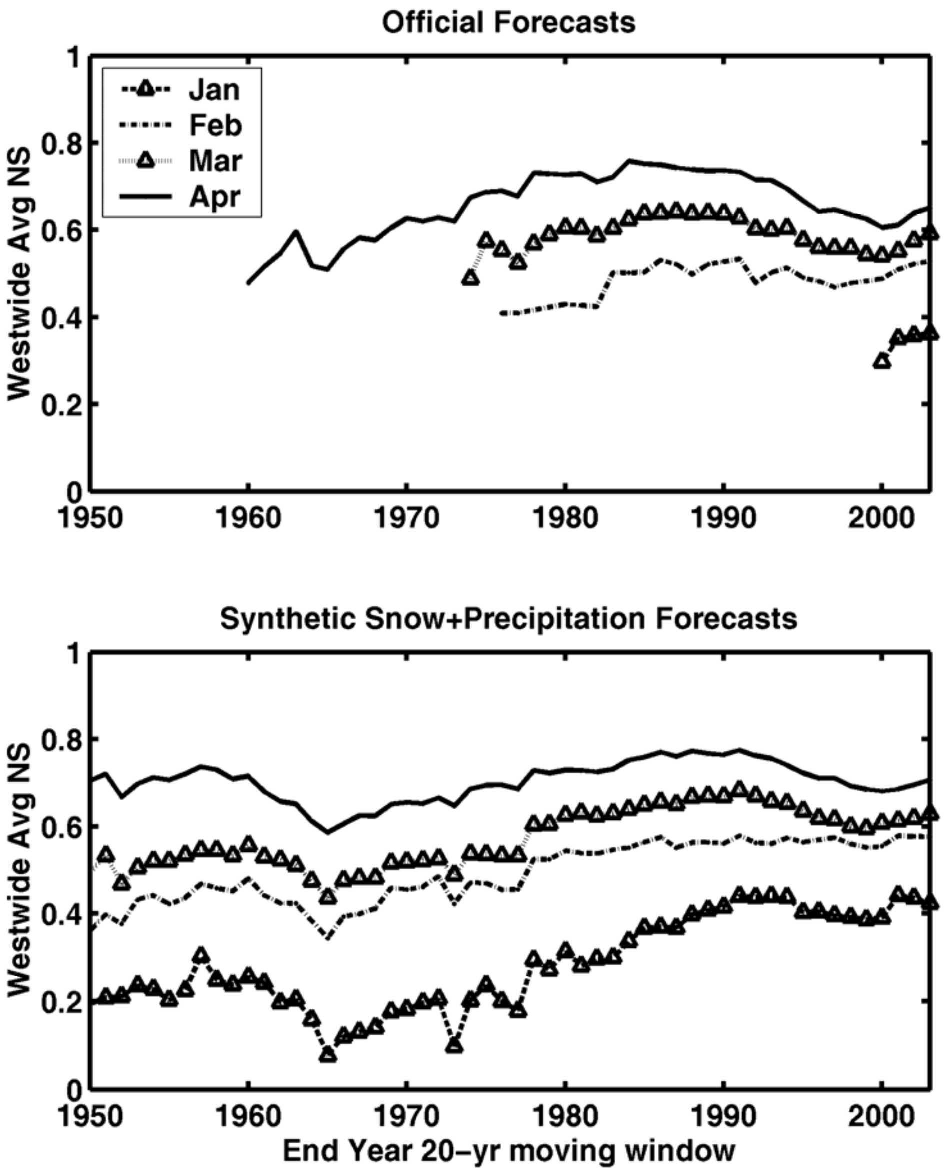


Figure 8.3. Trends in westwide official (top) and synthetic (combined snow+precipitation, bottom) forecast skill over 20 year moving windows, and by lead time. Data must be serially complete in 20-year window to be computed for a basin. At least 8 basins must have a valid NS for a value to be shown. Data is plotted at the end year of the 20-year moving window (i.e. 2000 = 1981-2000).

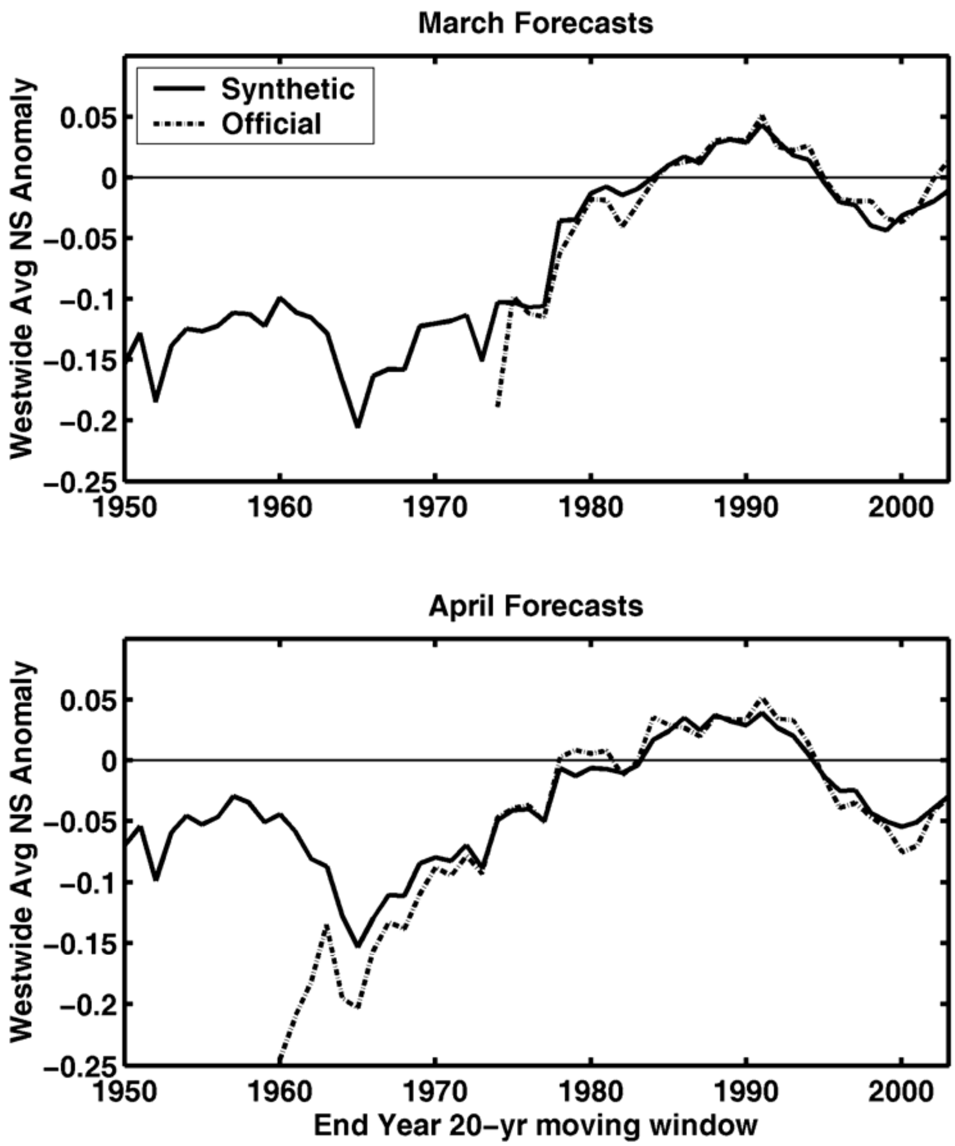


Figure 8.4 Trends in westwide official (dashed) and synthetic (solid) forecast skill anomalies over 20 year moving windows for March (top) and April (bottom) forecasts. Forecast skill anomalies are relative to the skill of each system during 1960-2000. Plot format follows that of figure 8.3.

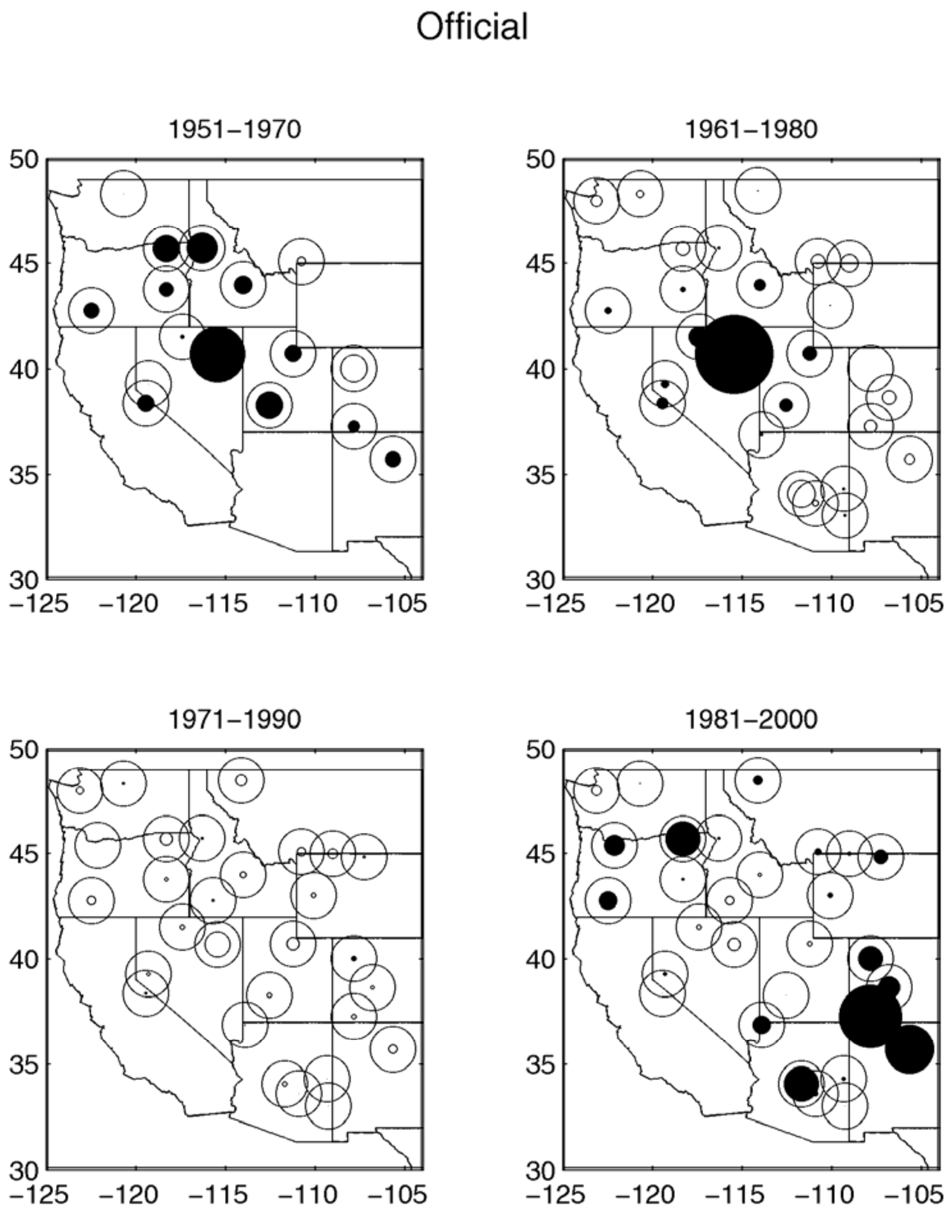


Figure 8.5 Skill anomalies of the official 1 April water supply forecasts for four 20-year periods, relative to 1961-2000. Filled circles indicate negative skill anomaly, empty circles indicate positive skill anomaly. Circle diameter is linearly proportional to anomaly magnitude. An outer reference circle of skill anomaly magnitude of 0.3 is provided. Missing outer circles (e.g. Arizona in 1951-1970) indicate insufficient forecasts available for analysis (i.e. the forecasts are not serially complete during this 20-year period).

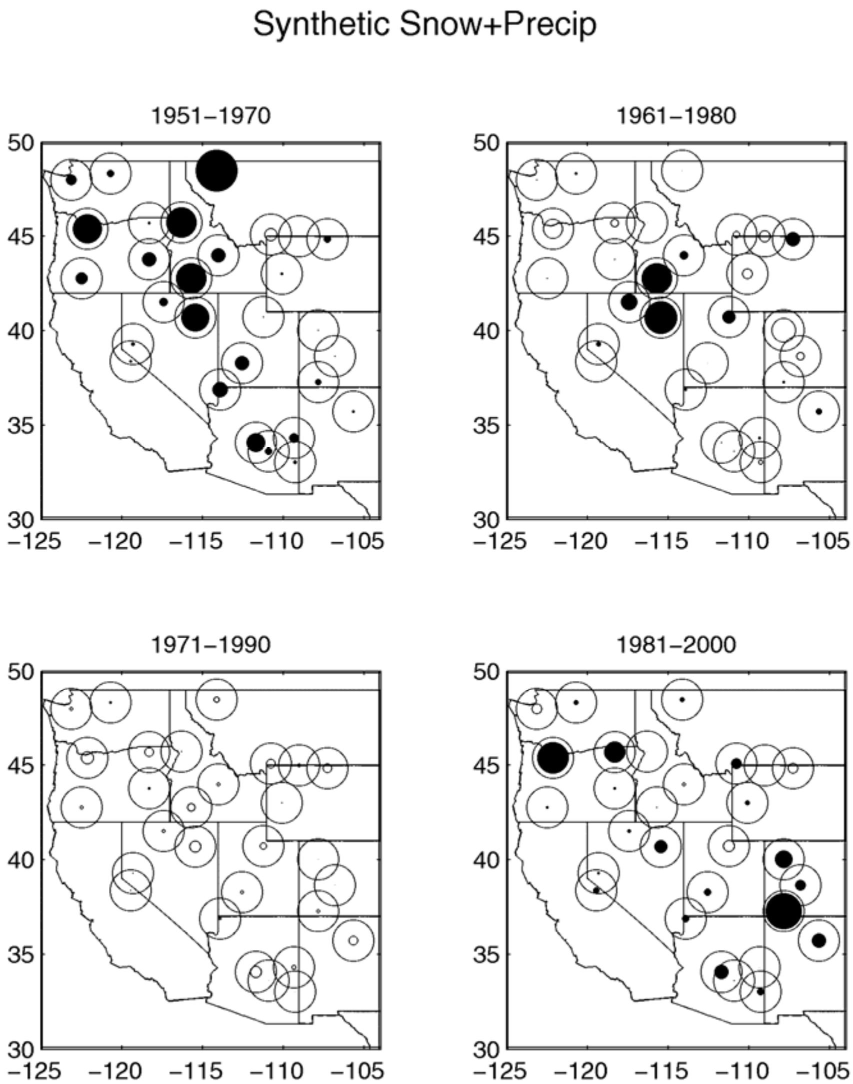


Figure 8.6. Same as figure 8.5 except for 1 April synthetic snow+precipitation synthetic hindcasts.

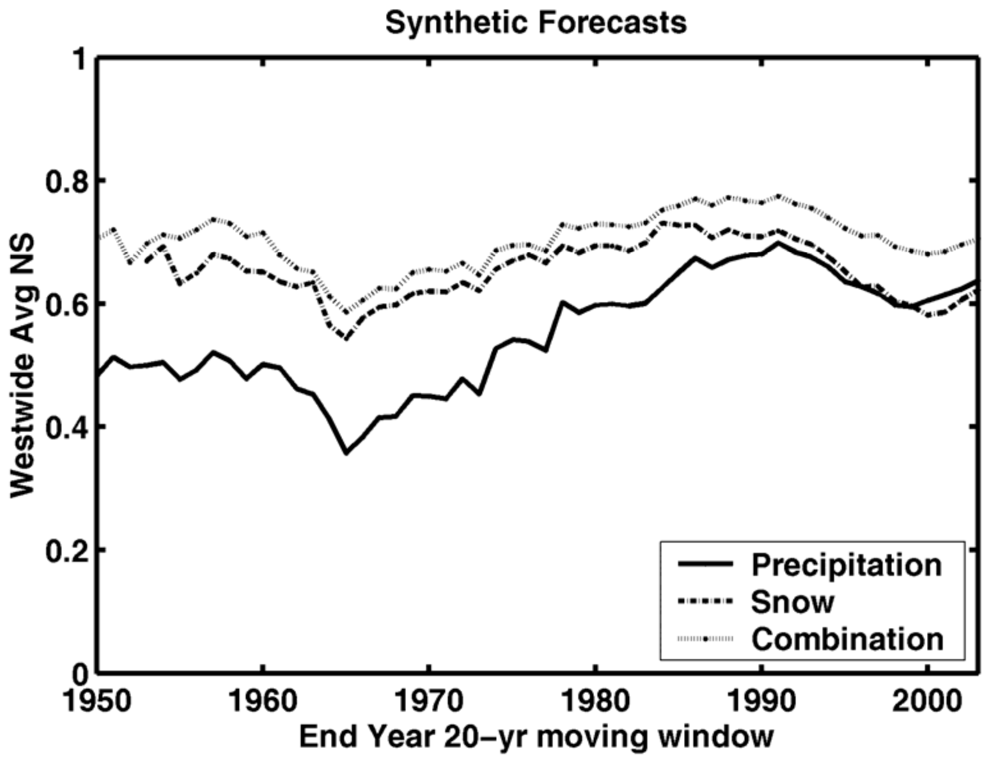


Figure 8.7. Time series plot of 1 April westwide average NS skill score during a 20-year moving window period. Skill is shown for the precipitation-based (solid) and the snow-based (dashed) synthetic hindcast systems. The performance of the hybrid precipitation+snow based hindcast system is shown by the top-most line.

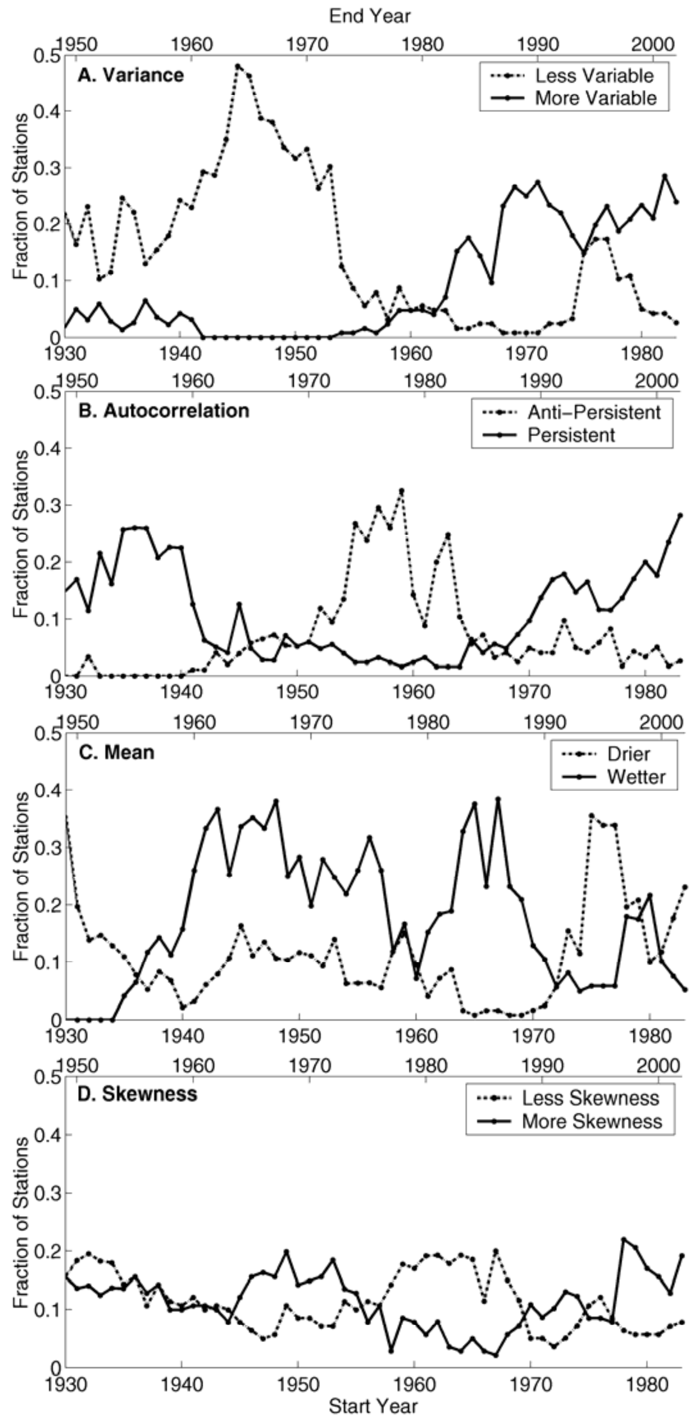


Figure 8.8. A: Time series of the fraction of western US streamflow stations reporting significant increases (solid) or decreases (dashed) in 20-year moving window variance compared to the period of record. Significance is defined as the 10^{th} and 90^{th} percentiles. B: Fraction of stations reporting interannual lag-1 year autocorrelation of greater than 0.3 (solid) or less than –0.3 (dashed). C. As panel A for the mean. D. As panel A for the skewness. The bottom X-axis indicates the start year of the 20-year moving window and the top X-axis indicates the end year.

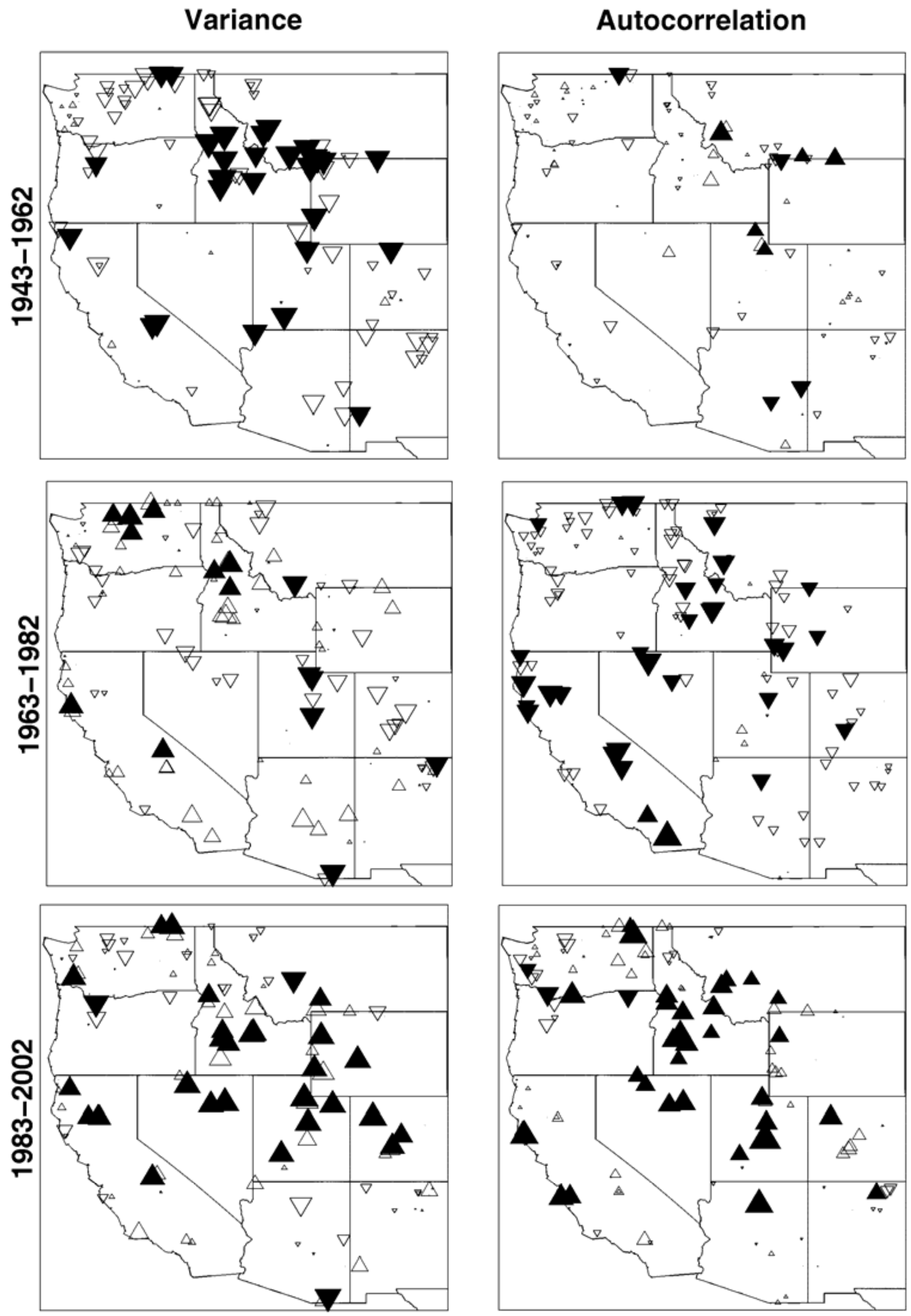


Figure 8.9 Maps of streamflow variance ratio significance (left) and autocorrelation (right) for three 20-year epochs (top, middle and bottom). Upward pointing triangles indicate positive autocorrelation or increased variance relative to the period of record. Downward pointing triangles indicate negative autocorrelation or decreased variance. Filled symbols indicate autocorrelation greater/less than +/-0.3 or statistically significant variance departures. The size of the symbol is proportional to the magnitude of the departure.

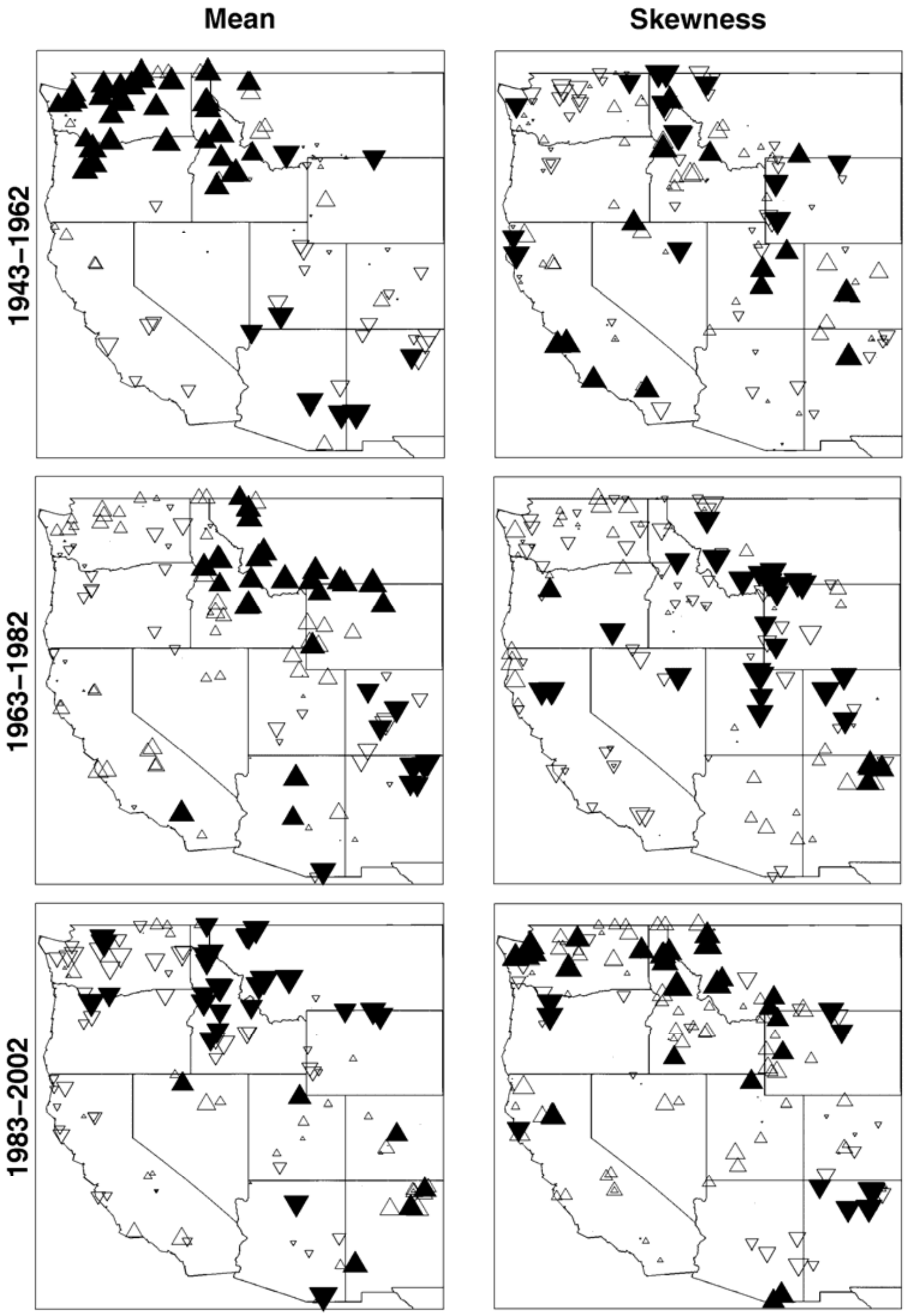


Figure 8.10 As figure 8.9 except for the mean significance (left) and skewness significance (right). See figure 8.9 for symbol definitions.

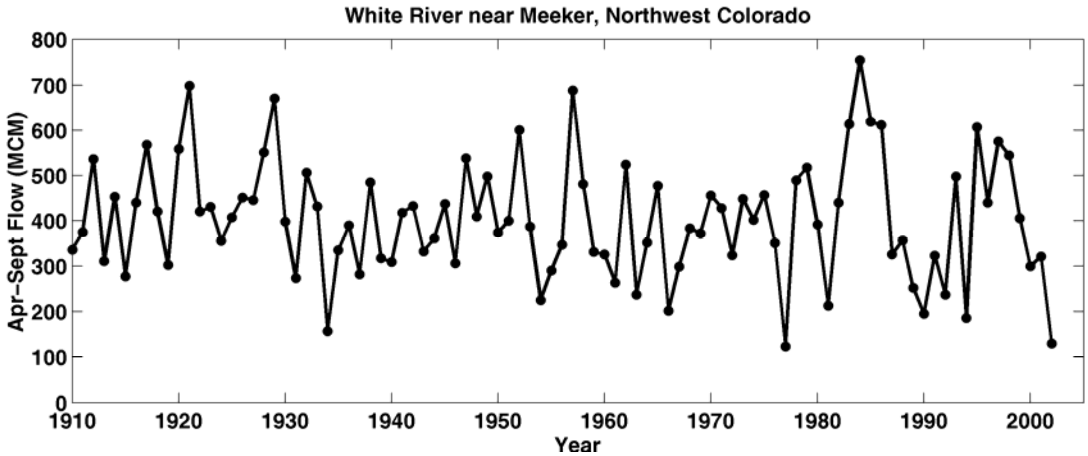


Figure 8.11 Time series of April-September flow at the White River near Meeker, CO (USGS ID 09304500) from 1910-2002. Units are millions of cubic meters (1.2335 MCM = 1 k-ac-ft)

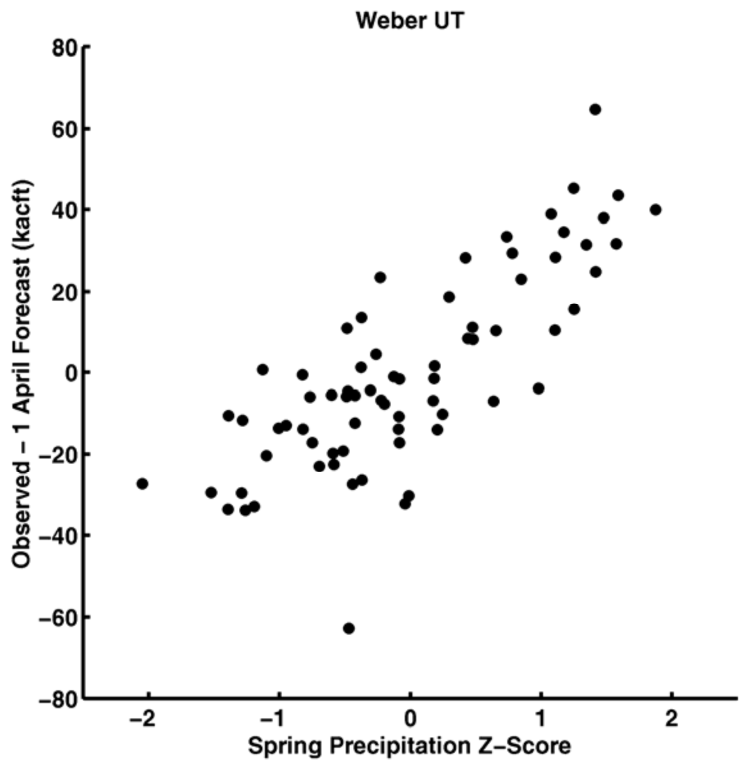


Figure 8.12. Representative scatter diagram of spring precipitation with 1 April forecast error (observed minus forecast) for the synthetic snow+precipitation hindcasts. See text for explanation of the spring precipitation index. Hindcasts are for the Weber River in Utah.

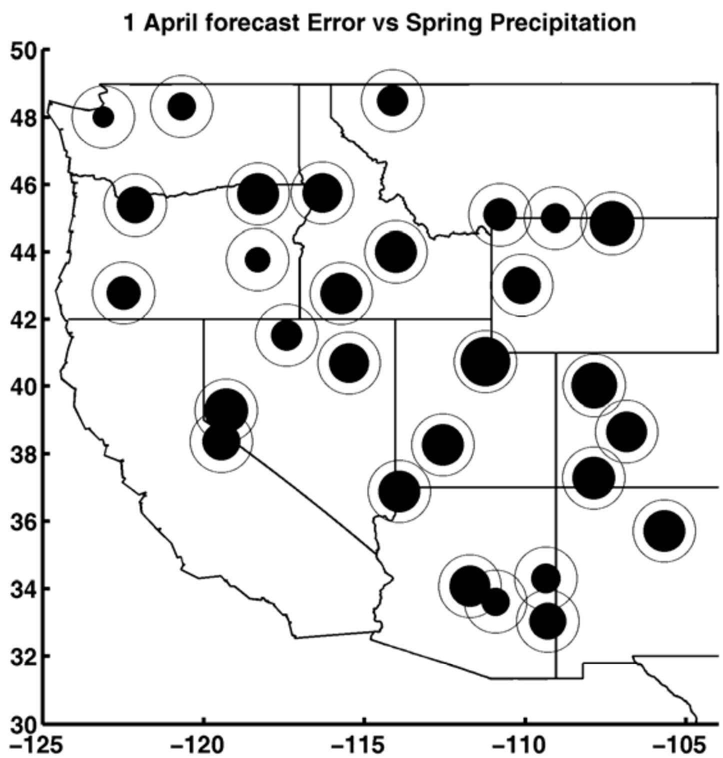


Figure 8.13. Map of correlation coefficients between 1 April synthetic forecast error and a spring precipitation index. Outer reference circle indicates correlation of 1.0. The majority of correlations fall in the range of 0.5-0.65. Map follows the convention of figure 7.5

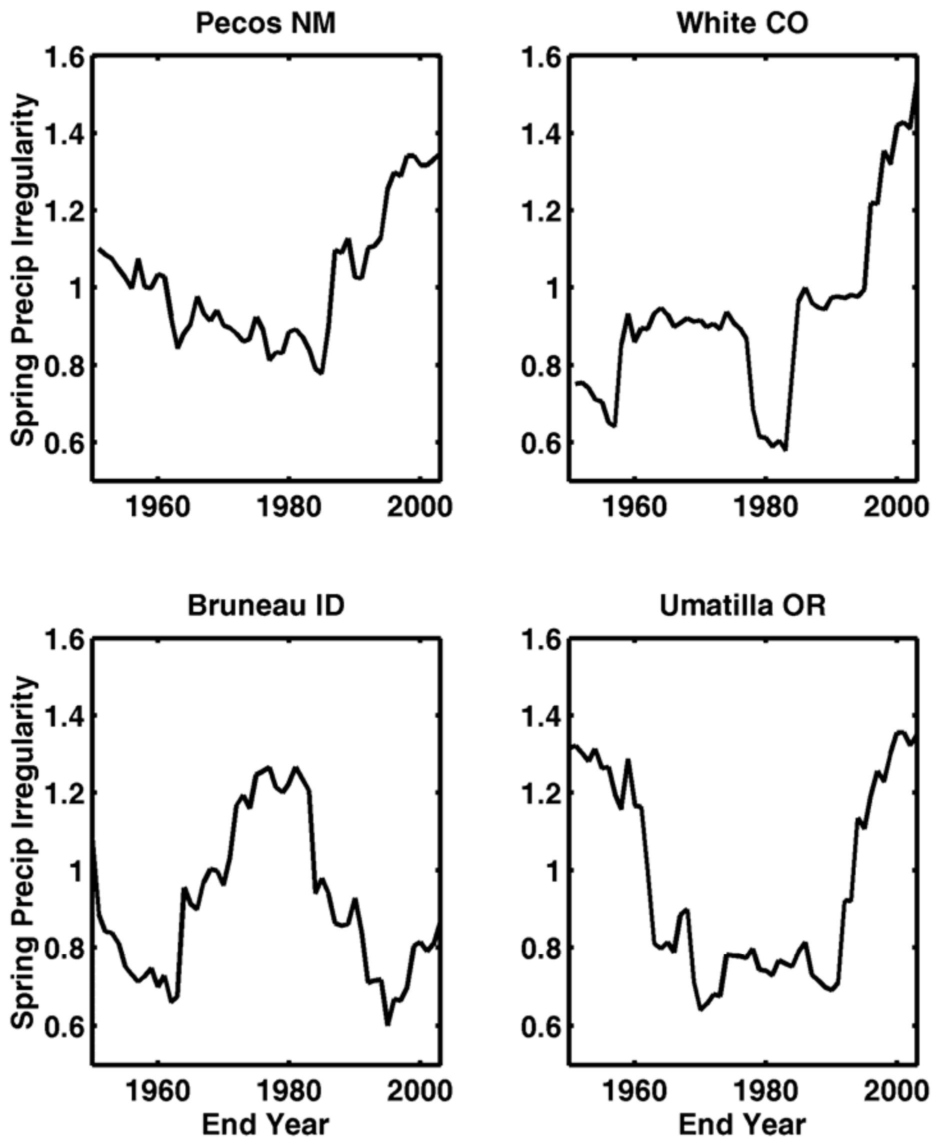


Figure 8.14 a Time series of the 20-year moving window average of a spring precipitation irregularity index (see text for definition) for four locations around the Western. On average, a value of 1.0 is considered normal variability. High values indicate unusual/extreme spring precipitation events (wet or dry) whereas low values indicate spring precipitation reliably near normal.

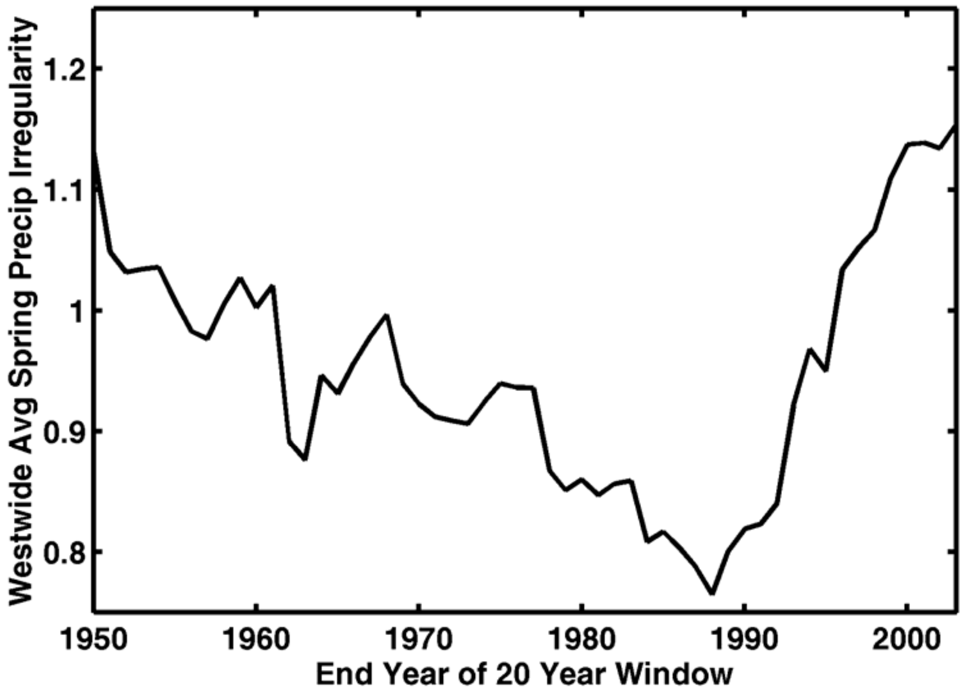


Figure 8.15. Westwide average of the 20-year moving window average spring precipitation irregularity index. Compare with figure 8.4 and 8.7. Spring precipitation was most “calm” in 1969-1988 and most “extreme” in 1983-2002.

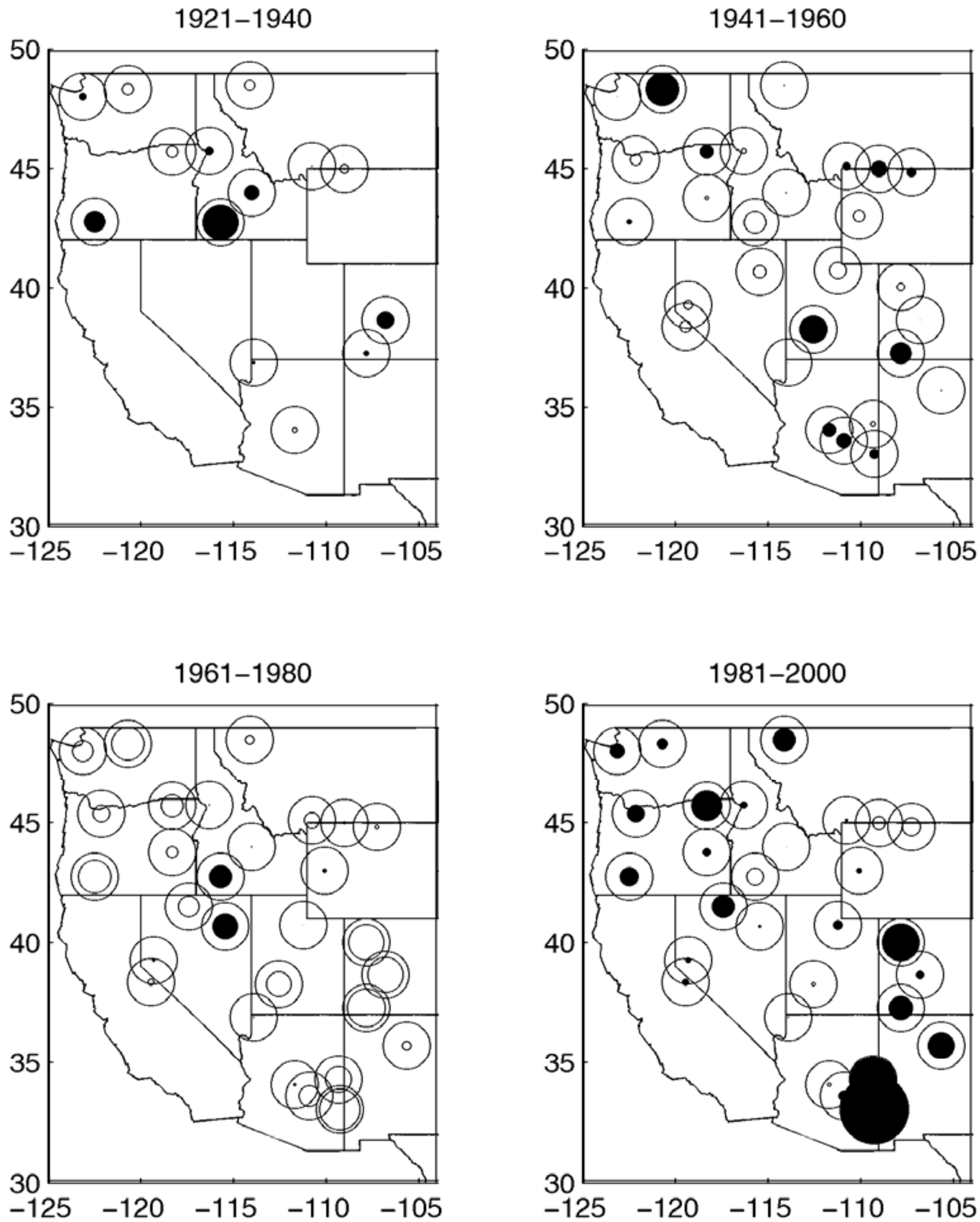


Figure 8.16. Spring precipitation irregularity for four 20-year periods. Filled circles indicate extreme spring conditions (SI>1.0), empty circles indicate calm spring conditions (SI<1.0). The outer circle is a reference circle for SI 0.45 or 1.55. This figure shows that spring conditions in the Southwest, Colorado and Pacific Northwest were exceptionally calm in 1961-1980, but in 1981-2000 extreme springs have occurred in the Colorado River Basin and Pacific Northwest Cascades. Compare panel 4 with forecast skill anomalies in figure 8.5-8.6

9. AN OPERATIONAL CLIMATE-BASED WATER SUPPLY OUTLOOK

9.1 Introduction

The previous section documented the very important role of climate variability in water supply forecast skill. Forecast skill is at the mercy of a sometimes placid but oftentimes erratic climate. The effect of human factors on forecast skill, such as the choice of forecast methodology or changes in data quality, are minor in comparison to the effects of the character of precipitation after the forecast issue date. Understanding and being able to predict future precipitation variability should be a key priority of water supply forecasting agencies. Indeed, as mentioned in chapter 5.2, hydrologists have been trying to take advantage of climate forecast skill for almost as long as climate forecasts have been issued.

This dissertation has recognized the existing scientific basis for climate forecasts and has measured the expected skill of a climate-based long-lead water supply outlook. Scientific understanding and technical expertise, however, are only some of the key ingredients to a successful operational product. Conceptually and technically, very little prevents the NRCS from issuing water supply outlooks with longer leadtimes than those currently issued. Perceptually, several barriers remain. This chapter describes some of the operational issues associated with the development of a new pre-season water supply outlook. It draws from Pagano and Garen (2005a) and its intended audience is operational water supply forecasters and forecast agencies.

9.2 “Climatologists are from Venus, Hydrologists are from Mars”

Although some aspects of forecasting are universal, climate based streamflow forecasting is very different from traditional snow based streamflow forecasting in several respects. Although nothing as dramatic as a “paradigm shift” is necessary, the hydrologist may need to develop additional skills, or develop a different frame of reference while issuing climate-based water supply outlooks.

For example, figure 7.10 effectively illustrates a contrast between climate and snowmelt hydrology. The correlation between snowpack and streamflow is so strong that it is relatively easy to believe that the relationship is (almost) deterministic. By 1 April, an exceptionally heavy snowpack is virtually guaranteed to produce proportionately high streamflow, and similarly, low snowpack yields low streamflow. Before December or January, the hydrologist has little information “on the ground” upon which to base a forecast.

In contrast, the correlation between climate and streamflow is marginally significant, and one **must** think of the relationship probabilistically. While climate forecasters may find it loathsome to produce long-lead deterministic climate forecasts, many hydrologists fear that users might not accept a probabilistic streamflow forecast of this skill level, thinking that users would view them as “vague”, “hedging”, or “non-committal”.

Hydrologists simply cannot describe a streamflow forecast issued in September in terms of “102% of normal” as they would a forecast issued in April. It becomes too tempting to describe such small shifts as a forecast of “we’re looking at near-normal conditions this year”, a gross distortion of the true situation of large forecast uncertainty. Hydrologists should ask themselves if they are comfortable conversing in probabilistic terms as a climatologist would. The ultimate test occurs when the hydrologist must issue an “equal chances” or “climatology” forecast (i.e., the outlook is completely uncertain), without a sense of failure, of “giving up” too easily, and “letting down” the users.

Hydrologists are also aware of the institutional barriers to using probabilistic forecasts. Many reservoir operating rules require a deterministic streamflow value. Water managers seeking to implement new dynamic operating procedures based on probabilistic forecasts encounter resistance from decades of tradition and external pressures to maintain consistency in operations. Such resource management, in the face of many highly conflicting interests, can result in rigid agreed-upon management practices, lest one party believe a new course of action is being taken at their expense to the benefit of others. Although some sophisticated water managers do consider risk and appreciate information about forecast uncertainty, a number of difficult challenges remain to those attempting to communicate probabilistic streamflow forecasts effectively. Some hydrologists would prefer not to issue a forecast that they suspect the user could not use or would misinterpret (Pielke Jr 1999). The nature and scope of these challenges are explored further in the next section.

There is also a spatial scale contrast between climate forecasting and streamflow. The strong correlation between snowpack and streamflow requires close scrutiny of small spatial variations in snowpack when forecasting. As a result, hydrologists generally frown upon forecasting using snow measurements outside a basin's boundary. To the most extreme case, forecasters balk at using snow in central Arizona to forecast New Mexico streamflow although a weak correlation exists. Hydrologists lack knowledge about what new climate indices (e.g., the Arctic Oscillation, the Quasi Biennial Oscillation, Solar Cycles, etc.) are "in" their basin so that it makes sense to consider them when forecasting or which ones are "outside" of their basin and are spuriously correlated.

At the opposite end of the spectrum, climate forecasters typically focus on large-scale continental, if not global, patterns when making their forecasts. If the contours on a national forecast map match the observed contours, except that they are displaced, for example, 1000 km to the east, it is generally thought of as a successful forecast. If the connection between climate and streamflow is to be made, streamflow forecasters will need to think "bigger" than they typically have.

9.3 On cultivating skepticism, combating pessimism, retaining credibility

While very long-lead water supply forecasting requires proficiency in climate variability, it also demands expertise in probabilistic forecasts and concepts. Although nothing prevents the generation of short lead-time probabilistic forecasts, the uncertainty in long-lead forecasts brings the issue to a head.

The water supply forecaster who issues a highly uncertain probabilistic climate-based streamflow forecast should be prepared to engage users who demand that the forecaster "come clean" and tell them "what the forecaster **really** thinks is going to happen". This discussion is, of course, ill-framed because all forecasts are at their root probabilistic. Deterministic forecasts are probabilistic forecasts with zero error bounds (i.e., complete confidence). A deterministic forecast may also be some point along the probabilistic forecast distribution, arbitrarily chosen by the forecaster (e.g., the mean, median, or mode).

The danger in allowing the forecaster to choose the "one number" is that the internal risk model of the forecaster may be different from that of the user. Unless the forecaster is intimately familiar with the user's operations, the forecaster is not qualified to judge what level of risk the user should accept. It is not the role of the forecaster to determine if and how water managers should use probabilistic forecasts to manage risk. Ultimately, the forecaster's efforts should be focused on quantifying and issuing the most unbiased, informative, and useful forecast possible (as discussed by Murphy 1993). At the same time, Pielke Jr (1999) asserts that the forecaster should not necessarily be completely independent of the user in that "A view held by some– that forecasters forecast and what others do with the forecasts is their responsibility– no longer seems tenable".

While the scientific literature has repeatedly shown that probabilistic forecasts are more appropriate and articulate than deterministic forecasts, the operational hydrology community is divided concerning the perceived low user demand for probabilistic forecasts and their inability to interpret them. In a recent case, a southwestern water manager, the Salt River Project, commissioned the development of an advanced climate-based water supply forecasting tool, but the user then developed a post-processor to convert the probabilistic output into a deterministic forecast. As a counter example, Baker (1995, quoted in Pielke Jr 1999) found that "people are more capable of comprehending and using at least certain types of probability information than is usually noted in the information processing and subjective risk literature". In other words, hydrologists underestimate users' exposure to and understanding of probabilistic forecasts.

If confronted with a user demanding a deterministic forecast, the hydrologist should consider if the user, in asking for the uncertainty to be removed from the forecast, tacitly wants the uncertainty to be removed from nature. After all, given enough time, money, satellites, and climate indices, one **should** be able to come up with the perfect forecast. The user, dissatisfied with the agency forecasts' large uncertainty, may seek out alternate opinions among, for example, the outputs of individual forecast tools or private consultants.

While it can be difficult to distinguish this user from the sophisticated user who accesses as much information as possible to supplement the official forecast, the former may suffer from "confirmation bias". This is a type of natural selective thinking encountered in a variety of contexts whereby one tends to notice and to look for what confirms one's beliefs and to ignore, not look for, or undervalue the relevance of contradictory evidence (Kahneman et al. 1982). Regrettably, one hydrologists' "confirmation bias" is another's "human expertise". It is difficult to objectively self-assess whether one is "gifted" or "fooling oneself".

The most dangerous combination is a user with a confirmation bias who relies upon forecasters who suffer from their own form of confirmation bias and who thus are willing to "go out on a limb" to attract customers with very confident (and thus presumably skillful) forecasts. When this water manager uses a "one number" deterministic forecast, which then greatly differs from the observed, the user is likely to foist responsibility for any negative outcome back onto the forecaster who presumably "read the signals wrong" or did not try hard enough (i.e., "they blew it big", Foster 1997).

Some operational water supply forecasters are skeptical of climate forecasts, often because of an instance in which the individual put faith in a climate outlook, and this resulted in undesirable consequences and regret. The episode of the 2000/2001 La Niña/cool PDO and the ensuing Pacific Northwest drought described in chapter 5.4 is an excellent example. Although all objective climate-based guidance pointed towards a wet winter, 2001 tied or broke records for the driest year on record in the Pacific Northwest. Many streamflow forecasters have a "What about 2001?" anecdote readily available as a justification as to why they do not rely on climate forecasts more heavily.

Forecasters and users alike must accept that, since the relationship between streamflow and climate is probabilistic, "No one can win them all." The threat of having a forecast "bust", however, strikes fear into all but the most steeled hydrologists. As Lewitt (1995) describes this situation: "[The event is not] entirely predictable, though it is possible to calculate the ranges of probability. Still, in every range there is the one in a billion chance, the blind shot that seems so improbable that we ordinarily discount it. And when it does happen, our sense of fair play is often more injured than our actual conditions." Who accepts responsibility when nature does not obey the predictions – the climate forecaster, the hydrologist, or the user? Given sufficiently negative consequences, even a long record of appropriate decisions can be negated by a single "bad" decision. Over the long term, however, if the climate information is properly used, the streamflow forecasts should improve in general.

While important, the Pacific Northwest example should not be overstated. At the opposite end of the spectrum from the user trying to strip nature of its uncertainty is the one who believes that long range predictability is impossible. One might encounter a hydrologist who perceives that "making a streamflow forecast in September, before any snow has accumulated, amounts to swinging before the ball has been pitched. One is bound to strike out."

Such hydrologists may feel *Schadenfreude* (malicious joy) when a forecast disagrees with the observed because it confirms their negative impressions of climate forecasts and releases them from any need to change their current operations. A forecast user may adopt the same mis-perspective that if the future is completely uncertain, there is no need to deviate from business as usual. Even if a catastrophic event occurs, such users feel absolved of responsibility, as the disaster was an unforeseeable "Act of God". The use of fixed reservoir operating "rule curves" operates under the principle of minimizing risk in the face of complete future uncertainty. The reality of climate forecasts lies somewhere in between the extremes of complete uncertainty and complete predictability.

One key to interpreting and using probabilistic forecasts is to have a clear quantitative understanding of forecast uncertainty. Often, users have only a subjective notion of how close the observed ought to be to the forecast to consider it acceptable. If the observed deviates too far from this subjective tolerance, then the user denotes this forecast as a "bust". Whether a forecast is a "bust" or not, however, depends on whether the observed lies outside reasonable error bounds, which themselves depend on the forecast uncertainty. Users must be fully cognizant of this interrelationship to understand the magnitude of possible deviations of observed from forecast. In the end, there are no "bad" probabilistic forecasts, only unlikely outcomes (of course, Murphy (1978,1993) and Murphy and Epstein (1967) might argue that there are bad, more specifically **inappropriate,** forecasts when hedging occurs but this is a separate issue).

A second key to understanding and using probabilistic forecasts is to realize that the chance of the observed ever equaling the deterministic forecast is essentially zero. Even under the best circumstances, one will always observe more or less than the forecast quantity, with probabilities described by the error distribution. Once this is understood, users can then develop, and when necessary implement, contingency plans in the event that more or less water is received than the forecast. This is true regardless of the chosen exceedence probability of the forecast quantity. Difficulties can arise if users and managers base their plans only on a single forecast quantity, ignoring the possibilities described by the forecast distribution. The danger in interpreting the "one number" forecast as “destiny” is particularly serious when involving long-range climate-based streamflow forecasts because the likely error is much higher than late-season snow-based forecasts.

9.4 Practical advice to water supply forecasters

Climate forecasts have long represented an opportunity to improve seasonal water supply forecasts. For decades, however, climate forecasts have been perceived as having insufficient skill and specificity for use in the operational hydrology environment. While climate forecasts may not significantly improve water supply outlooks during the snowmelt period, they possess great strength in providing information prior to snowpack accumulation, as early as September. While these pre-season forecasts are highly uncertain, they remain an improvement over the next best alternative (i.e., no information at all).

Although some technical barriers to incorporating climate outlooks into the water supply forecasts exist, the primary challenge is a perceptual barrier. To utilize such highly uncertain climate information properly, forecasters and users both must understand water supply forecasts in probabilistic (rather than deterministic) terms. Regrettably, operational hydrologic, climate, and weather forecasters have struggled for decades to communicate forecast uncertainty effectively (O’Grady and Shabman 1990; Sarewitz et al. 2000). Some progress has been made, particularly in the past decade or so, in the tabular and graphical display of forecasts to communicate more clearly the probabilistic nature of the forecasts. Continued efforts along these lines in both the academic and operational communities are needed.

While it is outside of the scope of this dissertation to determine if water managers should use long-lead yet uncertain climate-based water supply forecasts, it is safe to say that operational forecast agencies will inevitably start issuing them. Water supply forecasts were originally issued first in April, with March forecasts beginning in the 1950s, February forecasts in the mid-1960s and January forecasts in 1980. The historical trend towards longer lead-time forecasts suggests that the advent of December (or earlier) forecasts is overdue. The question remains not whether but how best to implement this system.

Operational forecast environments typically have several forecasters, each responsible for a limited subset of basins within the office's larger forecast area. At least one of these forecasters should have good to excellent proficiency in interannual climate variability, with a working knowledge of the tools used by the official climate forecasters at the Climate Prediction Center (CPC). During the forecast season this individual is encouraged to monitor and/or participate in the forecast development teleconferences CPC holds. This hydrologist can then brief the other hydrologists on the climate outlook, field questions about the forecast and develop a collective strategy on the implications for local streamflow. It might be possible for the climate-savvy forecaster to develop the pre-season forecast for all areas, alone, with subjective input from the other hydrologists.

This forecaster should be able to provide practical advice on using climate information in forecast equation development. For example, climate signals are typically large scale in nature (e.g., larger than 500 km across) except in coastal regions where the effects can be isolated. Therefore, if no streams in a region are correlated with climate except one, the correlation is likely spurious. Climate phenomena typically contain much persistence from month to month, and their high frequency variability usually does not contain relevant information. Three-month averages (such as September-November) of climate indices should suffice. In the end, the water supply forecaster must use sound hydrologic judgment and avoid the “garbage can” and “hunting and pecking” approaches of statistical forecasting (i.e., exhaustively fitting a historical streamflow time series to dozens to hundreds of candidate variables to find the best fit).

Also, one should choose only climate indices that will be available at forecast time; currently the Southern Oscillation Index is operationally supported, whereas the Pacific Decadal Oscillation is updated irregularly by an academic institution.

Each office within the water supply forecast environment would benefit from an individual also proficient in advanced statistics and probability concepts as well as someone with an interest in visual display and communication of uncertain information. These members can develop a regionally appropriate strategy for emphasizing forecast uncertainty without overly discouraging users. They can also address whether early-season forecasts require a fundamentally different format from those issued throughout the regular season (as discussed in the following section). Depending on availability, the agency may partner with the local NOAA Regional Integrated Sciences and Assessments project to serve as a user liaison. These projects have the interest and resources to develop and quantitatively test alternative forecast delivery formats. All forecasters should have a working knowledge of basic statistics and probability concepts; popularly accessible works such as Bernstein (1998), Gilovich (1993), Kahneman et al. (1982), Plous (1993), or Pollack (2003) can also assist in giving forecasters basic non-technical tools and concepts to help communicate forecast uncertainty to users.

The forecast environment should already be capable of historical forecast archival for the evaluation of forecast accuracy. There is no reason why this system cannot also include more uncertain, early season climate-based forecasts. Retrospective evaluations can measure the relative improvements of using climate information over existing practices (as was done in chapters 6 and 7). Hindcasting and simulated forecasting exercises (such as Baldwin 2001) can help streamflow forecasters build realistic expectations (that is, neither overly inflated nor unnecessarily pessimistic) of what will occur when using climate forecasts. If effective, there is a good chance that the climate forecasts will be properly applied, without regrets.

9.5 Display formats

Currently, as described in chapter 4.3.1, the primary products of the water supply forecasts are tables containing 10%, 30%, 50%, 70% and 90% probability of exceedence forecasts, and a map of the 50% probability of exceedence forecast expressed as percent of the long term average. While it is possible to display highly uncertain climate-based streamflow forecasts in this format, other formats may convey more useful information. The hydrologist is strongly discouraged from simply guessing the users needs, picking a favorite format and operationalizing it. Such an approach has a low probability of discovering and meeting user needs. Professional user and usability tests to determine the actual (versus hydrologist perceived) ability of users to interpret, comprehend and accept the products are a must (Nielsen 1994). There is a well-developed body of literature on the effective visual display of quantitative information, and Tufte (2001), for example, provides many general principles and practical guidelines. While it is tempting to do so, hydrologists should not assume that product development is just “common sense”. Trained professionals exist for this task.

The format that contains the most quantitative information for a single location, variable and target period is the probability of exceedence graph (Barnston et al. 2000). This graph displays the volume of seasonal streamflow on one axis and its probability of exceedence on the other. Several lines can be drawn. One line for the climatology distribution (i.e., the last 30 years of data) and one line for the current forecast are generally the necessary minimum. The climatology distribution places the forecast in context. The NRCS currently supports horizontal bar charts that approximate the probability of exceedence graph (figure 9.1). However, without any historical context for the different streamflow levels, the user is left wondering, “How unusual is this forecast for 200 kac-feet? Does that streamflow happen often?” CPC also places additional information on its POE graphics, such as text about the probabilities that the observed will fall in particular categories, or a time series at the bottom showing the observed precipitation for the past several years.

Figure 9.2 is a mock POE graph for what the Pecos NM climate-based streamflow forecast system (described in chapter 7.5) might have produced on 1 November 1999 for March-July streamflow. The solid stair line is the ranking of the historical observations from 1961-1990 (using the Gringorten plotting position). The dashed smooth line is a normal fit to this climatology distribution (mean 49, standard deviation 30). Other distributions, such as a log-Pearson distribution, provide a better fit for streamflow data, but the normal distribution is shown here because it corresponds to the shape of the error distribution used in the forecast equation. The smooth solid line is the forecast distribution ("most probable" 29, jackknife standard error 26.7). This line lies to the left of the climatology distribution, indicating a higher chance of dry conditions, due to PDO's cool state and the La Niña. Along the bottom of the graph is a listing, by 2-digit year, of the volumes of the past 10 years of streamflow. For example, March-July 1999 ("99") had observed flow of 57 kac-ft, and 1997 ("97") had 110 kac-ft.

Several problems are immediately evident in this graph. The normal distribution is a poor fit to the 1961-1990 climatology, especially below 20 or above 80 kac-ft. The forecast is also unrealistic at the tails of the distribution. This forecast indicates that there is a 15% chance of less than zero flow, a physically impossible outcome.

Currently, the forecast development and display format is such that NRCS operational forecasters would only know that the 90% of exceedence bound is negative (see table 4.1). They would "fix" this problem, often by improvising a new lower bound of the forecast. They might use a common rule of thumb, e.g., half of the value of the 70% exceedence bound. In this instance, the 70% exceedence volume is 15 kac-ft, and the improvised 90% exceedence volume would be in the neighborhood of 8 kac-ft. Unfortunately the NRCS forecast environment is not integrated with a database of historical streamflow data and the hydrologists would not even realize that 8 kac-ft is still too low of a forecast because it is much less than the lowest observed flow on record.

If the water supply forecaster decides to publish a continuous POE graph such as this, the "fix" described above is no longer a viable option. The forecast distribution must automatically take a form that produces positive and realistic values across its entire range. The CPC encountered this issue in their precipitation forecasts, especially when the distributions are very highly skewed (e.g., southern California in summer). CPC forecasters have converted their tercile forecasts into complete distributions by assuming gamma distributions (e.g., Briggs and Wilks 1996; Wilks and Eggleston 1992). It might be somewhat more difficult to apply their methodology to streamflow forecasts because Wilks and others assumed little to no skill-based contraction of the forecast distribution. Streamflow forecasts issued in April will have very highly contracted distributions and this assumption would not be valid. Regardless, the above example illustrates how the current operational environment remains focused on the deterministic streamflow forecast and is not designed to help the hydrologist accurately articulate probabilistic guidance. This will need to change if climate-based water supply forecasts are to be issued.

Many different products can be derived from the POE plot just shown. For example, additional information might be displayed in a table such as Table 9.1. Of course, the language in this table is cumbersome and other terminology should be field tested with users to see what has the highest chance of correct interpretation.

Table 9.1. Sample streamflow guidance based on 1 November 1999 synthetic climate based hindcast

Pecos River near Pecos, NM
March-July Volume 1000s acre-feet

Period of Record mean/median:	52		48
1961-1990 mean/median:	49		42
Previous 10 year (1990-1999) mean/median:	69		71
Dry/Wet tercile boundaries for 1961-1990:	36		62
Driest/Wettest year on record:	11 (1950)		153 (1941)
Forecast:			
Percent chance of below/above median (42 kac-ft)	69%		31%
Percent chance of dry/near normal/wet:	60%	30%	10%
Percent chance of falling below Pecos Compact (18 kac-ft)			34%
50% chance of above:	29 kac-ft (60% of 1961-1990 average)		
Forecast Confidence:	Low		
1961-1990 standard deviation (C):	30 kac-ft		
Forecast standard deviation (F):	27 kac-ft		
Forecast distribution compression (1-F/C):	0.10		

This table begins with several historical statistics to place the current forecast in context. For example, compared to this historical record, the 10 years prior to 2000 were extremely wet. An optimistic user might expect a continuation of the trend towards wet conditions. Another user might look at recent years as anomalous and that dry years in the near future would mean a return to normal.

The forecast in Table 9.1 and figure 9.2 indicates a 69% chance of flows below the median of flows during 1961-1990. If the historical record is broken into three equally likely categories (terciles), as CPC displays its climate forecasts, there is a 60% chance of dry. This “probability anomaly” of 26.7% (60% - 33.3% = 26.7%), is unusually strong for a climate forecast. Probability anomalies in the range of 5-15% will be more typical in other years for other locations. If the user desires a deterministic point forecast (e.g., the “most probable” [sic] forecast) that too is possible. With an interactive forecast processor through, for example, the Internet, users would be able to specify the threshold they are interested in and find the probability that it will fall below that threshold. For example, a fictitious threshold (18 kac-ft) at which junior irrigators in New Mexico might get cut off is provided.

The quantitative information about the compression of the forecast distribution is most likely to appear as meaningless jargon to the user. For example, the historical standard deviation of flows is known, as is the root mean squared error of the forecast equation during jackknife calibration. The ratio of the forecast distribution width (F) to the climatology distribution width (C) is a useful measure of the confidence and expected skill of the forecast. Forecasters could develop a look-up table translating a quantitative measure of forecast distribution compression (1-F/C) into a qualitative phrase. For example, 0.00-0.1 means “No” forecast confidence, 0.1-0.25 means “Low” forecast confidence, 0.25-0.5 means “Moderate”, 0.5-0.75 means “High” and 0.75-1.00 means

"Very High". The forecaster would have the liberty to degrade the "confidence" rating by one category if there are known data quality issues specific to this year's forecast or if other circumstances arise.

In terms of visual map-based forecast products, any one of the parameters in Table 9.1 could be displayed spatially. Figure 9.3 shows how Dettinger et al. (1999) display their climate-based streamflow forecasts. Shown are probability anomalies relative to three equally likely categories. The large filled circles in the Pacific Northwest in the top diagram indicate a ~60% chance of winter flows in the wettest third of record. This dual-map system provides more flexibility than the single-map system of CPC's forecasts, in that one is not required to follow any fixed rules about adding probability to one category and removing it from another.

IRI's climate forecast format (chapter 4.2.1) is also another candidate for how streamflow forecasts might be displayed. While it makes sense to discretize a precipitation forecast map into 2x2 degree cells, the interpretation of such a streamflow forecast map can be difficult. For example, a grid cell over central New Mexico could refer to local conditions such as the Pecos or Santa Fe rivers. It could also refer to the Rio Grande mainstem, which has its headwaters in southern Colorado but bisects the city of Albuquerque NM. Hydrologic Unit Code (HUC) basins may be a more appropriate discretization, as is currently done with the NRCS westwide forecast maps (chapter 4.3.1).

9.6 Summary

In determining what products to serve to users, there are several tradeoffs to consider. For example, does one provide as much quantitative information as possible and let the user sort out what they require? A danger in providing too much quantitative information is that it implies precision and confidence. For example, a statement such as "there is a 50% chance that this year will be more than 43.2734 kac-ft" may leave the user with the impression that exactly 43.2734 (+/- 0 0.0005) kac-ft is the expected outcome. The user is given a different impression if the forecast is framed in terms of "a 50% chance that flows will be between 30 and 70 kac-ft". Even if both statements are true, the second framing reminds the user of the broader range of possibilities.

In order to not imply complete precision in uncertain forecasts, one can choose formats that restrict the users ability to derive quantitative information, much as an engine governor prevents a driver from driving an automobile too fast. For example, the most a user on the Salmon River at Whitebird Idaho, could derive from figure 9.3 is that there could be a +10 and +30% higher than usual chance of falling in the wettest tercile. One does not know exactly what "wet" means in terms of local streamflow. The water manager may end up qualitatively using this information to contemplate the implications of a wet (how wet?) scenario but may not act.

Ultimately, the forecast agencies must give serious consideration to the costs and benefits of adopting forecast formats that hobble some users in order to help others. Once the agencies determine their objectives in producing long-lead uncertain streamflow forecasts, they must consult with the users to determine the best format to achieve those objectives. Lastly, forecast agencies should know that they do not need to "reinvent the wheel" when it comes to developing and communicating these new products. With over 70 years of experience providing users with water supply forecasts, it may be institutionally difficult to convince the NRCS that it has anything new to learn about its customers. Nonetheless, the NRCS would benefit from the existing wealth of social science literature on the subject of communicating uncertain climate forecasts to a broad range of users (e.g., NRC 1999; IRI 2001).

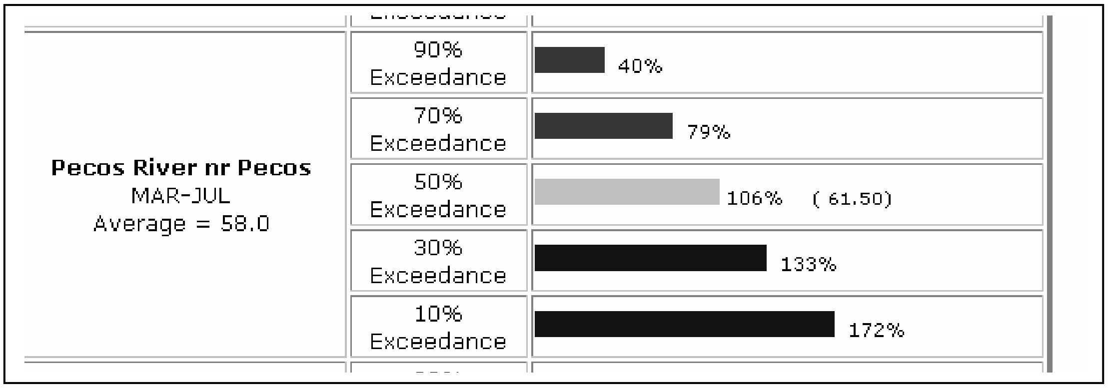


Figure 9.1 NRCS probability of exceedence bar chart for the Pecos River near Pecos on 1 January 2003. The length of the bar is proportional to the volume associated with the forecast at various probability of exceedence levels. The numbers to the right of each bar are the volumes as percent of the long term average flow (58.0, as shown to the left). The number in parenthesis to the right of the gray bar is the volume of the 50% probability of exceedence forecast, in kac-ft.

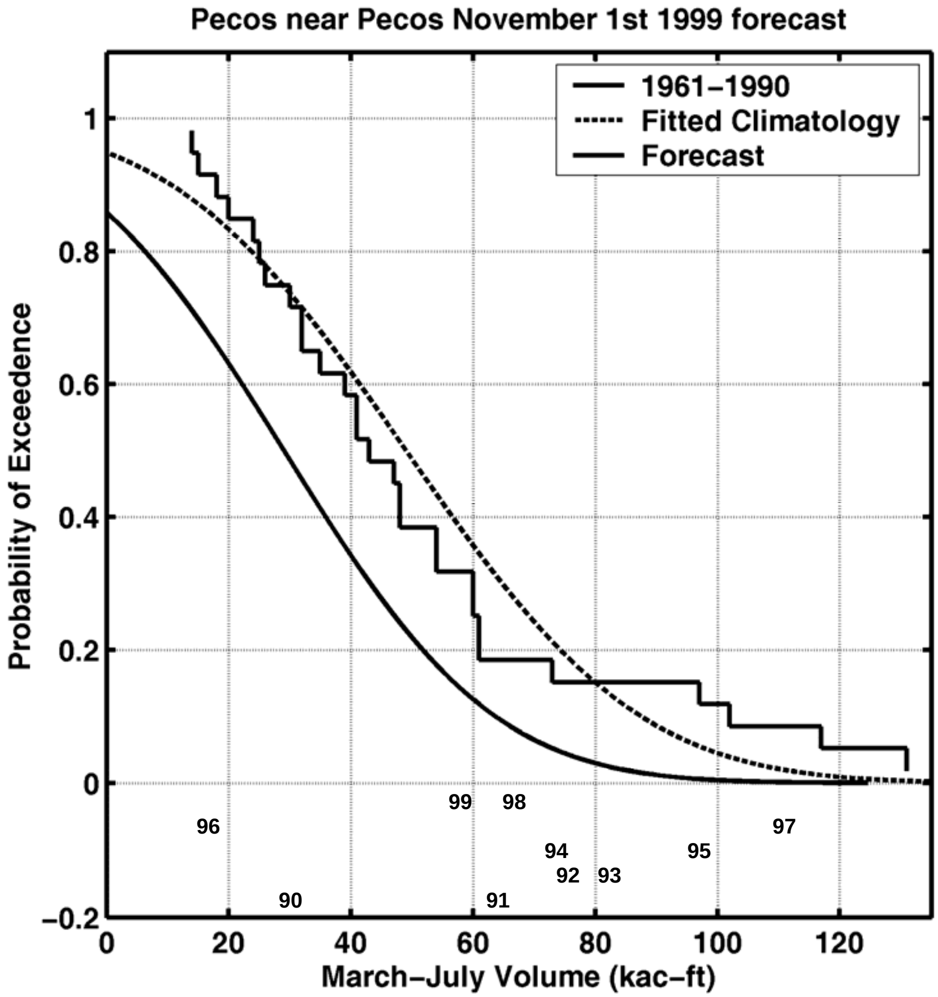


Figure 9.2 Example alternative water supply probability of exceedence diagram. See text for further discussion. Volume is on the X-axis and probability of exceedence is on the Y-axis. Shown are the long-term climatology, the climatology fitted to a normal distribution as well as the shifted forecast distribution (based on the synthetic climate-based hindcast that would have been issued 1 November 1999). Along the bottom are the previous 10 years (i.e. 99 = the year 1999) of flow plotted on the X-axis at their respective observed seasonal flow volumes.

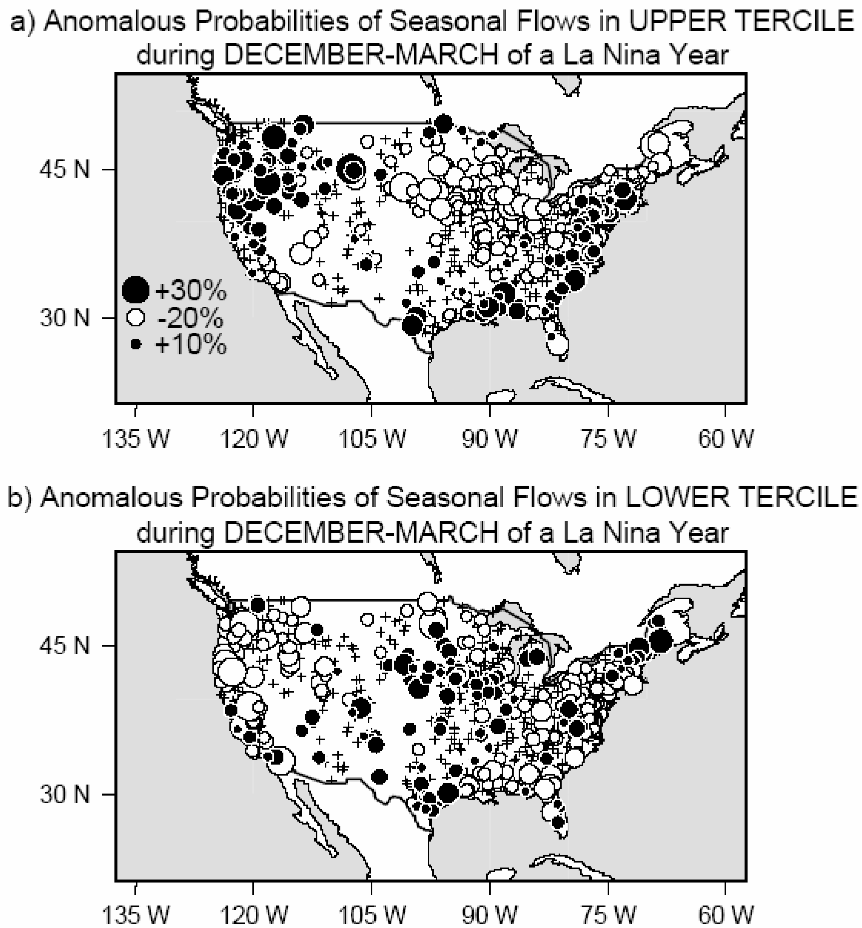


Figure 9.3. Example of a climate-based streamflow forecast map. Original caption from Dettinger et al (1999): Anomalous probabilities of occurrence of seasonal-mean December-March flows in the (a) upper tercile and (b) lower tercile of historical flows at selected rivers in the conterminous United States, based on historical flows during tropical La Niña episodes. Black dots indicate rivers with probabilities at least 10% greater than the 33% probabilities expected of a random sampling (that is, greater than 43%); white dots indicate rivers with probabilities at least 10% less than the expected 33% (that is, less than 23%); dots are scaled so that the radii are proportional to the probability in excess of 33% and small crosses indicate rivers that do not reach probability thresholds for the dots.

10. CONCLUSIONS AND RECOMMENDATIONS

This dissertation investigated the relationship between climate and operational hydrology and the NRCS. It reviewed past studies of the impacts of major climate phenomena on the Western US, as well as past attempts to link climate forecasts and water supply forecasts in the operational environment. El Niño, in particular, has a well-understood influence on western hydroclimatology, favoring wet conditions in the Lower Colorado Basin and New Mexico and favoring dry conditions in the Pacific Northwest and Cascades. El Niño impacts in the central Western US from California through Nevada, Utah and Colorado remain under study but the general sentiment is that the effect in those regions is weak at best or only isolated to special circumstances. The Pacific Decadal Oscillation is a popular but not well-understood phenomenon that, spatially and seasonally, has approximately the same influence on Western US precipitation as El Niño. A limited number of studies suggest the North Atlantic Oscillation may have upstream impacts on the Western US, but this dissertation found unpromising correlations with water supplies.

This dissertation also documented the history of operations oriented research in hydrology and climatology, described the roster of operational forecasting tools and traced the revolution of a forecast cycle. Methodologically, climatologists are more sophisticated in their statistical forecasting tools and have had much more success than hydrologists at operationalizing dynamic simulation modeling. In this respect, hydrologists can make rapid technological progress by adopting (versus developing anew) advanced simulation modeling or statistical forecasting methodologies. Both operational environments are time critical and harried, and if climate information is to be integrated into water supply forecasts, it must be done seamlessly and conveniently.

This dissertation evaluated and diagnosed the accuracy of existing operational water supply forecasts. This evaluation is significant in that it is the first westwide scientific evaluation of the official forecasts in the 70-year history of the NRCS to appear in the peer-reviewed literature (Pagano et al. 2004b). It reviewed previous studies and outlined a useful methodology for meaningful evaluations of water supply forecasts. The results confirmed and quantified intuitive notions about the evolution of forecast skill versus leadtime. It documented the strong relationship between forecast error and subsequent precipitation after the forecast issue date. It discovered a previously unknown tendency for low forecast skill in the Oregon cascades and the western Great Plains.

Next, an objective system to generate synthetic hindcasts that mimic the behavior of the operational forecasts was developed. Using accumulated precipitation and snow data, this system produced hindcasts with an exceptional correspondence with the official forecasts and had comparable skill. With this system, sensitivity tests were conducted to determine the relative merit of including climate information at various times of the year. In April, climate information has very little to offer the already highly skillful snow and precipitation-based water supply outlooks. In January, climate contributes a considerable amount of skill to the forecasts and by November, climate information accounts for almost all available forecast skill. The skill of these pre-season forecasts is very low but not zero in many locations. The forecasts are highly uncertain but they represent an improvement over the next best alternative of not having any information at all.

This study addressed the methodological question of whether developing separate forecasting equations is an effective strategy for dealing with decadal variability in climate. The PDO index does improve forecasts in the Pacific Northwest and Southwest when it is used as a variable in a regression equation. In contrast, subsetting calibration periods by PDO state and then selecting among different forecasting equations based on the current PDO state is a poor strategy. In some regions, always accidentally selecting the "wrong" PDO equation yielded better forecasts than always choosing the "correct" PDO equation. The PDO subsetting strategy does not appear to be a viable methodology for producing operational climate-based water supply outlooks.

Contrary to conventional wisdom about the long-term improvement in water supply forecasts due to investments in observational and methodological technologies, this study found that forecast skill was highest from the 1960-1980s. After this period, skill dropped sharply, returning to levels last seen in the 1950-1960s. In particular, there has been sharp a decline in forecast skill in Oregon and the Colorado/Rio Grande basins. The synthetic hindcasts were able to exactly reproduce this decadal variability in forecast skill in both space and time, indicating that the rise and fall of forecast skill is independent of the human forecaster. Indeed, the synthetic hindcasts showed that in the context of the long-term record, forecasts during the 1960-1980s were anomalously skillful; the recent decline is more of a return to normal.

To diagnose the causes of this skill variability, this study found that there are strong decadal trends in observed seasonal streamflow variability and persistence (autocorrelation). This result, while previously unknown to the scientific literature, is one of the expected impacts of anthropogenically induced climate change. Specifically, the 1950-1960s was a very calm period with high year-to-year persistence in streamflows. The period 1960-1980 had moderate variance but anti-persistence (a tendency to rapidly switch between wet and dry years). Flows during the most recent 20 years, however, are the most variable of the entire period of record across the Western US. Flows during the past 20 years are also more persistent of any others previously seen.

These trends in seasonal streamflow variability, alone, however, do not sufficiently explain the decadal trends in forecast skill because their spatial patterns do not match. For example, streamflow variability is on the rise in California, a region where forecast skill has remained steady. Forecast skill has declined in New Mexico, where streamflow variability is not on the rise.

Analysis in this dissertation and in previous studies suggested that changes in the behavior of spring precipitation might be responsible for the trends in forecast skill. This study found that the relationship between spring precipitation variability and 1 April forecast error is very strong, explaining 30-60% of the error variance. Declines in forecast skill are, indeed, linked to a higher frequency of extreme precipitation events (both wet and dry) in the most recent 20 years. Specifically, spring precipitation extremity was high from 1930-1950 steadily decreasing over the next 40 years and reaching its most calm period around 1960-1980s. During this period, almost no part of the Western US had increased springtime precipitation irregularity and many locations had considerably more invariant conditions than usual. In just over 20 years, Western US springtime precipitation irregularity soared to levels not previously observed in the modern record. Spring precipitation irregularity has increased the most in Oregon, the Upper and Lower Colorado and Rio Grande basins, precisely those regions that have experienced a sharp decline in forecast skill. No explanation is currently available as to why spring precipitation extremity has increased, however.

The final chapter of this dissertation dealt with issues related to the operationalization of climate-based streamflow forecasts. There are many thorny and challenging issues that hydrologists will have to address with regards to their perceptions of and abilities to effectively communicate highly uncertain probabilistic streamflow forecasts. Also provided are some practical advice to hydrologists who would seek to generate such forecasts, and gave suggestions on display formats.

As such, this study has several recommendations:

1) **Begin issuing experimental NRCS 1 December water supply forecasts**. There is a sound scientific and methodological basis for these forecasts, and this study has outlined the steps towards developing such a product operationally. If these forecasts are well received by users, they may become an official product, and experimental October or November water supply outlooks could follow.
2) **Develop climate literacy training within the NRCS**. This training may or may not be modeled around similar training currently being developed within the

National Weather Service. It may also include issue briefing papers and training materials for NRCS State Water Specialists and users. Water Supply forecasters may also benefit from "gaming" simulations with regards to climate forecasts, similar to Baldwin (2001). This training could be developed jointly with university groups.

3) **Address issues of uncertainty and probabilistic versus deterministic water supply forecasts.** This may include articulating the role of the water supply forecaster or State Water Specialist in the decision making process of the user. The NRCS should also review the social science literature or engage social scientists with regards to the effective communication of highly uncertain information.

4) **Appeal to the physical research community to determine the underlying causes for the decadal variability in streamflow and spring precipitation**. Similarly, the performance of the NRCS as an agency should not be tied to the performance of its forecasts, unless the performance measure incorporates the effects of natural variability. The NRCS should be keenly interested in knowing if the variability in spring precipitation will keep rising as it means that the skill of the forecasts may continue to fall.

5) **Encourage the climate change community to further explore the implications of long term climate change on water supply forecasts**. In particular, will snow-based streamflow forecasts remain relevant under a warmer climate? This question could be answered by forcing a Global Climate Model with observed

20th century climate and regressing a model variable (e.g., snowpack) with observed and simulated streamflow. The model would then be run forward in time to a warmer climate and tests done to see if the relationship between, for example, model snowpack and simulated streamflow remain constant.

6) **Track methodological developments within the climate community**. Researchers and operational climatologists are making excellent progress in the fields of ensemble dynamic simulation model forecasting. If and when the NRCS develops a simulation model, it could receive knowledge about bias adjustment and model combination techniques. It could also benefit from directly linking the outputs of the atmospheric models to hydrology models (e.g., Hay et al. 2002; Clark et al. 2003), or climate-weighting its ESP traces.

11. APPENDIX A: DERIVATION OF THE RELATIONSHIP BETWEEN THE COEFFICIENT OF VARIATION AND THE EXPECTED FREQUENCY OF ERROR WITHIN A PARTICULAR BOUND.

Let o be a collection of observations. The coefficient of variation, CV is

$$CV = (std(o)/ \bar{o})$$

where std(o) is the standard deviation of a collection of observations and $\bar{o}$ is the mean of the observations. The standardized anomaly, Z is

$$Z = (o–\bar{o})/std(o) = (o–\bar{o})/(cv * \bar{o}).$$

Let perc be the percentage error being analyzed. Let

$$o_{upper}= (1+perc)*\bar{o}$$

$$o_{lower}=(1-perc)*\bar{o}.$$

Then,

$$Z_{upper} = ((1+perc)* \bar{o} – \bar{o})/(cv * \bar{o}) = +perc/cv.$$

$$Z_{lower} = ((1-perc)* \bar{o} – \bar{o})/(cv * \bar{o}) = -perc/cv.$$

The Cumulative Standard Normal Distribution is shown as NormalCDF(Z). NormalCDF(Z_{upper})-NormalCDF(Z_{lower}) is the frequency that a climatology forecast would have an error between Z_{upper} and Z_{lower}. This also has the form,

$$\text{NormalCDF}(+\text{perc}/\text{cv}) - \text{NormalCDF}(-\text{perc}/\text{cv})$$

NormalCDF does not have a closed form, but in the example where cv = 1/3 and perc = 0.1, the expected frequency is 0.236 or 23.6% of the time.

12. REFERENCES

Allan, R., J. Lindesay, and D. Parker, 1996: *El Nino Southern Oscillation and Climatic Variability*. Collingwood, VIC, Australia, CSIRO Publishing, 408 pp.

Andrade Jr, E. R. and W. D. Sellers, 1988: El Nino and its effect on precipitation in Arizona and western New Mexico. *J. Climatol.,* **8(4)**, 403-410.

Ararbanel, H., and five others, 1980: *Regional-seasonal weather forecasting*. JASON Tech. Rept. JSR-80-05, SRI International.

Army Corps of Engineers, 1956: *Snow hydrology: Summary report of the snow investigations.* North Pacific Division US Army Corps of Engineers, 443 pp.

Arvola, W. A. 1975: *Long range weather forecasting*. California Department of Water Resources.

Baker, E.J., 1995: Public Response to hurricane probability forecasts. *Prof. Geogr.*, **47(2)**, 137-147.

Baldwin, C 2001. Final *Report: Phase II Long-Range Streamflow Forecasting Using Climate Information.* Prepared for Denver Water, Denver, Colorado. [Available at http://www.engineering.usu.edu/uwrl/CBaldwin/DenverWaterForecasting.pdf accessed 8/10/04].

Barnett, T. P., and R. W. Preisendorfer, 1978: A multifield analog prediction of short-term climate fluctuations using a climate state vector. *J. Atmo. Sci.*, **35**, 1771-1787.

Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate,* **7(10)**, 1513–1564.

Barnston, A.G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Weather Rev.,* **115**, 1083-1126.

______, and _______, 1989: An operational multifield analog/anti-analog prediction system for United States seasonal temperatures. Part II: Spring, Summer, Fall, and intermediate 3-month period experiments. *J. Climate*, **2,** 513-541.

______, and C. F. Ropelewski, 1992: Prediction of ENSO episodes using Canonical Correlation Analysis. *J. Climate,* **5(11)**, 1316–1345.

______, and twelve others, 1994: Long-lead seasonal forecasts—Where do we stand? *B. Am. Meteorol. Soc.,* **75(11)**, 2097–2114.

______, and seven others, 1999: NCEP forecasts of the El Niño of 1997-98 and its US impacts. *B. Am. Meteorol. Soc.,* **80(9)**, 1829–1852.

______, Y. He, and D. Unger, 2000: A forecast product that maximizes utility for state-of-the-art climate prediction. *B. Am. Meteorol. Soc.,* **81(6)**, 1271-1279.

______, S.J. Mason, L. Goddard, D. G. DeWitt, and S. E. Zebiak, 2003: Increased automation and use of multi-model ensembling in seasonal climate forecasting at the IRI. *B. Am. Meteorol. Soc.* **84**, 1783-1796.

Barton, M., 1983: Reorganization of the USDA - Soil Conservation Service - snow survey and water supply forecast activity. *Proc. Western Snow Conf.*, Vancouver, WA, 151-154.

Bernstein, P. L., 1998: *Against the Gods: the remarkable story of risk*. John Wiley and Sons, 400 pp.

Bitz, C. M. and D. S. Battisti, 1999: Interannual to decadal variability in climate and the glacier mass balance in Washington, Western Canada and Alaska. *J. Climate,* **12**, 3181-3196.

Blanchard, F. B., 1955: Operational economy through applied hydrology. *Proc. Western Snow Conf.,* Portland, OR, 35-48.

Bohr, G. S., and E. Aguado, 2001: The use of April 1 SWE measurements as estimates of peak seasonal snowpack and total cold-season precipitation. *Water Resour. Res.*, **37**, 51-60.

Brandon, D.G., 1998: Forecasting streamflow in the Colorado and Great Basins using "El Niño" indices as predictors in a statistical water supply forecast system. *Winter '97/'98: Year of the great El Niño?* San Diego, CA, Floodplain Management Association.

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

Briggs, W. M., and D. S. Wilks, 1996: Estimating monthly and seasonal distributions of temperature and precipitation using the new CPC long-range forecasts. *J. Climate,* **9**, 818-826.

Brown, R. D., 2000: Northern Hemisphere snow cover variability and change, 1915-97. *J. Climate*, **13(13),** 2339-2355.

Brubaker, K., A. Rango, and W. Kustas, 1996: Incorporating radiation inputs into the Snowmelt Runoff Model. *Hydrol. Process.*, **10**, 1329-1343.

Burke, J., and D. Stevens, 1984: The effects of runoff forecasting on Colorado River operations at Hoover Dam. *A critical assessment of forecasting in water quality goals in western water resources management,* J.J. Cassidy, and D. P. Lettenmaier, eds., American Water Resources Association, 47-53.

Cadavid, L. G., R. Zee, C. White, P. Trimble, and J. T. B. Obeysekera, 1999: Operational planning in south Florida using climate forecast. *Proc. 19th Annual Am. Geophysical Un. Hydrology Days*, Ft. Collins, CO, Colorado State University.

Cahill, A. T., 2002: Determination of changes in streamflow variance by means of a wavelet-based test. *Water Resour. Res., **38***(**8)**, 1-14.

Cayan, D. R., 1996: Interannual climate variability and snowpack in the western United States. *J. Climate*, **9**, 928-948.

______, and D.H. Peterson, 1989: The influence of the North Pacific atmospheric circulation and streamflow in the West. *Aspects of climate variability in the western Americas,* D. H. Peterson, ed., Geophysical Monograph, 55, American Geophysical Union, Washington, D.C., 375-397.

______, and R. H. Webb, 1992: El Niño/Southern Oscillation and streamflow in the western United States. *El Niño: historical and paleoclimatic aspects of the Southern Oscillation,*H. F. Diaz and V. Markgraf, Eds., Cambridge University Press, 29-68.

______, K. T. Redmond, and L.G. Riddle, 1999: ENSO and hydrologic extremes in the Western United States. *J. Climatol.*, **12(9)**, 2881-2893.

CBIAC, 1955: *Use of 30-day weather outlook in forecasting runoff in the Columbia River basin.* Columbia Basin Inter-agency Committee, Portland, OR.

CBIAC, 1961: *Review of procedures for forecasting inflow to Hungry Horse Reservoir, Montana.* Columbia Basin Inter-agency Committee, Portland, OR.

CBIAC, 1964: *Derivation of procedure for forecasting inflow to Hungry Horse Reservoir, Montana*. Columbia Basin Inter-agency Committee, Portland, OR.

Christensen, R. A., and R.F. Eilbert, 1985: Seasonal precipitation forecasting with a 6-7 month lead time in the Pacific Northwest using information theoretic model. *Mon. Weather Rev.*, **113**, 502-518.

Church, J. E., 1935: Principles of snow surveying as applied to forecasting stream flow. *J. Agr Res.*, **51(2)**, 97-130.

______, 1937: The human side of snow: the saga of the Mount Rose observatory. *The Sci. Mon.*, **XLIV**, 137-149.

______, and H. P. Boardman, 1944: Effect of high water-table in exaggerating streamflow: Further analysis of unusual Year 1942-43 and analysis for year 1943-44 on Upper Humboldt River, Nevada. *EOS T. Am. Geophys. Un.,* **25(1)**, 96-101.

Clark, M.P., M. C. Serreze, G. J. McCabe, 2001: Historical effects of El Nino and La Nina events on the seasonal evolution of the montane snowpack in the Columbia and Colorado river basins. *Water Resour. Res.*, **37**, 741-758.

______, and five others, 2003: Use of weather and climate information in forecasting water supply in the Western United States. *Water and Climate in the Western United States,* W. M. Lewis, ed., University of Colorado Press, 69-92.

Clemen, R. T., 1989: Combining forecasts: a review and annotated bibliography. *Int. J. Forecasting,* **5**, 559-583.

Codd, A.R. and P.E. Farnes, 1960: Application of the electronic computer to seasonal streamflow forecasting. *Proc. Western Snow Conf.,* Santa Fe, NM, 21-23.

Croley, T. E. II., 2000: *Using meteorology probability forecasts in operational hydrology*. American Society of Civil Engineers, 206 pp.

______, and H. C. Hartmann, 1987: Near real-time forecasting of large lake supplies. *J. Water Res. Pl.-ASCE*, **113(6)**, 810-823.

Daly, C., R. P. Neilson, and D.L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteorol.,* **33**, 140-158.

Delecluse, P., M.K. Davey, Y. Kitamura, S.G.H. Philander, M. Suarez, and L. Bengtsson, 1998: Coupled general circulation modeling of the tropical Pacific. *J. Geophys. Res.*, **103(C7)**, 14357-14373.

Dettinger, M.D., D.R. Cayan, and K.T. Redmond, 1999: United States streamflow probabilities based on forecasted La Niña, winter-spring 2000. *Exp. Long-Lead Forecast B., Cent. Land-Ocean-Atmos.*, **8(4)**, 57-61.

______, ______, G. J. McCabe, and J. A. Marengo, 2000: Multiscale streamflow variability associated with El Niño/Southern Oscillation. *El Niño and the Southern Oscillation--Multiscale Variability and Global and Regional Impacts.* H. F. Diaz and V. Markgraf, eds., Cambridge University Press, 113-146.

Deutsch, M., D. R. Wiesnet, and A. Rango, eds., 1979: *Satellite Hydrology*: *Fifth annual William T Pecora memorial symposium on remote sensing.* American Water Resources Association Technical Publication Series TPS81-1, St Anthony Falls Hydraulic Laboratory, Minneapolis, MI.

Doesken, N.J., T. B. McKee, and J. Kleist, 1991: *Development of a Surface Water Supply Index for the western United States*. Climatology report, 91-3. Colorado Climate Center, Department of Atmospheric Science Colorado State University, Ft Collins CO.

Douglas, E.M., R.M. Vogel, and C.N. Kroll, 2000: Trends in floods in the United States: impact of spatial correlation. *J. Hydrol.*, **240**, 90-105.

Dracup, J. A. 1977: Impact on the Colorado River Basin and Southwest water supply. *Climate, Climatic Change and Water Supply*, Panel on water and climate. National Research Council, Washington DC, 132pp.

______, D. L. Haynes, and S. D. Abramson, 1985: Accuracy of hydrologic forecasts. *Proc. Western Snow Conf.,* Boulder, CO, 13-24.

Drosdowsky, W., and L. Chambers, 1998: Near global sea surface temperature anomalies as predictors of Australian seasonal rainfall. Bureau of Meteorology Research Centre Research Report, 65.

Easterling, W.E., 1986: Subscribers to the NOAA monthly and seasonal weather outlook. *B. Am. Meteorol. Soc.,* 67(4), 402–410.

Eischeid, J.K., P. Pasteris, H.F. Diaz, M. Plantico, and N. Lott, 2000: Creating a serially complete, national daily time series of temperature and precipitation for the Western United States. *J. Appl. Meteorol.,* **39**, 1580-1591.

Enfield, D. B., A. M. Mestas-Nuñez, 1999: Multiscale variabilities in global sea surface temperatures and their relationships with tropospheric climate patterns. *J. Climate,* **12(9)**, 2719–2733.

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

______, 1988: Long range weather prediction: limits of predictability and beyond. *Weather Forecast.*, **3(1)**, 69-75.

Fassnacht, S. R., K. A. Dressler, and R. C. Bales, 2003: Snow water equivalent interpolation for the Colorado river basin from snow telemetry (SNOTEL) data. *Water Resour. Res.*, **39(8)**, 1208.

FEMA (Federal Emergency Management Agency), 1995: *National Mitigation Strategy; Partnerships for building safer communities.* Mitigation Directory, 2, Federal Emergency Management Agency, Washington DC.

FOCI (Fisheries-Oceanography Coordinated Investigations), 2001: *Ecosystem Indicators and Trends Used by FOCI.* [Available at http://www.pmel.noaa.gov/foci/ecotrends/Ecosystem_trends_2000for2001.pdf previously accessed 12/26/03].

Foster, D., 1997: Blame crests with Red River. *Rocky Mountain News*, April 28, p 47.

Fraser, J, 1999: Welcome. *Proc Climate Change and Salmon Stocks*, Vancouver, BC, Pacific Fisheries Resource Conservation Council [Available at http://www.fish.bc.ca/html/fish4012.htm previously accessed on 8/12/04].

Garen, D. C., 1992: Improved techniques in regression-based streamflow volume forecasting. *J. Water Res. Pl.-ASCE*, **118(6)**, 654-670.

______, 1993: Revised Surface Water Supply Index for western United States. *J. Water Res. Pl.-ASCE,* **119(4)**, 437-454.

______, 1998: ENSO indicators and long-lead climate outlooks: Usage in seasonal streamflow volume forecasting in the western United States. *Proc. Am. Geophysical Union Fall Meeting,* San Francisco, American Geophysical Union.

Gilman, D. L. 1982: The New Look of the Monthly and Seasonal Weather Outlook. *Proc. Climate Diagnostics Workshop,* Palisades, NY, National Oceanic and Atmospheric Administration.

______ 1985: Long-range rorecasting: the present and the future. *B. Am. Meteorol. Soc.,* **66(2)**, 159–164.

______, 1986: Expressing uncertainty in long-range forecasts. *Namias Symposium,* J.O. Roads, ed., Scripps Institution of Oceanography Reference Series, 86-17, La Jolla, CA, 174-187.

Gilovich, T., 1993: *How we know what isn't so: the fallibility of human reason in everyday life.* Free Press, 224 pp.

Goddard, L, and five others, 2000: *Current Approaches to Seasonal to Interannual Climate Predictions.* IRI Technical Reports, 00-01. International Research Institute for Climate Prediction, Palisades, NY.

Graham, N.E. and W.B. White, 1988: El Nino cycle: A natural oscillator of the Pacific ocean-atmosphere system. *Science*, **240**, 1293-1302.

Groisman, P. Ya., R. W. Knight, and T. R. Karl, 2001: Heavy precipitation and high streamflow in the contiguous United States: Trends in the Twentieth Century, *B. Am. Meteorol. Soc.*, **82(2)**, 219-246.

Gutzler, D. S., D. M. Kann, C. Thornbrugh, 2002: Modulation of ENSO-based long-lead outlooks of southwestern US winter precipitation by the Pacific Decadal Oscillation. *Weather Forecast.,* **17(6)**, 1163–1172.

Hamlet, A. F., and D. P. Lettenmaier, 1999: Columbia river streamflow forecasting based on ENSO and PDO climate signals. *J. Water Res. Pl.-ASCE.* **125(6),** 333-341.

______, and ______, 2000: Long range forecasting and its use for water management in the Pacific Northwest region of North America. *J. Hydroinformatics*, **2**, 163-182.

Hare, S. R., and N. J. Mantua, 2000: Empirical evidence for North Pacific regime shifts in 1977 and 1989. *Prog. Oceanogr.,* **47**,103-145.

Harshburger, B., H. Ye, and J. Dzialoski, 2002: Observational evidence of the influence of Pacific SSTs on winter precipitation and spring stream discharge in Idaho. *J. Hydrol.,* **264**, 157-169.

Hartmann, H. C. 2001: Stakeholder driven research in a hydroclimatic context. Ph.D. Dissertation, Department of Hydrology and Water Resources, University of Arizona.

______, R. Bales, and S. Sorooshian, 1999: *Weather, Climate and Hydrologic forecasting for the Southwest US.* Climas report series CL2-99, University of Arizona, Tucson, AZ.

______, T. C. Pagano, S. Sorooshian, and R. Bales, 2002a: Confidence builders: evaluating seasonal climate forecasts from user perspectives. *B. Am. Meteorol. Soc.*, **83**, 683-698.

______, R. Bales, and S. Sorooshian, 2002b: Weather, climate, and hydrologic forecasting for the US Southwest: a survey. *Climate Res.,* **21**, 239-258.

Hay, L. E., and seven others, 2002: Use of regional climate model output for hydrologic simulations. *J. Hydrometeorol.,* **3(5)**, 571–590.

Heidke, P., 1926: Berechnung des erfolges und der gute der windstarkevorhersagen im Sturmwarnungsdienst [Measures of success and goodness of strong wind forecasts by the gale-warning service]. *Geogr. Ann.,* **8**, 301-349.

Held, L.J., and J. J. Jacobs, 1990: Evaluating water supply outlook for fertilizing hay under variable streamflows. *J. Prod. Agric.*, 3(4), 429-435.

Helms, D., 1992: Snow surveying comes of age in the West. *Proc. Western Snow Conference.* Jackson WY, 10-17.

Hidalgo, H. G., and J. A. Dracup, 2003: ENSO and PDO effects on hydroclimatic variations in the Upper Colorado River Basin. *J. Hydrometeorol.,* **4**, 5-23.

Hsieh, W. W., Li. J. Yuval, and A. Shabbar, 2003: Seasonal prediction with error estimation of Columbia river streamflow in British Columbia. *J. Water Res. Pl.-ASCE,* **129(2)**, 146-149.

Huang, J., H. M. van den Dool, A. G. Barnston, 1996: Long-lead seasonal temperature prediction using Optimal Climate Normals. *J. Climate,* 9(4), 809–817.

Hurrell, J.W., Y. Kushnir, G. Ottersen, and M. Visbeck, eds., 2003: *The North Atlantic Oscillation: climate significance and environmental impact*. Geophysical Monograph Series, 134, 279pp.

IRI, 2001: Communication of climate forecast information. *Communication of Climate Forecast Information Workshop Proc.*, Palisades, N.Y. International Research Institute for Climate Prediction.

Ji, M., A. Kumar, and A. Leetmaa, 1994: A multiseason climate forecast system at the National Meteorological Center. *B. Am. Meteorol. Soc.,* **75(4)**, 569-577.

Johnson, L. F., 1960: Use of the electronic computer for streamflow analysis. *Proc. Western Snow Conf.*, Santa Fe, NM, 15-20.

Jones, E.B., B. A. Shafer, A. Rango, and D. M. Frick, 1981: Application of a snowmelt model to two drainage basins in colorado. *Proc. Western Snow Conf.*, St. George, UT, 43-54.

Jones, P.D., T. Jónsson, and D. Wheeler, 1997: Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and South-West Iceland. *Int. J. Climatol.,* **17**, 1433-1450.

Joyce, T. M., 2002: One hundred plus years of wintertime climate variability in the eastern United States. *J. Climate,* **15(9)**, 1076-1088.

Kahneman, D., P. Slovic, and A. Tversky, eds., 1982: *Judgement under uncertainty: heuristics and biases.* University of Cambridge Press, 544 pp.

Kahya, E., and J. A. Dracup, 1993: US streamflow patterns in relation to the El Niño/Southern Oscillation. *Water Resour. Res.*, **29(8)**, 2491-2503.

Keables, M. J. 1992: Spatial variability of midtropospheric circulation patterns and associated surface climate in the United States during ENSO winters. *Phys. Geogr.,* **13(4)**, 331-348.

Kim, Y-O, and R. N. Palmer, 1997: Value of seasonal flow forecasts in Bayesian stochastic programming. *J. Water Res. Pl.-ASCE,* **123(6)**, 327-335.

Koch, R. W., and A. R. Fisher, 2000: Effect of inter-annual and decadal-scale climate variability on winter and spring streamflow in western Oregon and Washington. *Proc. Western Snow Conf.*, Port Angeles, WA, 1-11.

Kohler, M. A., 1959: Preliminary report on evaluating the utility of water supply forecasting. *Proc. Western Snow Conf.*, Reno NV, 26-33.

Korzun, V. I., 1978: *World Water Balance and Water Resources of the Earth.* Studies and Reports in Hydrology, 25, UNESCO, English Translation; original, 1974.

Landscheidt, T., 2000a: River Po discharges and cycles of solar activity. *Hydrol. Sci. J.,* **45**, 491-493.

______, 2000b: Solar forcing of El Niño and La Niña. *The solar cycle and terrestrial climate*. M. Vazquez, and B. Schmieder, eds., European Space Agency Special Publication, 463, 135-140.

______, 2001a: *Solar eruptions linked to North Atlantic Oscillation.* [Available at http://www.john-daly.com/theodor/solarnao.htm last accessed 11/24/03].

______, 2001b: *Trends in Pacific Decadal Oscillation subjected to solar forcing*. [Available at http://www.john-daly.com/theodor/pdotrend.htm last accessed 11/24/03].

Laurmann, J. A., 1975: On the prospects for seasonal climate forecasting. *B. Am. Meteorol. Soc.,* **56(10)**, 1084-1088.

Lee, S., A. Klein, and T. Over, 2002: Effects of the El Nino/Southern Oscillation on temperature, precipitation, snow water equivalent and resulting streamflow in the Upper Rio Grande. *Proc. Eastern Snow Conf.*, Stowe, VT.

Leathers, D.J., B. Yarnal, and M. A. Palecki, 1991: The Pacific/North American teleconnection pattern and United States climate. Part I: regional temperature and precipitation associations. *J. Climate,* **4(5)**, 517-528.

Lettenmaier, D. P., 1984: Some issues in assessing the accuracy of hydrologic forecasts. *A critical assessment of forecasting in water quality goals in western water resources management,* J.J. Cassidy, and D. P. Lettenmaier, eds., American Water Resources Association,106-116.

______, 2004: The role of climate in water resources planning and management. *Water: Science, Policy and Management,* R. Lawford, et al., eds., Water Resources Monograph #16, American Geophysical Union, 247-266.

______, and D. C. Garen, 1979: Evaluation of streamflow forecasting methods. *Proc. Western Snow Conf.,* Sparks, NV, 48-55.

______, E.F. Wood, and J.R. Wallis, 1994: Hydro-climatological trends in the continental United States, 1948-1988. *J. Climate*, **7**, 586-607.

Leung, L. R., and M. S. Wigmosta, 1999: Potential climate change impacts on mountain watersheds in the Pacific Northwest. *J. Am. Water Resour. As.*, **35(6)**, 1463-1471.

Lewitt, S., 1995: *Momento Mori.* Toc Books, 286 pp.

Liles, C. 1999: Pacific Decadal Oscillation and New Mexico precipitation. [Available at http://www.srh.noaa.gov/abq/feature/PDO_NM.htm last accessed 11/24/03].

Lins, H., and J.R. Slack, 1999: Streamflow trends in the United States. *Geophys. Res. Lett.*, **26**, 227-230.

Livezey, R. E. 1990: Variability of skill of long-range forecasts and implications for their use and value. *B. Am. Meteorol. Soc.,* **71(3)**, 300-309.

______ and A. G. Barnston, 1988: An operational multifield analog/anti-analog prediction system for United States seasonal temperatures 1. System design and winter experiments. *J. Geophys. Res.,* 93(D9), 10953-10974.

______, ______, and B. K. Neumeister, 1990: Mixed analog/persistence prediction of seasonal mean temperatures for the USA. *Inter. J. Climatol.,* **10**, 329-340.

Madden, R.A., 1981: A quantitative approach to long-range prediction. *J. Geophys. Res.*, **86**, 9817-9825.

______, and P. R. Julian, 1994: Observations of the 40-50 day tropical oscillation: a review. *Mon. Weather. Rev.*, **122**, 814-837.

Mantua, N.J. and S. R. Hare, 2002: The Pacific Decadal Oscillation. *J. Oceanogr,* **58(1)**, 35-44.

______, ______, Y. Zhang, J.M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *B. Am. Meteorol. Soc.*, **78(6)**, 1069-1079.

Marr, J. C. 1936: Status of coordination and standardization of snow-surveying. *Trans Am Geophysical Un Meeting: Western Interstate Snow-Survey Conference Joint Meeting with Section of Hydrology*, Pasadena, CA, 530-533.

Marron. J. K., 1986: Parameter estimation for the precipitation runoff modeling system using snow telemetry system data. *Proc. Western Snow Conf.,* Phoenix, AZ ,154-157.

Marshall, J., and eight others, 2002: Atlantic climate variability. *Int. J. Climatol.,* **21**, 1863-1898.

Mason, S. J., and G. M. Mimmack, 2002: Comparison of some statistical methods of probabilistic forecasting of ENSO. *J. Climate*, **15**, 8-29.

______, A. M. Joubert, C. Cosijn, and S. J. Crimp, 1996: Review of seasonal forecasting techniques and their applicability to Southern Africa. *Water SA*, **22(3),** 203-209.

______, and five others, 1999: The IRI seasonal climate prediction system and the 1997/98 El Niño event. *B. Am. Meteorol. Soc.*, **80(9)**, 1853–1874.

Mauget, S. A., 2003: Multidecadal regime shifts in U.S. streamflow, precipitation, and temperature at the end of the Twentieth Century. *J. Climate*, **16(23)**, 3905-3916.

Maxwell, K. D., and V. P. Holbrook, 2002: *Pacific Decadal Oscillation and Arizona precipitation.* Western Regional Technical Attachment 02-08. Western Regional Headquarters National Weather Service. [available at http://www.wrh.noaa.gov/wrhq/02TAs/0208/ last accessed 11/24/03].

McCabe, G. J. and D. R. Legates, 1995: Relationships between 700 hPa height anomalies and 1 April snowpack accumulations in the Western USA. *Int. J. Climatol.,* **15**, 517-530.

______, and M. D. Dettinger, 1999: Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States. *Int. J. Climatol.,* **19**, 1399-1410.

______, and ______, 2002: Primary modes of predictability of year-to-year snowpack variations in the western United States from teleconnections with Pacific Ocean climate. *J. Hydrometeorol*, **3**, 13-25.

______, and D. M. Wolock, 1999: Effects of potential climatic change on annual runoff in the conterminous United States. *J. Amer. Water Resour. As.*, **35**, 1341-1350.

______, and ______, 2002: A Step Increase in Streamflow in the Conterminous United States, *Geophys. Res. Lett.*, **29(24)**, 2185.

______, M. A. Palecki, and J. L. Betancourt, 2004: Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *P. Natl. Acad. Sci. USA*, **101**, 4136-4141.

McFarlane, G.A., J.R. King, and R.J. Beamish, 2000: Have there been recent changes in climate? Ask the fish. *Prog. Oceanogr*., **47(2-4)**, 147-169.

Medford Mail Tribune, 1959: Forecasts of water supply for southwest region vary. *Medford Mail Tribune*, Medford OR. February 8, 1959.

Meinke, H and R. C. Stone, 2004: Seasonal and interannual climate forecasting: a new tool for increasing preparedness to climate variability and change in agricultural planning and operations. *Climatic Change* (submitted).

Modini, G. C., 2000: Long-lead precipitation outlook augmentation of multi-variate linear regression streamflow forecasts. *Proc. Western Snow Conf.*, Port Angeles, WA, 57-68.

Mote, P. W., 2003a: Trends in temperature and precipitation in the Pacific Northwest during the twentieth century. *Northwest Sci.*, **77(4)**, 271-282.

______, 2003b: Trends in snow water equivalent in the Pacific Northwest and their climatic causes. *Geophys. Res. Lett.,* **30(12)**, 1601.

______, 2005: Declining mountain snowpack in western North America. *B. Am. Meteorol. Soc*. **86(1)**, 39-49.

Murphy, A.H., 1978: Hedging and the mode of expression of weather forecasts. *B. Am. Meteorol. Soc*., **59**, 371-373.

______, 1993: What is a good forecast? An essay on nature of goodness in weather forecasting. *Weather Forecast.*, **8**, 281-293.

______ and E.S. Epstein, 1967: A note on probability forecasts and "hedging". *J. Appl. Meteorol*. **6**, 1002-1004.

Myers, M. F., 1997: Trends in floods. *Proc. Workshop on the Societal and Economic Impacts of Weather,* Boulder, CO, National Center for Atmospheric Research, 77-86.

Namias, J. 1962: Influences of abnormal heat sources and sinks on atmospheric behavior. *Proc. Int. Symp. on Numerical Weather Prediction,* Tokyo, Japan, Meteor. Soc. Japan, 615- 627.

______, 1965: Macroscopic association between mean monthly sea surface temperature and the overlying winds. *J. Geophys. Res.*, **70**, 2307-2318

______, 1968: Long range weather forecasting – history, current status and outlook. *B. Am. Meteorol. Soc.,* **49(5)**, 438-470.

______, 1985a: Remarks on the potential for long-range forecasting. *B. Am. Meteorol. Soc.*, **66(2)**, 165-165.

______, 1985b: Some empirical evidence for the influence of snow cover on temperature and precipitation. *Mon. Weather. Rev.*, **113**, 1542-1553.

______, and D. R. Cayan, 1984: El Nino: implications for forecasting (USA). *Oceanus*, **27(2),** 41-47.

Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models, Part I: A discussion of principles. *J Hydrol.,* **10**, 282-290.

NAST (National Assessment Synthesis Team), 2000: The potential consequences of climate variability and Change: overview water. *Climate Change Impacts on the United States*, US Global Change Research Program, 6 pp.

Naujokat, B., 1986: An update of the observed quasi-biennial oscillation of the stratospheric winds over the tropics. *J. Atmos. Sci.,* **43**, 1873-1877.

Neal, E.G., M. T. Walter, and C. Coffeen, 2002: Linking the pacific decadal oscilation to seasonal stream discharge patterns in Southeast Alaska. *J. Hydrol.,* **262**, 188-197.

Newman, M., G. P. Compo, and M. A. Alexander, 2003: ENSO-forced variability of the Pacific Decadal Oscillation. *J. Climate,* **16**, 3853-3857.

Nicholls, N., 1980: Long range weather forecasting: value, status, and prospects. *Rev. Geophys. Space Ge.,* **18(4)**, 771-788.

______, 1999: Cognitive illusions, heuristics, and climate prediction. *B. Am. Meteorol. Soc.,* **80(7)**, 1385-1397.

Nielsen, J., 1994: *Usability Engineering*. Morgan Kaufmann, 362 pp.

NRC, 1999: *Making climate forecasts matter*. Panel on the Human Dimensions of Seasonal-to-Interannual Climate Variability, National Research Council. National Academy Press, 192 pp.

______, 2001: *A climate services division: first steps towards the future*. Board on Atmospheric Sciences and Climate, National Research Council. National Academy Press, Washington DC.

NRCS, 1997: *Southern Oscillation Index statistical correlation with spring runoff.* [Available at http://www.wrcc.dri.edu/enso/soiwsf2.pdf accessed 1/21/03].

______, 2004: *Guidelines for Development of Statistical Streamflow Forecasting Models.* USDA Natural Resources Conservation Service, National Water and Climate Center, Portland, OR.

O'Grady, K., and L. Shabman, 1990: Communicating the probability of Great Lakes water levels and storms. *Proc. Great Lakes Water Level Forecasting and Statistics Symp.,* H. Hartmann and M. Donahue, eds., Ann Arbor, MI, Great Lakes Commission, 197-204.

Pagano, T. C. and D. C. Garen, 2005a: Integration of climate information and forecasts into western US water supply forecasts. ASCE Monograph on the use of climate information in water management. (accepted)

______, and ______, 2005b: A recent increase in western US streamflow variability and persistence. *J. Hydrometeorol.* (accepted).

______, H. C. Hartmann, S. Sorooshian, and R. C. Bales, 1999: *Advances in seasonal forecasting for water management in Arizona: A case study of the 1997-98 El Niño*. Department of Water Resources, University of Arizona, HWR 99-040, 221 pp.

______, ______, and ______, 2000: Climate forecasts: A new tool for hazard management in the southwestern US. *Nat Hazards Observer*, **24(6)**, 7-8.

______, ______, and ______, 2001: Using climate forecasts for water management. *J. Am. Water Resour. As.*, **37(5)**, 1139-1153.

______, ______, and ______, 2002: Factors affecting seasonal forecast use in Arizona water management: A case study of the 1997-98 El Niño. *Climate Research*, **21**, 259-269.

______, P. Pasteris, M. Dettinger, D. Cayan, and K. T. Redmond, 2004a: Water year 2004: Western water managers feel the heat. *EOS T. Am. Geophys. Un.*, **85(40)**, 385-400.

______, D. C. Garen, and S. Sorooshian, 2004b: Evaluation of official Western US seasonal water supply outlooks, 1922-2002. *J. Hydrometeorol.*, **5**, 896-909.

Paget, F. H., 1940: Comparison of forecast and actual results, 1939. *EOS T. Am. Geophys. Un.*, **I-B**, 99-108.

Papineau, J.M., 2001: Wintertime temperature anomalies in Alaska correlated with ENSO and PDO. *Int. J. Climatol.*, **21**, 1577-1592.

Perica, S., D.-J. Seo, E. Welles, and J. Schaake, 2000: Simulation of precipitation fields from probabilisitic quantitative precipitation forecast. *J. Hydrol.*, **239**, 203-229.

Perkins, T. R., 1988: Seasonal streamflow forecasting in the Upper Rio Grande Basin by incorporating the use of SNOTEL data in the SSARR hydrologic model. *Proc. Western Snow Conf.*, Kalispell, MT, 58-67.

Peters, H. J., 1984: Long-range streamflow forecasting: A state agency perspective. *A critical assessment of forecasting in water quality goals in western water resources management*, J.J. Cassidy, and D. P. Lettenmaier, eds., American Water Resources Association, 3-10.

Peterson, T.C., and R. S. Vose, 1997: An overview of the Global Historical Climatology Network temperature database. *B. Am. Meteorol. Soc.*, **78**, 2837-2849.

Philander, S.G.H, 1992: Ocean atmosphere interactions in the tropics- a review of recent theories and models. *J. Appl. Meteorol.,* **31(8)**, 938-945.

Piechota, T. and J. A. Dracup, 1996: Drought and regional hydrologic variation in the United States: associations with El Nino/Southern Oscillation," *Water Resour. Res.*, **32(5)**, 1359-1373.

______, ______, and R. G. Fovell, 1997: Western U.S. streamflow and atmospheric circulation patterns during El Niño-Southern Oscillation (ENSO). *J. Hydrol.*, **201**, 249-271.

Pielke Jr, R. A. 1999: Who decides? Forecasts and responsibilities in the 1997 Red River Flood. *Appl. Behav. Sci. Rev.*, **7(2)**, 83-101.

______, and M.W. Downton, 1999: US trends in streamflow and precipitation: Using societal impact data to address an apparent paradox. *B. Am. Meteorol. Soc.,* **80**, 1435-1436.

Plous, S., 1993: *The psychology of judgement and decision making.* McGraw-Hill, 352 pp.

Pollack, H. N., 2003: *Uncertain science ... uncertain world.* Cambridge University Press, 256 pp.

Poulton, H. J. 1964: *James Edward Church: Bibliography of a Snow Scientist.* Bibliographical Series, 4, University of Nevada Press.

Preisendorfer, R.W. and C.D. Mobley, 1984: Climate forecast verifications, United States mainland 1974-83. Mon. Weather Rev., **112**, 809-825.

Pulwarty, R., and K. T. Redmond, 1997: Climate and salmon restoration on the Columbia River Basin: the role and usability of seasonal forecasts. *B. Am. Meteorol. Soc.*, **78(3)**, 381-397.

______, and Melis, T., 2001: Climate extremes and adaptive management on the Colorado River. *J. Environ. Manage.*, **63(3)**, 307-324.

Rasmusson, E. M., 1984: El Nino: the ocean/atmosphere connection. *Oceanus*, **27(2)**, 5-12.

______, and T. H. Carpenter, 1982: Variations in tropical sea surface temperatures and surface wind fields associated with the Southern Oscillation/El Nino. *Mon. Weather Rev.*, **110(5)**, 354-384.

Redmond, K. T., 2003: Climate Variability in the West: Complex spatial structure associated with topography, and observational issues. *Water and Climate in the Western United States,* W. M Lewis Jr., ed., University Press of Colorado, 29-48.

______, and R. W. Koch, 1991: Surface climate and streamflow variability in the western United States and their relationship to large scale circulation indices. *Water Resour. Res*, **27(9)**, 2381-2399.

Reek, T., S. R. Doty, and T. W. Owen, 1992: A deterministic approach to the validation of historical daily temperature and precipitation data from the Cooperative Network. *B. Am. Meteorol. Soc.*, **73**, 753-762.

Reigle, D., 1998: Weather and climate forecasting at SRP. *Forecast assessment workshop*, Tucson, AZ, Southwest Climate Assessment Project, Institute for the Study of Planet Earth, University of Arizona.

Rhodes, S., D. Ely, and J. Dracup, 1984: Climate and the Colorado River: Limits to management. *B. Am. Meteorol. Soc.*, **65(7)**, 682-691.

Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño Southern Oscillation (ENSO). *Mon. Weather. Rev.*, **114**, 2352-2362.

______, and ______, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Weather Rev.*, **115**, 1606-1626.

Sarewitz, D., R. A. Pielke Jr., and B. Radford, eds., 2000: *Prediction: Science, decision making, and the future of nature*. Island Press, 405 pp.

Schaake, J. C., 1978: The National Weather Service extended streamflow prediction techniques: description and applications during 1977. *Climate Diagnostics Workshop*, Miami, Fl, National Oceanic and Atmospheric Administration.

______, and E. L. Peck, 1985: Analysis of water supply forecast accuracy. *Proc. Western Snow Conf.*, Boulder, CO, 44-53.

Schneider, J. M., and J. D. Garbrecht, 2003: Temporal disaggregation of probabilistic seasonal climate forecasts. *Proc American Meteorological Society 14th symposium on global change and climate variations*, Long Beach, CA, American Meterological Society.

Schwing, F.B. and W.T. Peterson, 2003: A new climate change in northeast pacific ecosystems, *Geophys. Res. Lett.,* **30(17)**, 1986.

SCS, 1988: *Snow survey and water supply forecasting program: productivity improvement program.* Soil Conservation Service (Natural Resources Conservation Service), 87 pp.

Serreze M.C, M. P. Clark, R. L. Armstrong, D. A. McGinnis, and R. L. Pulwarty, 1999. Characteristics of the Western US snowpack from snowpack telemetry (SNOTEL) data. *Water Resour. Res.,* **35**, 2145-2160.

______, ______, and A. Frei, 2000: Characteristics of large snowfall events in the montane western United States as examined using snowpack telemetry (SNOTEL) data. *Water Resour. Res.,* **37**, 675-688.

Shafer, B. A. and L. E. Dezman, 1982: Development of a surface water supply index (SWSI) to assess the severity of drought conditions in snowpack runoff areas. *Proc. Western Snow Conf.,* Reno, NV, 164-175.

______, and J. M. Huddleston, 1984: Analysis of seasonal volume streamflow forecast errors in the western United States. *A critical assessment of forecasting in water quality goals in western water resources management,* J.J. Cassidy, and D. P. Lettenmaier, eds., American Water Resources Association, 117-126.

Shukla, J and thirteen others, 2000: Dynamical seasonal prediction. *B. Am. Meteorol. Soc.,* **81(11)**, 2593-2606.

Skindlov, J., D. Phillips, and C. Dempsey, 2000: Seasonal precursors to winter runoff and summer rain on the Salt-Verde watershed. *Southwest Weather Symposium*, Tucson, AZ, University of Arizona Department of Atmospheric Sciences.

Slack, J. R., and J. M. Landwehr, J. M., 1992: *Hydro-Climatic Data Network: A U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1974-1988.* U.S. Geological Survey Report. 92-129, 193 pp.

Smith, S. R. and J. J. O'Brien, 2001: Regional snowfall distributions associated with ENSO: Implications for seasonal forecasting. *B. Am. Meteorol. Soc.,* **82(6),** 1179-1192.

Sonka, S. T., P. J. Lamb, S. A. Changnon, and A. Wiboonpongse, 1982: Can climate forecasts for the growing season be valuable to crop producers: some general considerations and an Illinois pilot study. *J. Appl. Meteorol.,* **21(4),** 471–476,

Taylor, G. H., C. Daly, G. Johnson and P. Pasteris, 2004: Trends in snowfall and snow water equivalent in the Pacific Northwest and their relation to temperature and precipitation variations. *Mountain Climate Science Symposium*, Lake Tahoe, CA, USDA Forest Service. [Available at http://www.x-cd.com/mcss04/P64.html accessed 8/11/04].

Thompson, D. W. J, and J. M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297-1300.

Torrence, C., and G. P. Compo, 1998: A practical guide to wavelet analysis. *B. Am. Meteorol. Soc.,* **79(1)**, 61-78.

Tufte, E. R., 2001: *The visual display of quantitative information*. 2d ed., Graphic Press, 197 pp.

Unger, D. A., 1996a: Long lead climate prediction using screening multiple linear regression. *Proc. Climate Diagnostics Workshop.* Seattle, WA, National Oceanic and Atmospheric Administration, 425-428.

______, 1996b: Skill assessment strategies for screening regression predictions based on a small sample size. *Preprints, Thirteenth Conference on Probability and Statistics in the Atmospheric Sciences.* San Francisco, CA., American Meterological Society, 260-267.

Van den Dool, H. M. and Toth, Z., 1991: Why do forecasts for “near normal” often fail? *Weather Forecast.,* **6**, 76-85.

______, J. Huang, and Y. Fan, 2003: Performance and analysis of the constructed analogue method applied to US soil moisture over 1981–2001. *J. Geophys. Res.,* **108(D16),** 8617.

Vogel, R. M., Y. Tsai, and J. F. Limbrunner, 1998: The regional persistence and variability of annual streamflow in the United States. Water Resour. Res., **34(12)**, 3445-3459.

Voortman, R.L., 1998: Recent historical climate change and its effect on land use in the eastern part of West Africa. *Phys. Chem. Earth*, **23**, 385-391.

Wagner, A.J., 1989: Medium and long-range forecasting. *Weather Forecast.*, **4(3)**, 413-426.

Walker, G.T., and E. W. Bliss, 1934: World weather V. *Mem. Roy. Meteorol. Soc.*, **4(36)**, 53-84.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere Winter. *Mon. Weather Rev.,* **109**, 784-812.

Webb, R. H. and J. L. Betancourt, 1992: *Climatic variability and flood frequency of the Santa Cruz River, Pima County, Arizona.* Water-Supply Paper 2379. US Geologic Survey, 40 pp.

Werner, K., D. Brandon, M. Clark, and S. Gangopadhyay, 2004: An analysis of weighting schemes using climate indices for seasonal volume forecasts produced from the ensemble streamflow prediction system of the National Weather Service. *J. Hydrometeorol.*, **5(6)**, 1076-1090.

Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences: An introduction*. Academic Press, 467 pp.

______, 2000a: On interpretation of probabilistic climate forecasts. *J. Climate*, **13**, 1965-1971.

______, 2000b: Diagnostic verification of the climate prediction center long-lead outlooks, 1995-98. *J. Climate*, **13**, 2389-2403.

______, and K. L. Eggleston, 1992: Estimating Monthly and Seasonal Precipitation Distributions Using the 30- and 90-Day Outlooks. *J. Climate*, **5(3)**, 252–259.

Woodhouse, C. A., 1997: Winter climate and atmospheric circulation patterns in the Sonoran Desert region. *Int. J. Climatol.*, **17(8)**, 859-873.

Work, R. A., 1940: Standards and terminology. *EOS T. Am. Geophys. Un.,* **III-B**, 979-981.

______, and R. T. Beaumont, 1958: Basic data characteristics in relation to runoff forecast accuracy. Proc Western Snow Conf., Bozeman MT, 45-53.

Young, K. C., and R. L. Gall, 1992: A streamflow forecast model for central Arizona. *J. Climate*, **31**, 465-479.

Zettlemoyer, R. H., 1982: *Experimental water supply forecasts based on seasonal precipitation forecasts*. California Department of Water Resources.