SEASONAL STOCHASTIC STREAMFLOW FORECASTS FOR THE YAKIMA RIVER BASIN AND IMPLICATIONS TO SALMON SURVIVAL AND STREAMFLOW MANAGEMENT

by

SARAH CORRINE STAPLETON

B.S. Phy., Creighton University, 2002

A thesis submitted to the

University of Colorado in partial fulfillment

for the requirement for the degree of

Master of Science

Department of Civil, Environmental, and Architectural Engineering

2004

This thesis entitled:

Seasonal Stochastic Streamflow Forecasts

for the Yakima River Basin and Implications for

Salmon Survival and Streamflow Management

written by Sarah C. Stapleton

has been approved for the Department of Civil, Environmental, and

Architectural Engineering

_____

Diane McKnight

_____

Balaji Rajagopalan

_____

Edith Zagona

_____

Date

The final copy of this thesis has been examined by the signatories,
and we find that both the content and the form meet acceptable presentation
standards of scholarly work in the above mentioned discipline.

Stapleton, Sarah (M.S., Civil, Environmental, and Architectural Engineering)

Seasonal Stochastic Streamflow Forecasts for the Yakima River Basin and Implications to Salmon Survival and Streamflow Management

Thesis directed by Assistant Professor Diane McKnight

The Yakima River Basin is an arid basin with competing water needs such as instream fish flows for salmon, irrigation demand, and flood control. Water managers attempt to balance demands by relying on forecasts of the anticipated water supply for the upcoming water year. Current forecasting techniques employ multiple linear regression and use anticipated precipitation. Water managers are seeking methods that improve forecast skill and increase the lead time with which the forecasts can be issued. This thesis attempts to improve the forecasts by incorporating large-scale climate information and using a local regression technique that does not make assumptions about the underlying data distribution. The usefulness of these forecasts to water managers is discussed, as well as implications for salmon survival in the basin.

Acknowledgements

There are many individuals whose support, guidance, and advice, enable my research. I would like to thank my advisors: Diane McKnight, Balaji Rajagopalan, and Edith Zagona. It was Diane McKnight's work for the National Research Council examining the issue of salmon survival in the Columbia River Basin that prompted this choice of a research topic. Balaji Rajagopalan's Hydrology Class sparked my interest in climatology and streamflow forecasting. Martyn Clark and Suhbrendu Gangopadyay also provided enormous support with streamflow forecasting techniques. Edith Zagona provided much guidance about basin policy and modeling.

There are four individuals from the US Bureau of Reclamation who readily answered my questions and provided models and forecasts: Roger Sonnichsen, Warren Sharp, Christopher Lynch, and Jeff Rieker. While I did not use their data sets, three individuals from the US Geological Survey provided me with much insight into the Yakima River Basin: Laren Hayes, Mark Matsin, and John Vaccaro.

At CADSWES, the following people helped me with modeling: Katrina Grantz, Satish Regonda, James Prairie, J.D. Emmert, and Bill Oakley. At INSTAAR: the following people provided much advice: Rose Corey, Marcella Appel, Chi Yang, and Christopher Jaros.

Finally, this thesis would not have been possible without the support and coding advice of my fiancé, Jan Weiss. I would also like to thank my family and friends.

Contents

Tables

Figures

Chapter 1

Introduction

The Yakima River Basin, located in south central Washington State, exemplifies the difficulties facing water managers in many basins. Competing water uses, legal battles, and over-allocation have presented water managers in the basin with the challenge of meeting demand, anticipating future demand, and preparing for potential impacts of climate change. The basin is arid because it is located in the rain shadow of Mt. Rainier. Despite being dry, the area has become the nation's primary producer of pears, hops, mint, apples, and other soft fruits. Several wineries, such as the Kiona winery (Walton, 2000) have produced notable crops.

In recent years, communities in the basin have had to grapple with the decline of another economic livelihood, salmon. The Yakima basin once provided spawning grounds to numerous genetically distinct salmon stocks. The Yakama Nation, a sovereign Native American tribe living in the basin, derives its cultural identity and way of life from fishing the salmon. The decrease of the salmon returning to spawn is a problem facing the entire Pacific Northwest region. The factors contributing to the salmon's demise are many, including historical fishing practices, agricultural pollution, and the alteration of the Yakima streamflow regimes by people. Natural environmental variability, such as the cyclical nature of various ocean conditions and streamflow variations, has also affected salmon populations. A growing awareness of the problem facing salmon stocks and other species has prompted a shift in water management.

The United States Bureau of Reclamation is responsible for managing the majority of water in the Yakima Basin for agricultural use and instream flows. The past two decades have seen a shift in policy from the priorities of flood control, meeting demand, and power

generation to attempting to provide instream flows to protect the salmon. These competing uses are exacerbated by the problem that current storage capacity is able to hold only enough water for one operational year. Thus, the ability to forecast streamflow for the upcoming water season is of the utmost importance for managing resources during the water year. This thesis discusses the implications of a forecasting technique for water operations planning.

1.1 Site Description

The Yakima River begins its journey on the slopes of the Eastern Cascades in south central Washington. The river meanders some 215 miles through diverse terrain of mountains and lowlands until its confluence with the lower Columbia River near Richland, Washington. It is a major tributary of the Columbia River. The basin drains an area of 6,150 square miles. A map of the basin is shown in Figure 1.1.

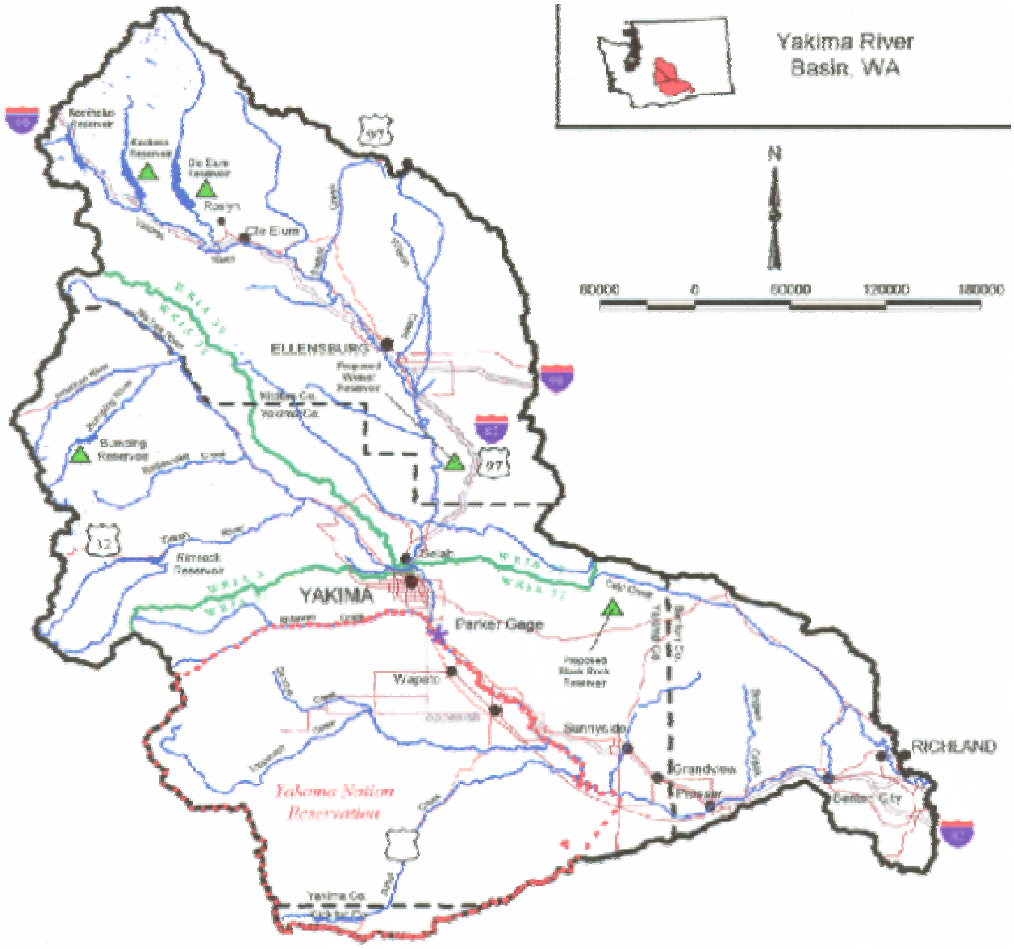


Figure 1.1: The Yakima River Basin

The topography of the basin varies greatly. The headwater regions are mountainous and encompass four natural lakes that were dammed by the Bureau of Reclamation (USBR, 2004) at the beginning of the 20th century to meet increasing irrigation demands. Pine and spruce forests dominate the headwaters. Land use in the nascent region is predominantly recreation, wildlife habitat, logging, and cattle grazing. The lower reaches of the river open into arid, flat plains that are heavily cultivated for agriculture.

The basin can be effectively divided into upper and lower halves based on the amount of precipitation each receives. Spatial variations in precipitation are large. Mt. Rainier is located just west of the basin and has significant orographic effects. Storm tracks from the Pacific Ocean split around the mountain and create an uneven distribution of precipitation. Overall, the whole basin receives an annual mean of 27 inches of precipitation (Matsin and Vaccaro, 2002). However, the majority of precipitation is concentrated along the mountainous western rim of the upper half of the basin. The lower half of the basin below Parker, WA, averages 8 inches of precipitation a year. Figure 1.2 presents the precipitation distribution in the basin.

Streamflow in the basin is predominantly snowmelt driven, though late winter rain-on-snow events can occur and cause rapid melting of the snow at some of the lower elevations in the Cascades. A large proportion, approximately 60 to 80 percent, of the annual precipitation falls as snow during the period from November to March. Water from snowmelt is diverted and stored in five main USBR reservoirs in order to meet agricultural, industrial, and domestic consumption. The five reservoirs hold a combined 1,070,700 acre-feet of water, which is enough storage for one operational water year only. Water managed by the USBR falls under the Yakima Project, which was initiated in 1905. The project provides irrigation water for close to 465,000 acres of land in the entire basin.

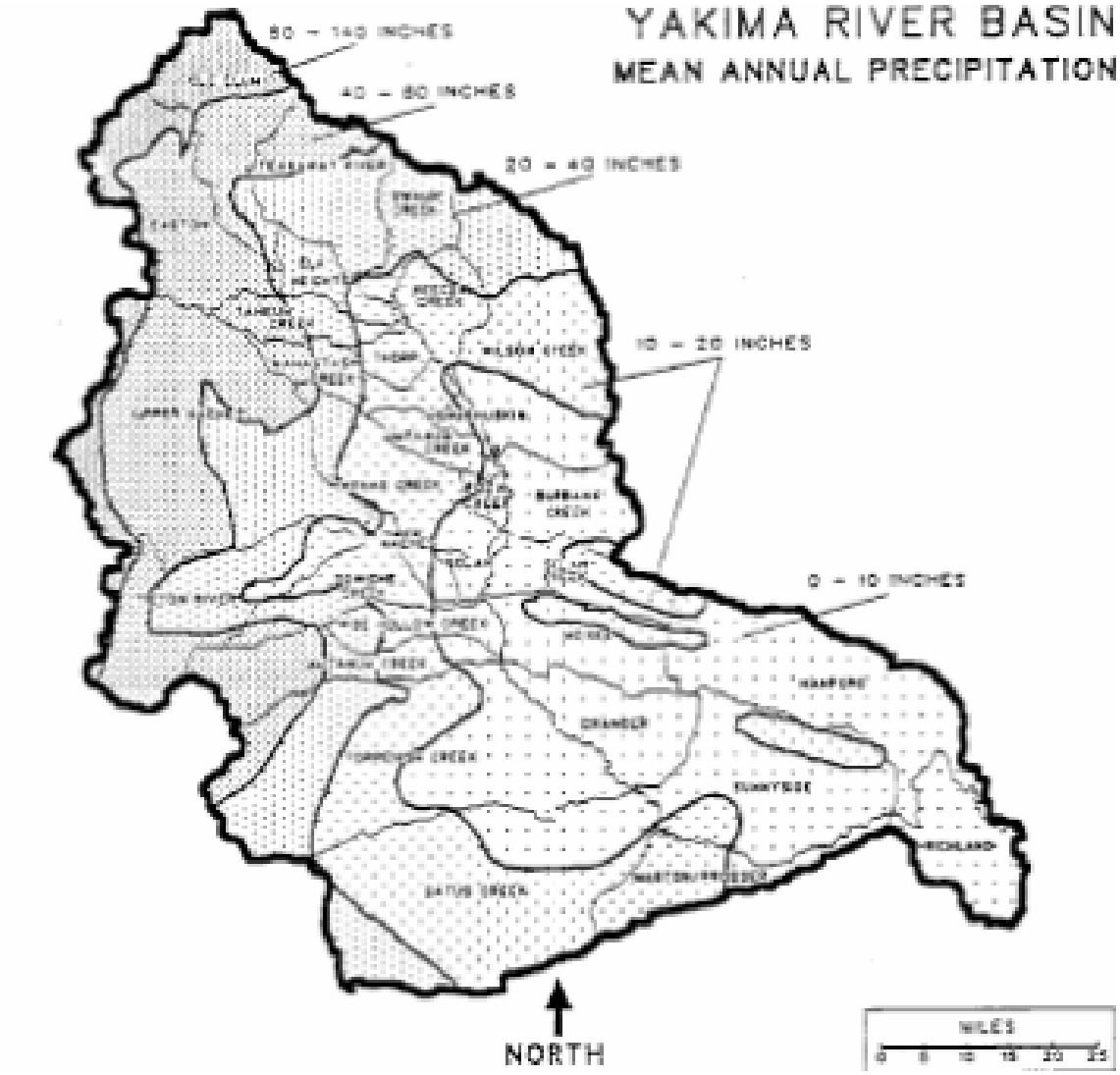


Figure 1.2: Precipitation Distribution Map

1.1.1 The Upper Subbasin

Precipitation is concentrated primarily in the more mountainous upper portion of the basin where elevations range from 8,000 to 2,000 feet. The headwater reaches average approximately between 80 to 140 inches of precipitation primarily as snow or rain on snow from November through March. The western rim of the basin, located south of the headwater reaches, receives 40 to 80 inches of precipitation on an annual basis.

Four natural glacial lakes: Kachess, Keechelus, Cle Elum, and Bumping, provided important natural storage to the Yakima River. The Yakima begins at the outflow of Keechelus Lake and reservoir. There are three major tributaries in the upper basin: Naches River, Teanaway River, and the Cle Elum River. The USBR took advantage of the four natural lakes and had dammed them by 1933. The fifth storage reservoir, Tieton (also known as Rimrock), was created by flooding McAllister Meadows in 1926 (IOP, 2002). The elevations of the five reservoirs range from approximately 2,000 feet to 3,400 feet (USGS, 2004). The five main reservoirs have the storage capacity to capture roughly one third of the total annual runoff volume of the Yakima River.

The majority of the land in the upper region is mountainous and forested and provides recreational opportunities and natural resources. Some cattle rangeland is also present in the upper subbasin. The land suitable for agricultural cultivation is located in the floodplain of the mainstem Yakima and as the mountains open onto flatter land. There are three irrigation divisions located in these areas: Kittitas, Easton, and Tieton. The Kittitas Division requires irrigation for 59,000 acres of land lying north and south of the Yakima River past Thorp, WA. The Tieton Division diverts water to irrigate some 28,000 acres of land west of the city of Yakima between Naches River and Ahtanum Creek.

1.1.2 The Lower Subbasin

The lower subbasin lies within the rain shadow of Mt. Rainier. Elevations range from 2,000 to 340 feet at the confluence with the Columbia River. Major tributaries to the Yakima in the lower subbasin are Toppenish and Satus Creeks. Precipitation is significantly reduced in this area. Yet, much of the arable land is located in this arid region. The USBR provides water to four irrigation divisions in the lower basin: Kennewick, Roza, Wapato and Sunnyside.

The Wapato Division is operated by the Bureau of Indian Affairs for the Yakama Nation whose reservation encompasses much of the land in the southwestern portion of the

basin. Figure 1.1 demarcates the boundaries of the reservation with a dotted red line. However, much of the 136,000 acres of irrigated land in the Wapato Division receives its water from the Yakima Irrigation Project. The Sunnyside Division is the next largest consumer of irrigation water in the lower basin. This division consists of approximately 103,000 acres of land north of the river between Parker and Benton City, Wash. The Roza Division irrigates nearly 72,500 acres of land north of the river from Pomona to Benton City. The Kennewick Division is the smallest, requiring water for some 19,000 acres and the Chandler Power plant. A combined 330,500 acres of land are irrigated in the lower basin, which is nearly three-fourths of the irrigated land in the entire basin (USBR, 2004).

1.1.3 Fishery Habitat

Historically, the Yakima River and its tributaries provided spawning habitat for six genetically distinct anadromous salmonids: spring, summer, and fall chinook, summer steelhead, early run coho and sockeye (Yakima River Subbasin Plan, 2004, from here referred to as Subbasin Plan). Numerous genetically distinct stocks of each of the six salmon species developed to adapt to conditions in different regions of the basin. Prior to large scale development and alteration of the natural flow regime, it is estimated that the total wild salmon runs ranged from 289,000 fish (Kreeger and McNeil, 1993 as cited in Subbasin Plan) to approximately 800,000 fish. The Yakama Nation, and other Native American tribes, centered their economy and way of life around the life cycle of the salmon.

The lifestyle of the Native American tribes and the life cycle of the salmon were seriously disrupted with the inception of mass immigration to the American West. The railroad connection between California and Washington and the Homestead Act of 1862 provided the mean and motive for large-scale settlement of the Pacific Northwest. Several other federal mandates: the 1866 and 1870 Mining Acts and the 1877 Desert Land Act began the process of legalizing the appropriation of water to make homesteads viable. The Desert Land Act in particular, provided for the appropriation of water to lands that were not

appurtanant to the water body in the arid portions of the west, including those in the state of Washington (Getches, 1997). The influx of large numbers of people led to over-fishing and alteration of the natural flow regime to meet agricultural and domestic demand. It was this demand for water in the arid Yakima basin that directly contributed to the anadromous fish decline. Current wild runs are estimated to be only 1% to 3% of the historic return (Subbasin Plan, 2004).

The upper Yakima basin, in particular, once provided important spawning and rearing habitat for several anadromous fish species. Sockeye salmon, bull trout and steelhead once spawned in the headwater reaches of the Yakima (IOP, 2002). Sockeye salmon, summer chinook, and coho salmon have been locally extirpated from the entire basin. While there are currently no salmon species listed as endangered on the ESA, bull trout and steelhead are listed as Threatened (ESA, 2004). Bull trout and the locally extirpated species once spawned in the upper basin; however, the reservoirs were constructed without fish ladders, blocking access to the headwater regions. Access to spawning beds in the upper subbasin has also been blocked by numerous agricultural diversions that lack fishways and smaller district dams that are too high for the fish to jump (Subbasin Plan, 2004). Some fluvial bull trout populations have been observed in Teanaway River and Ahtanum Creek, but not in the mainstem river or other tributaries (WDFW, 1998). With the inception of the dams, some lentic populations of bull trout were trapped in four reservoirs and have persisted.

Spring chinook have been observed spawning in upper reaches of the basin between the Cle Elum River and the Teanaway River. This prompted the Quakenbush Decision in 1980. There were fears that as reservoir releases wound down with the end of the irrigation season, the salmon redds would be dewatered in these areas. The decision implemented a flip-flop in releases between the upper three reservoirs and the lower two reservoirs. Through about the first week of September, the majority of irrigation demand is met through releases from the upper three reservoirs. Releases from the lower reservoirs are minimized as much as

possible during this time. After this period, a gradual shift or flip-flop in the releases occurs such that the lower reservoirs meet the remaining irrigation demand. Water is retained in the upper reservoirs to provide enough flows to keep redds covered in these reaches during the winter (IOP, 2002).

Currently, much of the suitable habitat for anadromous fish spawning and rearing is located in the lower subbasin. Low instream flows and fluctuating flows present the greatest flow impediment to salmon recovery. Low flows elevate water temperatures and increase toxin concentrations (metals, PCBs, DDT, and pesticides) to levels that are fatal to fish (Watershed Assessment 2003, Ecology 2003, EPA 1998). High sediment loads also have been noted during winter/spring runoff or when large releases from the reservoirs are made. Severe channelization has reduced the amount of suitable streambed for salmon redds (Watershed Assessment, 2003).

1.2 Basin History

By 1902, nearly one-fifth of the land currently irrigated was used for agriculture and demanded a consistent water supply (SOAC, 1999). Many of the tributaries and large portions of the mainstem were over-appropriated to the extent that some sections would dry up by mid-summer (Subbasin Plan, 2003). Congress passed the Reclamation Act in 1902 and created the Bureau of Reclamation with the intent of providing irrigation water to those farms established under the Homestead Act. Water delivered to users under the Reclamation Act is subordinate to existing state laws unless the state laws are directly in conflict with a provision of the act (Getches, 1997). The state of Washington adopted a hybrid approach to the issuance of water rights. Many of the earliest water rights are riparian in nature in the water abundant portions of the state. Riparian rights, which require the water claimed to be appurtunant to the land on which it is applied, did not seem applicable in the arid Yakima Basin, and prior appropriation became the dominant water law of the area. Under the Water

Code of 1917, Washington State formally recognized a hybrid water system in which the majority of water rights are based on the system of prior appropriation (Ecology, 2002).

Water supply deficiencies in the early 1940's led to disputes over water rights in the basin. The 1945 Consent Decree sought to clarify the issue of water distribution and rights in the basin (IOP, 2002). The decree specified two types of water users in the basin: those with nonproratable rights and those with proratable rights. Water users that have a priority date after 1905 have rights that are proratable. In years of drought, the senior nonproratable rights must be fulfilled from the available water supply first. The remaining water must be equitably split among users with proratable entitlements according to their water right (Watershed Assessment, 2003). The fraction of the entitlement each proratable user receives is the proration fraction. Approximately half of the total water entitlements are nonproratable. The Yakama Nation has the most senior water rights in the basin, with a priority dating back to time immemorial (Wash. Dept of Ecology v. Yakima Reservation Irrigation District, 1993). These rights are not subject to proration, the usual abandonment clause, and include the ability to fish in the accustomed places (Kittitas Reclamation District v. Sunnyside Irrigation District, 1993 as cited in the Primer to Washington Water Law). The Yakama Nation also holds substantial proratable entitlements.

Prior to large scale irrigation, the majority of streams in the basin were perennial. The natural water supply proved insufficient to meeting demand as more people moved to the basin. Streams began to dry up. Faced with increasing shortage and unreliable water supply, water users petitioned the federal government for an irrigation project in 1903 under the Reclamation Act. The USBR constructed the five dams for the Yakima Irrigation Project with the primary mandate of meeting irrigation demands, including new water rights, storage, and flood control. The Yakima Irrigation Project provides water for nearly 70% of the surface water rights in the basin (Watershed Assessment, 2003). The seven irrigation districts constructed a myriad of smaller diversion dams, canals, drains, and ditches to convey the

water stored under the project to their users. Few of the diversions were constructed with screens to prevent fish from straying into irrigation water ways or with ladders to ease passage (Watershed Assessment, 2003).

The Yakima Project severely altered the natural flow regime and the basin hydrology (SOAC, 1999). The following figure (Figure 1.3) presents an estimate of the unregulated flow regime at Parker on the Yakima River in comparison to what is currently measured. The figure is based on ten years of daily streamflow data from 1986 to 1995. It is apparent that the flow regime has been drastically attenuated by storage and human consumptive use. The peak flow periods of April through July and November through December are much smaller than the historic estimate. Summer low flows are also currently much lower than the unregulated values. Historically, diversions and inefficient application of water to crops, led to some tributary reaches and sections of the Yakima being completely denuded of water during the during parts of the year.

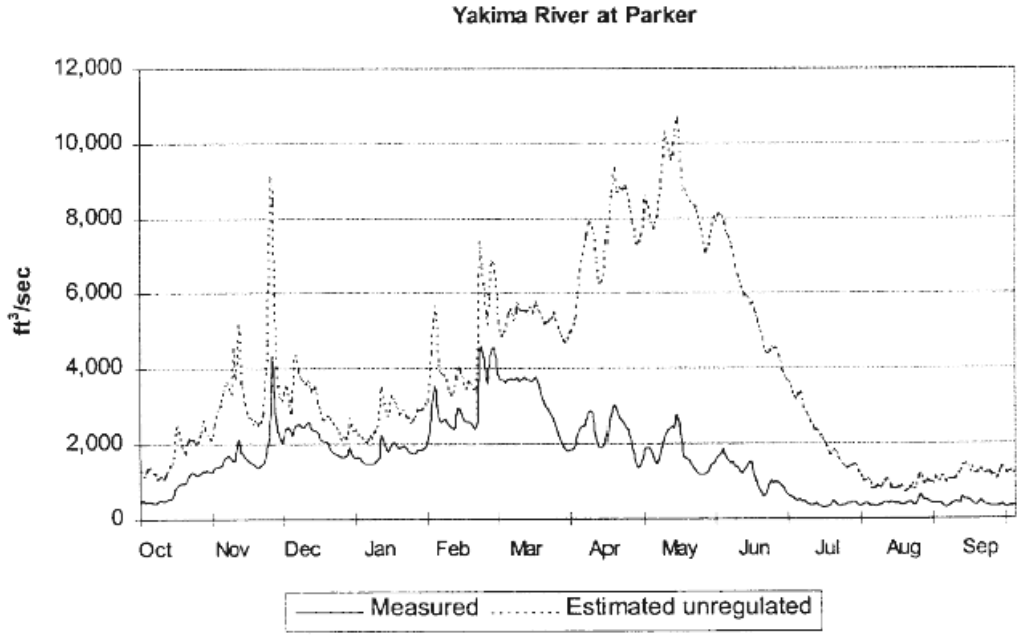


Figure 1.3: Average daily measured flows and estimated flows at Parker: 1986-1995. Measured flows are significantly attenuated.

Because of the irrigation practices and the early operational policy of the Yakima Project, salmon smolt survival within the basin decreased (Subbasin Plan, 2003). The outward migration of the juvenile salmon to the ocean and the successful return of the adults to spawn was, and still is, compromised by the maze of the irrigation networks and the highly altered flow regime. The precipitous decline of the various salmon populations culminated in the local extirpation of the coho, sockeye, and summer chinook salmon by the 1970's.

The loss of the fish prompted Congress to enact what is known as the Title XII legislation in 1994 (Watershed Assessment, 2003). Title XII provided for instream flows at two locations in the lower subarea, the Sunnyside (Parker) and Prosser Diversion Dams to be met from the Yakima Project and administered by the USBR. Minimum instream flows are not quantified per se, but are adjusted on an annual basin according to the water supply forecasts issued by the USBR.

1.3 Policy and Operations in the Yakima River Basin

Title XII presents the USBR with the unique challenge of ensuring a minimum instream flow amount that is subject to change with the forecasted total water supply available (TWSA), while fulfilling water rights, and providing flood control. Average total demand every year is approximately 2.5 million acre-feet of water. Central to the planning operations for determining instream flow and proration levels are the forecasts of the anticipated TWSA for the period of October 31st to July 1st. The TWSA is an estimate of unregulated flows at key basin points (such as the reservoirs), return flow from water users, and carryover storage. The TWSA is based on current reservoir storage, estimated runoff, and return flows. During years of normal flow, the TWSA can balance demand. In drought years, it cannot.

In all years, the Title XII target flows are set first based on the TWSA. Then proration levels are determined by subtracting the nonproratable entitlements and target flows

from the TWSA. The remaining forecasted water is then divided among proratable users as a fraction of their entitlements. The TWSA is updated each month of the irrigation season and the proration levels adjusted as necessary in light of the new forecast. In years where the TWSA is below 2.65 million acre-feet, the minimum flow is not allowed to drop more than 50 cfs at Prosser or be less than 65% of the target at Sunnyside in a twenty-four hour period. The proration levels will also be updated more frequently, usually biweekly, during a drought year (IOP, 2002). The following table lists the target flows at Sunnyside and Prosser depending on the estimated water supply.

Table 2-5
Target Flows at Sunnyside and Prosser Diversion Dams

Water Supply Estimate[1] for Period (million acre feet)				Target Flow (cfs) from date of estimate through October downstream of Sunnyside and Prosser Diversion Dams	
April through September	May through September	June through September	July through September	Without Basin Conservation Program	With Basin Conservation Program
3.2	2.9	2.4	1.9	600	900
2.9	2.65	2.2	1.7	500	800
2.65	2.4	2.0	1.5	400	700
<2.65	<2.4	<2.0	<1.5	300	300[2]

Source: Adapted from Title XII legislation as presented in USBR 1999

[1] "Estimate" refers to the Project Superintendent's water supply estimate.
[2] Only increased with reduced diversions below Sunnyside.

Figure 1.4: Title XII target flows in relation to the forecasts of the TWSA

The irrigation season usually commences around the beginning of April, depending on precipitation and temperature. This coincides with the average onset of the spring runoff. The USBR begins filling the reservoirs during runoff. Irrigation demands are met naturally from the spring runoff until flows are insufficient. At this point, the reservoirs switch modes from filling to storage control, which is when releases must be made to meet demand and the target flows. Releases are made through the end of October. A flip-flop in releases between

the upper three reservoirs and the lower two occurs toward the end of the irrigation season as specified earlier.

The SOAC, a consortium comprised of representatives from the Yakama Nation, US Fish and Wildlife, Washington Dept. of Fish and Wildlife, and representatives from the irrigation districts, was formed under the Title XII legislation to review the factors affecting salmon recovery in the Yakima basin and make recommendations as to the biologically based flow needs of salmon (SOAC, 1999). The legislation also provided funds for modernizing the irrigation systems and improving fish habitat under the Yakima River Basin Water Enhancement Project (YRBWEP). The various irrigation entities and the Yakama Nation bid for improvement funds based on needs identified by the SOAC. As more water efficient equipment is installed, conservation is expected to lead to an increase in the target flows as well as stabilize the water supply for irrigators. For every 27,000 acre-feet of water that is conserved, 50 cfs of water will be added to the target flows at Prosser and Sunnyside. If the YRBWEP is completely and successfully implemented, target flows will increase by 300 cfs (Watershed Assessment, 2003). The success of current efforts is measured in terms of smolt survival rates and access to spawning areas. Full implementation of the YRBWEP is not expected to be complete for some years.

1.3 Motivation

The Yakima Storage Project is able to store enough water to meet the demands of one operational water year. In the past, much of the consumptive demands were confined to agriculture. The Yakima River basin has become the nation's leading agricultural sector in the production of such high value crops as: apples, mint, asparagus, hops, and cherries. Many of these soft fruits bearing the sticker "Grown in Washington" are grown in the Yakima basin. Until roughly 1995, water management operations were conducted with the primary concerns of storage, meeting irrigation demand, flood control, and to a smaller extent, power

production. Consideration of environmental flow needs was not significant.

However, the decline of various salmon populations and growing national awareness of environmental issues forced policy makers to address the issue of minimum flows in the Yakima River basin. Title XII presents the USBR with the unique challenge of ensuring a minimum instream flow amount that is subject to change with the forecasted total water supply available (TWSA), while fulfilling water rights, and providing flood control. Central to the planning operations for determining instream flow and proration levels are the forecasts of the anticipated TWSA. The official forecast is issued to water rights holders on April 1st. Water managers are constantly seeking ways to improve the skill of the forecasts as well as the lead time of the forecasts. Agriculturalists would like a longer period to plan which crops to plant or to determine if leasing their water would be a better option. A longer planning period has the potential to save irrigation districts money and improve the ability of water managers to determine when storage of spring runoff should begin.

This thesis will attempt to address two questions concerning the water management of the basin.

1) Can a longer lead statistical streamflow forecast model be constructed of the coming spring runoff peak with enough skill to be useful to water operations planning?
2) What proration levels and target flows would be implied by forecasts generated for 1977 and 2001? These years marked two of the most severe droughts in the basin within the past 54 years

The formation of the ensemble forecasts generated for this research is discussed in the second and third chapters. The forecasts we develop use large-scale climate features that appear to influence the Yakima River Basin's hydroclimatology. Chapter 2 deals specifically with the methods used to determine relevant climate features and the formation of a predictor set from the climate information to be used in forecast generation. The formation of the

forecasts and their validity is discussed in the third chapter. The fourth chapter addresses our forecast ability in two drought year (1977 and 2001) scenarios and modeling the forecasts in a decision support system. Implications to water management are noted. The final chapter discusses the impacts of climate variability on salmon survival and streamflow in the Yakima Basin and concludes this thesis.

Chapter 2

Climate Diagnostics

Climate is a dynamical system controlled at all scales by such physical processes as thermodynamics and momentum conservation. Large-scale climate features are governed by the interactions between the ocean, atmosphere, and land. Due to the nonlinear nature of certain physical processes such as thermodynamics and internal variations, climate is not a strictly deterministic system. However, within certain characteristic timeframes, dominant modes of ocean-atmosphere, land-ocean, and land-atmosphere patterns emerge. The ocean-atmosphere patterns seem to exhibit the strongest influence because of the enormous heat capacity of the oceans and the resulting gradients between the ocean and the atmosphere.

The hydroclimatology of many regions around the world has been shown to be influenced strongly by large-scale climate features. The hydrology of the western United States exhibits a strong relationship with climate phenomena in the Pacific Ocean. The three dominant ocean-atmosphere patterns seen in the Pacific are the El Niño-Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), and the Pacific North American (PNA) pattern. These three teleconnection patterns are responsible for determining the winter atmospheric circulation that brings the Pacific Northwest (PNW) much of its precipitation. Statistical analysis is often employed to relate the primary modes of such features as sea surface temperature (SST) and sea level pressure (SLP) to hydrological variables such as precipitation or temperature. The strength of the relationships between climate phenomena and hydrological variables increases the predictive capabilities of streamflow or precipitation forecasts. Increased skill in forecasting is invaluable to water resources managers in the western U.S. who face increasing demands on limited water supplies.

2.1 Dominant Climate Features Affecting the Hydroclimatology of the Western United States

The variability in volume and timing of streamflow in many western basins is largely determined by precipitation. Yet, precipitation in the semi-arid western United States varies greatly spatially, inter-seasonally and inter-annually. Snowmelt provides an estimated 75% of the annual discharge of many of the streams in this region (Palmer, 1988 as cited in Cayan, 1996). Climate phenomena and topography play strong roles in the distribution of snow. Much research has been conducted in an attempt to understand the links large-scale climate features and the hydrology of the western U.S. Three teleconnection patterns are the predominant modes of climate phenomena affecting the hydroclimatology of the Pacific Northwest and are described below.

Teleconnection patterns are recurring, large-scale pressure and circulation pattern anomalies that can persist for several months to a few years. One of the primary modes of low-frequency variability in the Northern Hemisphere mid-latitudes is the PNA teleconnection pattern. The PNA is characterized by a quadripole pattern of anomalously high and low pressures in the height field. A blocking pattern of above average geopotential heights is observed with centers near Hawaii and northwest central Canada. Corresponding low pressure systems set up over the Aleutian Islands and the southeastern United States (Wallace and Gutzler, 1980). The pattern is an annual feature, occurring in almost every month except for June and July. It is the teleconnection pattern that dominates Pacific Northwest wintertime weather.

The location of the pressure systems and the ensuing circulation patterns dictate the movements of the jet stream. Circulation around a low pressure system is counterclockwise, and opposite around a high pressure system. The pressure gradient between high and low pressure systems induces winds. The resulting fast moving bands of winds in the troposphere are known as jet streams. Subtropical jet streams in the north Pacific follow the circulation patterns set by the location of the high and low pressure systems associated with the PNA.

The mid-latitude jet stream is the primary mover of storms to the Pacific Northwest and is strongest during the winter season when temperature gradients are the largest. During the winter, the jet transports moist air from the Pacific Ocean over the North American landmass. Refer to Figure 2.1 for the location of the PNA centers and the subtropical jet stream.

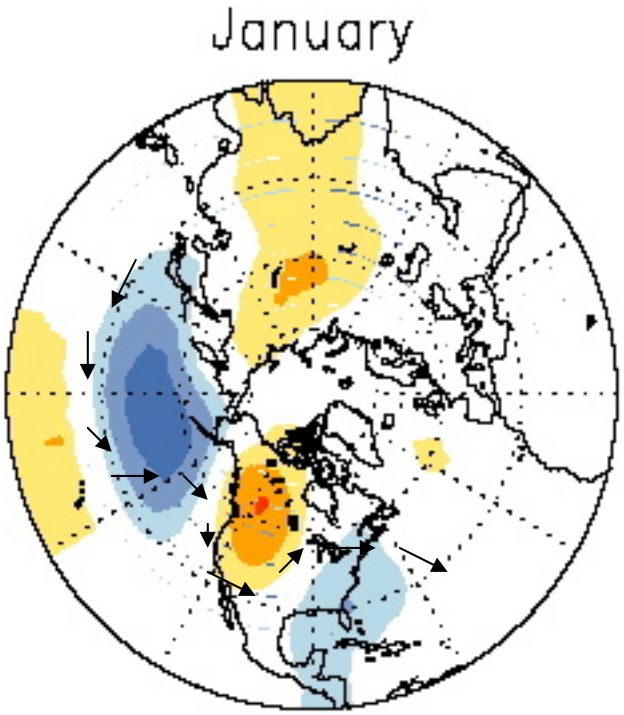


Figure 2.1: Positive PNA Pattern and the mid-latitude jet stream

The PNA is variable in strength and undergoes positive or negative phases which are modulated by ENSO or PDO events. The positive phase of the PNA is associated with an intensification of the Aleutian low and a strengthening of the high pressure center over the Canadian Rocky Mountains. When the relative pressure difference between the two pressure centers results in an overall high pressure, the winter storms deflect north toward Alaska which reduces precipitation in the Pacific Northwest. The negative phase causes a reverse of the positive situation and induces precipitation to increase as storm tracks are directed into

the northwestern U.S. Thus, the strength and location of the Aleutian low and the Canadian ridge are critical to the movement of storm tracks into the Pacific Northwest.

The primary variability modes of long persisting ocean-atmosphere circulation phenomena in the Pacific are ENSO and PDO. There is some debate as to whether ENSO and PDO are truly separate features, as they exhibit similar spatial manifestations in the North Pacific (Gershunov and Barnett, 1998). However, this debate is beyond the scope of this thesis. For our purposes, the two will be considered independent. ENSO is an interannular feature with a cyclical recurrence every two to seven years. PDO operates on an interdecadal time-scale of approximately twenty-five years. Both modulate the pressure systems associated with the PNA pattern.

The PDO phenomenon is characterized by a pool of abnormally cool water located in the northern central Pacific that persists for a couple of decades before reversing polarity. Zhang et al. (1997) identify the PDO as the leading principal component of Pacific SST. A narrow band of unusually warm water located of the west coast of North America often accompanies the cold pool (Zhang *et al.*, 1997, Latif and Barnett, 1996, Gershunov and Barnett, 1998). Sea level pressures (SLP) are affected by the various phases of the PDO. The warm phase of the PDO is characterized by the SST's noted previously and an anomalously intense Aleutian low and Canadian high. The cold phase of the PDO exhibits the opposite behavior. The modes of the PDO have also been noted as having great influence over the North Pacific currents and food web production (Polovina *et al.*, 1995, Aydin, 1999). Figure 2.3 shows SST and SLP regressed against the PDO index. The PDO index is the leading principal component of monthly SST in the North Pacific. The values are all standardized (Mantua, 2004).

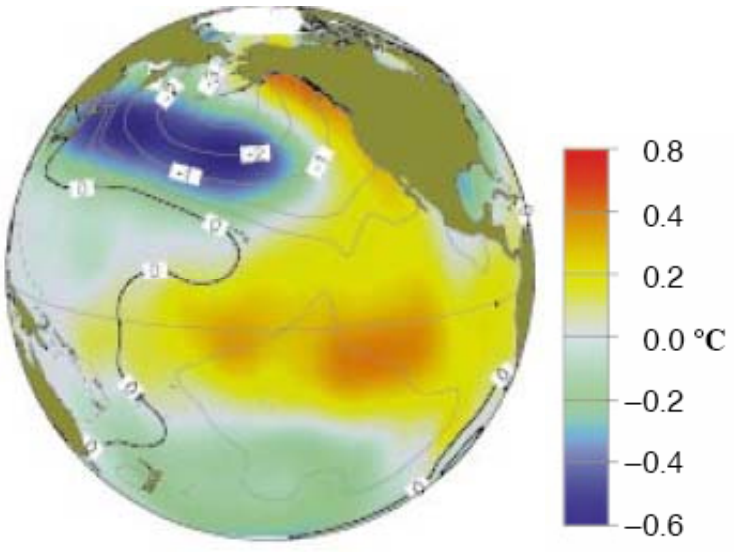


Figure 2.2: SST and SLP regressed against the PDO index

Given the basic physical mechanism under which convection occurs, it would seem that precipitation, and thus streamflow, would be enhanced in the Pacific Northwest (PNW) during the positive phase of the PDO. The warmer than normal waters off the coast tends to encourage warm, moist air over the region. However, as noted previously, the strengths of the Aleutian low and Canadian ridge are key to the amount of precipitation received. A relative pressure ridge enhances cyclonic winds over the region, blowing the moist air away from the coast. Therefore, precipitation and streamflow in the PNW tend to decrease during the PDO's warm regime. The opposite has been observed during the cold phase of the PDO. However, abnormally warm SSTs off the PNW coast can mitigate the effects of a relative ridge and produce normal winter precipitation. Likewise, abnormally cool waters can lead to decreased convection and a somewhat drier year.

Much focus in recent climate research has been on the impact of the ENSO phenomena on global hydroclimatology. The impact of the various phases of ENSO on the precipitation, temperature, and streamflow of the Pacific Northwest has been well demonstrated by several researchers (Hamlet and Lettenmaier, 1999; McCabe and Dettinger, 1998; Cayan, 1996. McCabe and Dettinger, 2001 and Clark *et al*. 2001, have noted that the

interannual variability of snowpack in the western U.S. is strongly influenced by the El Niño-Southern Oscillation.

ENSO is defined by changes in SST and SLP around the equatorial Pacific. The SST anomalies are recorded in a narrow band extending roughly 5º N to 5º S about the equator. Variations in SLP are measured as a phase shift between pressures measured at Darwin, Australia, and the island of Tahiti. When a pressure ridge is located over Darwin, a trough is located over Tahiti, and vice versa. The relative difference between the two pressure centers is referred to as the Southern Oscillation (McCabe and Dettinger, 1999). ENSO is marked as anomalous extremes in the boundaries mentioned, yet the influence of such tropical forcing extends well into the mid-latitudes.

The mean state of the tropical Pacific sea temperatures is denoted by a tropical west Pacific 6º to 8º C warmer than the eastern tropical Pacific. During an El Niño year, the relative temperature difference between the two decreases, whereas in a La Niña year, the difference is enhanced. Both El Niño and La Niña alter the tropical convection and rainfall patterns about the equator. These changes affect the average latent heat fluxes in the tropical atmosphere. The altered information propagates in a wave-like fashion through the troposphere up to the mid-latitudes, particularly during the winter. The primary effects of ENSO with concern to the PNW are manifested as an alteration in the direction of the trade winds and an intensifying or weakening Aleutian low and the Canadian Rocky Mountain ridge. The Aleutian low also tends to shift southward during the positive ENSO phase (and positive PDO phase). The shifting of the trade winds and the modifications in the Aleutian low transform the jet stream over North America. As demonstrated below, the jet stream behaves quite differently during ENSO events and can impact the winter precipitation in the PNW. Winters tend to be drier and warmer in the PNW during an El Niño year.

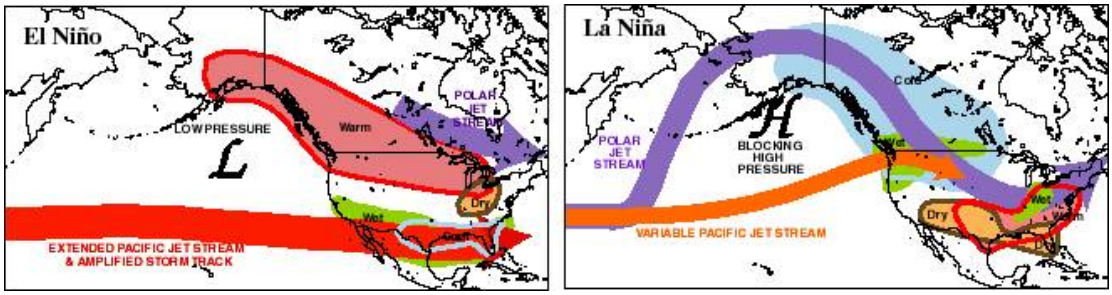


Figure 2.3: Jet stream patterns during El Niño and La Niña

Hamlet and Lettenmaier (1999) note that PDO and ENSO appear to act independently in affecting PNW hydroclimatology. When the two regimes are in phase, streamflow in this region tends to bias toward above or below normal. If the two events coincide out of phase, for example a La Niña year during the warm phase of the PDO, little consistent deviation from the long term mean is observed in the streamflow, and it becomes difficult to determine how the spring runoff flows might respond. Another difficulty in predicting the effects a particular El Niño or La Niña event might have on PNW hydrology is the fact that ENSO is a non-linear pattern (Hoerling *et al.*, 1997, Hamlet and Lettenmaier, 1999). No two ENSO events are alike in their geographical area of influence or strength. Each event generates a unique atmospheric response that alters the response in the PNW. Therefore, basing a forecast solely on an El Niño or PDO index is not likely to have great skill.

2.2 Analysis of Coupled Ocean-Atmospheric Features Relevant to the Yakima River Basin

The previous section illustrated the influence of coupled ocean-atmospheric circulation patterns on the hydroclimatology of the western U.S. and the PNW. This section examines the impact of such features on the Yakima River Basin and discusses the climate diagnostics performed in order to determine the relevant climate indices. Several data sets were employed in the generation of the streamflow forecasts.

1) Daily reconstructed naturalized streamflows from the USBR for the five reservoirs: Kachess, Keechelus, Cle Elum, Rimrock, and Bumping. The data sets encompassed the years 1949-2002 and were obtained from the Hydromet data base: http://www.usbr.gov/pn/hydromet/yakima/yakwebarcread.html

2) Daily natural streamflow data for the American River, a small tributary river to the Yakima, that met Slack and Landwehr's criteria for natural streamflow. This data set covered the years 1949-2002. The data was retrieved from the HCDN: http://water.usgs.gov/pubs/of/ofr92-129/content.html

3) Monthly SWE data was gathered for eight snow course and snotel stations in the Yakima Basin that had data extending from 1949-2002 from the NRCS National Weather and Climate Center (http://www.wcc.nrcs.usda.gov) and from Martyn Clark (CIRES).

4) Monthly values of large-scale ocean and atmospheric phenomena: SST, vector winds, and 500 mb geopotential heights for the Northern Hemisphere from the CDC website. These were derived from the NCEP/NCAR re-analysis data sets.

5) PDO index (Mantua, 2004) and Multivariate ENSO Index (Klaus, 2004)

No natural streamflow records exist for any water systems in the Yakima basin except for the American River. As noted in Chapter 1, water extractions for irrigation began as early as the 1880's in the basin, so it is difficult to gauge what the actual streamflow volume of the basin might have been without anthropogenic intervention. As the basin is heavily regulated and irrigated, the USBR can only reliably estimate the natural flows at the five reservoirs. Since the five reservoirs are located in the mountainous headwater regions, diversions upstream of the reservoir are minimal. Naturalized daily streamflow (Qu) is determined by gage measured reservoir releases (Qd), and change in storage of forebay elevation (AF) from the previous day. 1.98 is how many acre-feet 1 cfs over the course of a day would fill.

$$Qu = Qd + (AF2 - AF1)/ 1.98 \qquad \text{Eq. 2.1}$$

Considerable error potentially exists in these reconstructed values, as several factors such as evaporation and wind were not accounted for. Strong winds at the reservoirs can persist for several days, causing piling on one lake side and erroneous gage measurements. Another source of error was introduced into the older records when the paper records were digitized in the late 1970's. Forebay elevations were often rounded to the nearest ten acre-feet, resulting in a potential overestimation of streamflows from the early period (Christopher Lynch, personal communication). Little can be done to compensate for these errors.

The ability to observe the relationship between climate and streamflow is often hampered by the fact that many water systems are highly regulated. Such anthropogenic influence can alter the hydrology of a system such that the direct response to a climate input can be masked. It becomes inherently difficult to provide forecasts based on climate phenomena the farther the system is from its natural state. Slack and Landwehr (2002) developed the Hydro-Climatic Data Network (HCDN) for the USGS in order to identify river systems with long historical records (at least 20 years) that are free of anthropogenic modifications. The American River in the Yakima basin met Slack and Landwehr's criteria from the 1920's to 1988. The HCDN does not vouch for any systems after 1988. So, the American River's monthly streamflow record is known to closely reflect natural flows through at least, 1988. The quality of the records from 1988 to 2002 is not known. While the American River is not used as a forecast location in the USBR's TWSA forecast, we include it in our forecast as a comparison of a natural system to the reconstituted natural states of the reservoirs.

Snow water equivalent (SWE) data was obtained from six snow course and snotel stations in the Yakima basin that had record lengths covering the 1949-2002 period. Snow course and snotel data for the state of Washington were provided by Martyn Clark (CIRES) and cross-checked with the NRCS Washington data. All of the stations are located along the mountainous western rim of the basin where the majority of the precipitation contributing to

spring runoff occurs. Mt. Rainier has a strong orographic effect on the winter stormtracks entering the basin, causing a split in snow distribution between the upper and lower basin. Therefore, stations were divided according to geographic location and proximity to the forecast sites. Four stations fell into the upper basin and four into the lower basin.

To ensure that no one station is given more weight than the others, the SWE for each month of each year is divided by the long-term mean of that month for each station. For example, at one site the SWE on Feb 1. 1987 is 43 inches and the long-term Feb 1st mean at that station is 60 inches. The SWE value measured on Feb. 1, 1987 is 72% of normal. This is repeated at all sites for Feb. 1st, March 1st, and April 1st SWE. The percent values from the groups representing snow pack in the upper subbasin and the lower subbasin are averaged separately to figure out the overall SWE distribution in those two regions.. Stations with data missing for a particular year were not included in the basin average. Separate timeseries of the percent average SWE for the upper and lower subbasins were then created. These vectors of regional percent averages for January, February, and March were employed in the forecasts. As SWE data for a site is not available until the end of the month, the percent averages are actually of SWE measured on Feb. 1, March 1, and April 1.

The large scale climate variables were obtained from the NCAR/NCEP re-analysis data (Kalnay *et al*., 1996) for the 1949-2002 time frame. The phenomena such as 500 mb geopotential height and SSTs, are available for download as a seasonally and spatially averaged variable for the region of interest. The re-analysis data is available at 2.25° grid spacing, allowing for a fairly tight spatial average.

The PDO index was obtained from http://jisao.washington.edu/pdo/PDO.latest and was compiled by Mantua (2004). It is derived as the leading PC of SST in the North Pacific Ocean. The MEI is a weighted average of six variables such as SST and pressure and defines main ENSO features in the tropical Pacific (Klaus, 2004). It is available from http://www.cdc.noaa.gov/people/klaus.wolter/MEI/table.html.

2.2.2 Methodology

Given the physical mechanisms (teleconnections) mentioned earlier, it is expected that large-scale climate phenomena impact streamflow in the Yakima Basin and might lend some predictability to a forecast of such streamflows. Correlation and composite analysis were performed to verify the relationship between spring runoff and winter ocean-atmospheric patterns.

The basin hydrology is dominated by a wet season (roughly from Nov. to March) and is dry the rest of the year. The streamflow reflects the basin's extreme sensitivity to temperature. Two streamflow peaks are observed: one extending from mid-October to mid-December while it is warm enough for the precipitation to fall as rain or rain-on-snow, and the other corresponds to the April through mid-July runoff season. Roughly 60% of the annual streamflow occurs during the spring runoff period. Figure 2.5 exhibits the average monthly streamflows for the five reservoirs. During the dry season, streamflows are maintained primarily through groundwater infiltration, some glacier melt, and the remains of the snowpack. From approximately mid-December through the beginning of April, cold temperatures ensure that the majority of the precipitation falls as snow. However, extreme flood events have occurred in the January through March period when brief, rapid thaws cause rain-on-snow events and meltoff. Years with brief mid-winter thaws have the potential to be drought years, as the snow pack is eroded and floodwater must sometimes be passed through the reservoirs rather than be stored.

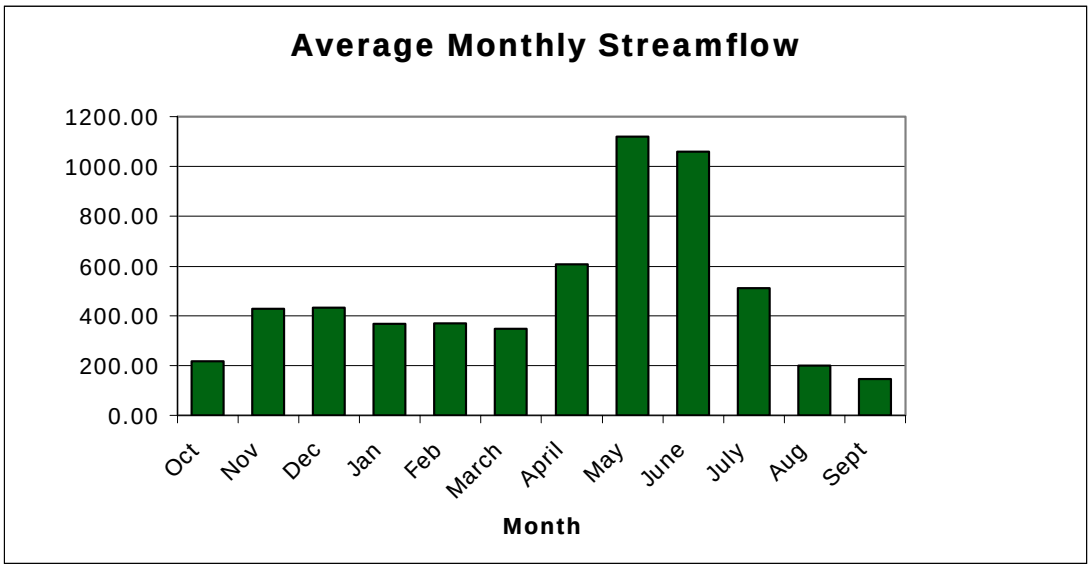


Figure 2.4: Average monthly streamflow for the five reservoirs (1948-2002)

The Yakima Basin is not a large basin from a climate variability perspective and each of the six forecast sites were expected to show similar relationships with large-scale climate influences. Therefore, in order to perform the correlations, and ultimately the forecasts, principal component analysis (PCA) was performed utilizing all six forecast sites to generate a set of variables containing the maximum streamflow variance. The idea behind PCA is that a multidimensional data set can be transformed into a linear combination of the underlying patterns. By performing PCA, it is hoped that the data set can be reduced to a smaller number of variables that contain the relevant information of the original data set. PCA is performed on the anomalous behavior of the data set, which is found by subtracting the mean. Our data set is comprised of six time series vectors corresponding to the streamflow at each site. The anomalies were calculated by subtracting the site specific mean from each individual site:

$$\vec{\mathbf{x}}'_{\mathbf{site1}} = \vec{\mathbf{x}}_{\mathbf{site1}} - \mu_{\text{site1}} \qquad \text{Eq. 2.1}$$

The anomaly vectors (one for each site) form the columns of our data set matrix. The anomaly matrix can then be expressed as a linear combination of variables that capture the maximum covariance between the sites. This is done with the assumption that the spring runoff at one of the sites will vary in a similar manner to the variation in spring runoffs at all the other sites, at the same time. The linear combination of the anomaly matrix is expressed as (Storch and Zweirs, 2001):

$$\vec{\mathbf{x}}_t = \sum_{i=1}^{k} \alpha_{i,t} \vec{\mathbf{e}}_i \qquad \text{Eq. 2.2}$$

The eigenvalues or principal components (α) account for the maximum joint variability in time in the streamflow between sites. The eigenvectors ($\mathbf{e}^i$) provide the spatial components and point in the direction of maximum joint variability. Each eigenvector is orthogonal which implies that the principal components are uncorrelated and independent. The number of eigenvector/value pairs is equal to the number of sites. The first eigenvector corresponds with the largest (and first) principal component (PC), which describes the most data variability. The variance explained by the remaining PCs decreases with each PC. The advantage to using PCA is that it preserves the joint patterns between the forecast sites in the forecasts, which would be lost if the forecasts were generated separately at each site. The disadvantage is that some information is lost in the process. The first PC was found to explain 95% of the variance. This PC was used in generating the correlation and composite maps of the relationship between climate variables and the streamflow.

Before the PCA was performed, spring streamflow volumes (April to July total) were calculated for the five reservoirs and the American River from the daily record. The resulting volumes were then normalized as the range of volumes is large from one site to the next. The following graph shows the principal component (PC) distribution. The first principal component captured 95% of the variance of the data and the second explained 4%. Correlation analysis was performed for both PC's, however, no significant climate

relationships emerged for the second PC and it was not kept in the ultimate analysis. Dry years generate a positive PC loading, while wet years are denoted by a negative PC loading.

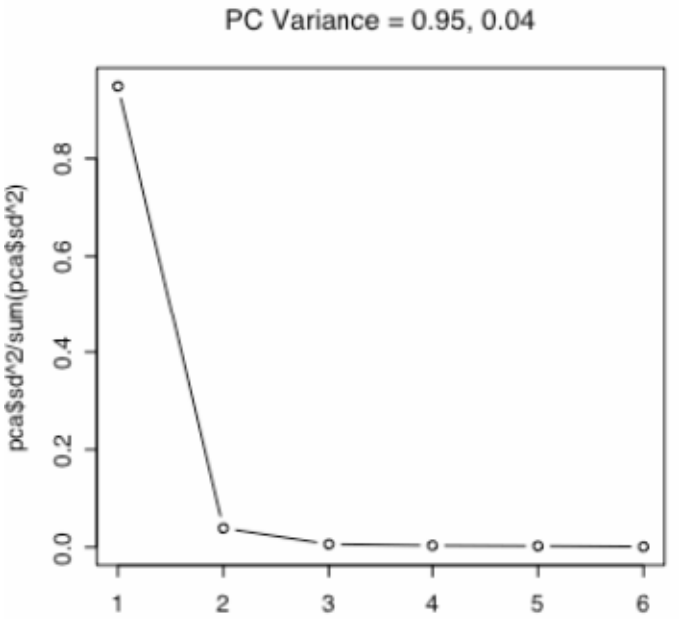


Figure 2.5: PCs of USBR AMJJ streamflow volumes. PC1 explains 95% of total variance, PC2 explains ~4%.

Correlation analysis was performed establishing the relationship between various climate indices and SWE against the first PC. As it was not known in advance what the relationships would be between the climate factors and the PCs, a non-directional assumption was made. In order for a correlation to be deemed statistically significant to a 95% level, it must be at least 0.377. A correlation coefficient of 0.436 indicates a correlation significant to the 99th percentile. Correlations were determined to be significant following the two-tail t-test methodology outlined by Lowry (2002).

Composite analysis was also performed to examine anomalous climate behavior during years of extreme streamflow conditions. In composite analysis, the average (composite) of the mean or anomalous behavior for a data set is found. The composites were performed for years in which the streamflow was below the 15th quantile and above the 85th

quantile. Relevant predictors were identified from the correlation and composite analysis. Finally, the forecasts attempted to predict the PC value for a particular year.

2.3 Climate Analysis

The following sections describe the techniques employed in determining the climate predictors that impact spring runoff in the Yakima basin. The relationships between the teleconnection patterns described in the previous section are explored through composite and correlation analysis, which are described below.

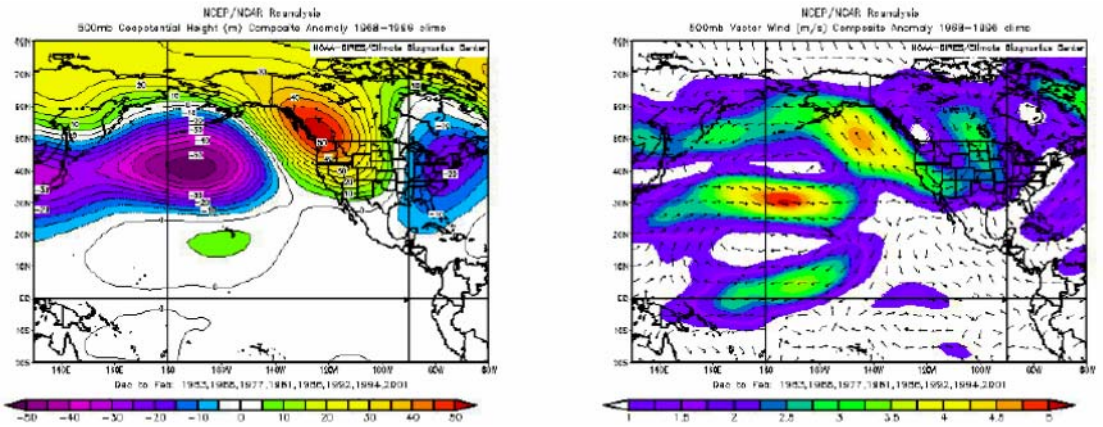
2.3.1 Composite Analysis

Composite analysis was performed first to provide an insight into the physical relationships that might exist between various climate indicators and spring streamflow. As noted previously, composite analysis was performed for years of extreme low flows or high flows to examine climate relationships in anomalous years. Eight years fell into each of the two quantiles. The composites examined the anomalous behavior of such climate phenomena as 500 mb geopotential height, vector winds, and SST that were hazarded to play a role in the Yakima's hydrology. The resulting composite maps indicate physical features common to drought years and to wet years.

The first set of figures (Figure 2.6) displays the atmospheric features common to low flow years. The second set (Figure 2.7) represents atmospheric features dominant during high flow years. The PNA pattern is striking in both sets of images. As noted in the description of relevant physical features at the beginning of this chapter, the Canadian Rocky Mountain ridge and the Aleutian low are the key pressure systems determining how stormtracks move into the PNW. Figures 2.6.a and 2.7.a display the 500mb geopotential height patterns associated with the PNA. Earlier, it was noted that an intense Canadian ridge and Aleutian trough would cause a strong clockwise rotation of the winds near the PNW coast, directing stormtracks and moist air from the Pacific away from the land interior. This

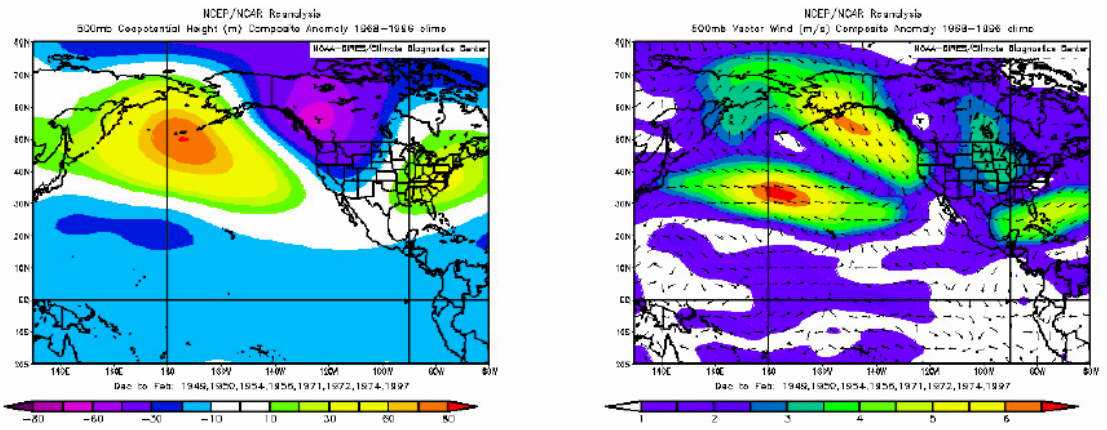
behavior is clearly evidenced in the 500 mb vector wind plot (Figure 2.6.b), which can be thought of as a visible proxy for the geopotential height patterns, and is associated with drought. The opposite behavior is observed during high flow years. A weaker Aleutian low and Canadian ridge set up a counter-clockwise rotation pattern, seen in the 500 mb vector wind plot (Figure 2.7.b), which moves storms into the PNW region.

Remarkably absent was any information about the SST patterns during the extreme flow years. Composite analysis of SSTs for both scenarios showed no anomalous deviations in the SSTs of the north central or PNW coastal Pacific waters. One would expect cooler than normal SSTs to the east of a low pressure region and warmer SSTs on the west side of a ridge. The majority of the information, however, seems to be supplied by the atmospheric phenomena. The strong atmosphere signal might negate any contribution warmer or cooler SSTs have to convection. Thus, SST's do not seem to have a significant role in anomalous years. However, we still examine SST as a possible predictor for all years in the correlation analysis.



(a) **(b)**

Figure 2.6: Low streamflow year composites: 500 mb geopotential heights (a) and corresponding vector winds (b)



(a) **(b)**

Figure 2.7: High streamflow year composites: 500 mb geopotential heights (a) and corresponding vector winds (b)

2.3.2 Correlation Analysis and Predictor Selection

The correlation analysis involves correlating the first PC with general ocean-atmospheric circulation patterns and examining which relationships are statistically significant. One forecasting goal was to see if the information provided by various predictors persisted and could be used to increase the lead-time of the forecasts, Therefore, correlation analysis was performed in moving three month seasonal intervals, starting with a lead time of eight months before the April-July spring runoff period. Correlation maps were generated using the Climate Diagnostic Center's (CDC) website. Seasonal correlations beginning in August of the previous year were performed and examined the physical link between the 500mb geopotential height field, SST's and the first PC. The region of interest was confined to the Pacific Ocean in the northern hemisphere and to the northern hemisphere land masses. This region was picked because the physical link between climate features and streamflow might be deemed more plausible to water managers than looking at relationships in the ENSO regions, which would certainly influence mid-latitude patterns.

Two criteria were used in determining which physical phenomena were real and significant as possible predictors for the forecasts: 1) the feature had to be statistically significant at least to the 95th percentile and, 2) the feature should persist for more than three

months. The second requirement comes from the observation that a short-term physical phenomenon is unlikely, though not improbably, to be responsible for sustaining the winter storms that provide the majority of the precipitation to the basin. There is also the slight chance that a short lived feature might be an artifact of the NCEP/NCAR re-analysis data and not a real physical phenomenon.

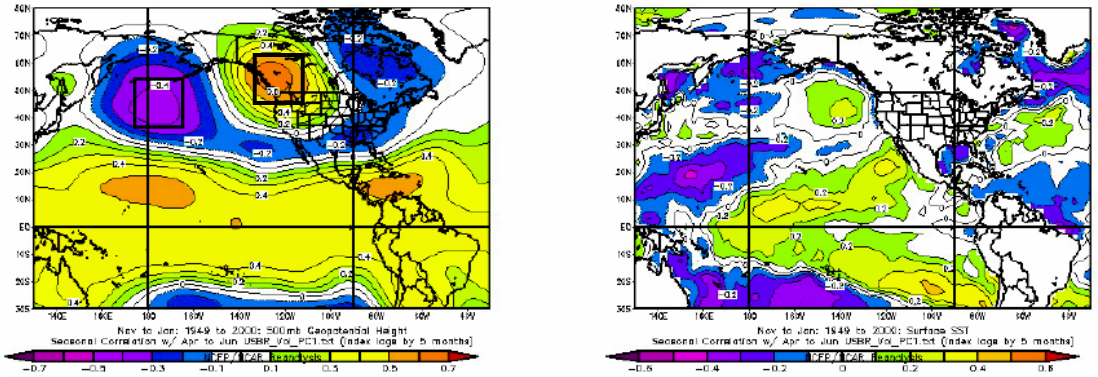
The following table (Table 2.1) presents the strength of select climate phenomena (on the left) for the indicated three month season (on top) with the first PC of the streamflow.

	Aug- Oct	**Sep- Nov**	**Oct- Dec**	**Nov- Jan**	**Dec - Feb**	**Jan - Mar**
Canadian Ridge	0.2	0.4	0.5	0.66	0.65	0.65
Aleutian Low	0	-0.2	-0.3	-0.46	-0.51	-0.52
SST	0.3	0.3	0.4	0.2	0.58	0.64

Table 2.1: Correlation of climate indices of select seasons with AMJJ streamflow. Values above +/- 0.4 are significant beyond the 95th%.

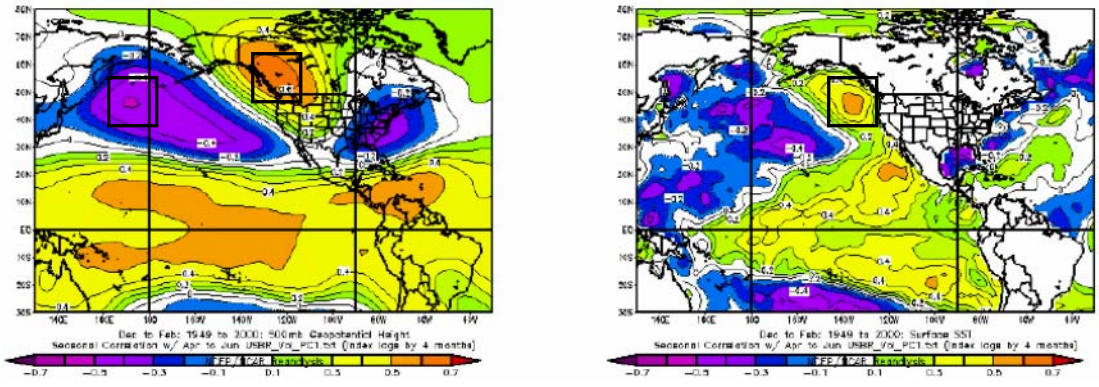
The PNA pattern begins emerging in the August to October timeframe, but does not develop its full strength until the October to December season. It is the dominant teleconnection pattern seen in the correlation analysis. The SST correlation becomes insignificant during the November through January (NDJ) time-step. That it, and the geopotential height indices are barely statistically significant during the October to December (OND) period, suggests that the climate indices during this time are only indirectly correlated with spring runoff. What is more likely is that the climate indices from OND would correlate strongly with the midwinter streamflow peak, which is a direct response to the rain at the beginning of the wet season. The OND relationship might potentially be viewed as a proxy for soil moisture conditions in the spring. If the soil is saturated from the early wet season precipitation, less of the spring melt will have to penetrate the soil before runoff occurs. This would lead to higher spring flows.

Of interest to water managers is the ability to generate reasonably skilled forecasts at longer lead times. Therefore, the seasons showing the strongest climate to streamflow correlations are examined in greater detail. Based on the composite analysis presented previously, we are interested in examining the strength of the relationships of the Canadian ridge, Aleutian low, and coastal SSTs with the first PC and the evolution of the relationship through the seasons. The (NDJ) and December to February (DJF) periods show the greatest predictor potential for forecasting spring streamflows. The first set of figures (Figure 2.8) shows the contour maps of the correlations for the 500mb height fields and SSTs occurring during the NDJ season with the first PCs. The second set (Figure 2.9) displays DJF correlations.



(a) **(b)**

Figure 2.8: NDJ correlations with PC1: 500mb height field (a) and SSTs (b)



(c) **(d)**

Figure 2.9: DJF correlations with PC1: 500mb height field (c) and SSTs (d)

The ocean-atmospheric correlations are quite strong in both the early winter (NDJ) and the winter season (DJF). Three of the pressure systems associated with the PNA pattern are strongly evident in the correlation maps. The areas of strongest correlation, outlined by the boxes, were selected as predictors for the formation of the streamflow forecasts. No statistically salient SST features were found in the NDJ correlations. The resulting SST patterns are the expected response to winds generated by the pressure systems.

Climate predictors for the forecast model were formulated by spatial averaging the climate indices over the areas of highest correlation, which are highlighted by the boxes in the correlation maps. This yields a time series vector for each of the indices that can be used in the forecast model.

SWE is even more directly related to spring runoff than the climate indices. The correlation between the various cumulative SWE, as measured at the beginning of the months of February, March, and April, and spring streamflow is quite strong and is presented in the following table. The correlations are negative because of the PC loadings. Recall that high flow years corresponded with negative PC loadings and the low flow years had positive loadings. Thus the relationship between SWE and the first PC is inverse. Below average SWE is related to low spring streamflows (high positive PCs) and vice versa.

	Feb 1 SWE	March 1 SWE	April 1 SWE
Correlation	-0.85	-0.88	-0.90

Table 2.2: SWE correlations with first PC of spring runoff. These values are significant beyond the 99th%.

Forecasts relying on climate predictors only could be generated using the OND time series. We do not feel that the climate correlations are strong enough at earlier time frames to formulate forecasts. While these forecasts might not provide a detailed resemblance to actual spring runoff, the values generated would give an indication as to whether or not a drought conditions were forming. Furthermore, water managers would be able to watch the

development of the PNA pattern, particularly the Aleutian low and the Canadian ridge, in conjunction with the ENSO index or PDO. Monitoring the location of the Aleutian low to determine if it was shifting southward might also provide an early indication of the likely position of the mid-latitude jet stream and the direction of winter storm tracks over the North American land mass.

Finally, to check that PDO and ENSO would not form suitable predictors, PC1 was correlated with the PDO Index and the Multivariate ENSO Index (MEI). The correlation was found to be stronger between the PDO and PC1 than PC1 with ENSO. The correlation between PDO and PC1 was 0.34, which is significant to approximately the 90^{th}%. However, the correlation between ENSO and PC1 was quite weak, only 0.11. While the correlation with the PDO is fairly strong, it is not significant to the 95^{th}%. An alternative forecast might be generated using the PDO index, which is a measure of SST, but we do not use it our forecast generation. This thesis will not speculate as to why one relationship is stronger than the other.

In the next chapter, we will examine forecasts generated using climate predictors beginning with the OND time frame. We will also compare the skill of forecasts utilizing climate information in conjunction with SWE against those forecasts generated using SWE alone.

Chapter 3

Nonparametric Stochastic Streamflow Ensemble Forecasting

Streamflow varies greatly from location to location and in the timing of the distribution from year to year. For example, the sites selected for forecasting in the Yakima basin range in altitude from 2,000 to a little over 3,000 feet. These elevations within the Cascade Mountains are highly susceptible to temperature fluctuations where a few degrees can mean the difference between rain or snow, early thaws and floods. The elevations of the southern reservoirs, Bumping and Rimrock, are approximately 1,000 feet higher than their northern counterparts. Onset of spring runoff is on average, two weeks later than the northern sites. Secondly, the catchments area and volumetric capacity of each reservoir is different. Because of the orographic impact of Mount Rainier, situated just to the west of the Yakima basin, winter stormtracks diverge around the basin. This leads to a large spatial unevenness in the distribution of precipitaion. Thus, the range of streamflows among sites varies greatly.

Given these substantial orographic impacts, any streamflow forecasting method must retain the particular spatial and temporal information of a system in order to provide a reliable estimate of flow conditions. The forecast must also be able to generate a range of streamflow scenarios that have not been historically observed but are statistically possible. Sequences exceeding the observed minimum and maximum are also desirable in order to allow water managers to develop operations for extreme events. Finally, the ability to disaggregate the seasonal forecast to smaller time steps is essential for use in water operations planning scenarios. The following chapter discusses the techniques employed in the formation of streamflow ensemble forecasts, the resulting forecasts, and evaluates their skill.

3.1 Introduction

Streamflows are traditionally forecast using either deterministic methods (physically-based), stochastic methods (empirically-based), or a combination of the two. Each method has unique advantages and disadvantages. Regression forecasts are relatively easy to develop and can provide the level of accuracy needed for employment in water decision support systems. Deterministic forecasts model the hydrological response of the basin (e.g. streamflow) to various sets of inputs (e.g., soil moisture, long wave radiation). Deterministic models can provide realistic response scenarios, including responses not seen in the historical record. These responses are generated by perturbing the input data.

Deterministic models rely on large sets of parameters to model basin hydrological response. Each parameter must be calibrated for the model and assumptions made about the calibration. For example, antecedent soil moisture may be an input to a deterministic model. Multiple samples must be taken from sites within the basin and calibrated with the assumption that the samples are representative of conditions within the basin. As soil type is generally not homogeneous across the basin, such averaging might not accurately reflect conditions. Secondly, deterministic models are computationally expensive (Martyn Clark, personal communication).

Stochastic techniques attempt to find a relationship between the predictand (e.g., streamflow) and the predictors (e.g., precipitation or temperature). Traditional parametric forecasts fit a linear regression between the predictand [Y_t] and a set of predictors [x]. These models take the form:

$$Y_t = a_1x_{1t} + a_2x_{2t} + .. + a_nx_{nt} + e_t \qquad \text{Eq. 3.1}$$

The coefficiencts (a_n) are derived from the data, often by minimizing the sum of the squares of the errors. The error, e_t, is assumed to have a Gaussian distribution with a mean of 0 and variance of σ_e^2. Helsel and Hirsch (1995) describe the theory of parametric models, equations for parameter estimation, and methods for testing the model's goodness of fit.

There are several drawbacks to parametric models: 1) the predictor data and errors are assumed to be normally distributed, 2) linearity is assumed between variables, and 3) non-portability across the data sets (i.e., sites).

Nonparametric techniques estimate a local fit between the predictors and predictands. No assumptions about the functional form of the relationship (e.g., normal distribution or linear relationships) are made, which alleviates the drawbacks of the parametric models. There are several nonparametric techniques widely employed, such as kernel-based estimators, splines, K-Nearest Neighbor (K-NN) local polynomials (Rajagopalan and Lall, 1999, Grantz, 2003), and local weighted polynomials (Loader, 1999). Nonparametric models take the form:

$$Y_i = \mu(x_i) + \varepsilon_i \qquad \text{Eq. 3.2}$$

In the K-NN and local weighted polynomial (LOCFIT) methods, the function $\mu(x_i)$ is estimated by fitting a polynomial locally to find Y. The value of $\mu(x_i)$ at any point 'x*' is found by fitting a polynomial of order 'p' to a small set of neighbors to x*. Once the neighbors have been identified, ensemble forecasts can be generated using one of two methods:

1) The neighbors can be resampled based on a weighting scheme giving more weight to the closest neighbors and less weight to the farthest, which generates an ensemble (Rajagopalan and Lall, 1999, Singhrattna *et al.*, 2004).
2) An estimate of the error variance and mean of the dependent variable can be found fitting a polynomial to the neighbors (Rajagopalan and Lall, 1998). The ensembles are formed by adding the estimate of the error variance to the mean fit estimate.

3.1.2 Current USBR forecast models

The USBR internally generates the forecasts to be used for its decision support system and to be issued to water users. The forecasts are used to determine the expected total water supply available (TWSA) for the upcoming dry season. The TWSA is a volumetric forecast that includes computed unregulated streamflow, return flow, and carryover storage (Watershed Assessment, 2003). The USBR employs both a stochastic model and a deterministic ensemble streamflow prediction (ESP) model (Christopher Lynch, personal communication) to determine runoff.

The statistical model is a multiple linear regression model that uses precipitation (anticipated as well as antecedent), April 1st SWE, and antecedent runoff conditions. By antecedent, we are referring to conditions from Oct. 1st through the day the forecast is generated. The model forecasts total runoff from October 1st to July 31st at the five reservoirs and three other locations to be used for internal decision making. Residual remaining runoff volume at the forecast site from the forecast date to July 31st is calculated by subtracting the runoff that has already passed the point from Oct. 1st to the forecast date (IOP, 2002). The April through July runoff is calculated using linear regression. The remainder of the runoff occurring in August and September is computed using recession. The statistical model is updated monthly and included in the TWSA that is officially released to stakeholders on April 1st of each year. Proration levels are adjusted monthly, or biweekly if needed, to reflect the new information. This model's forecast ability is compromised by the usage of estimates of future precipitation. The return flow portion of the TWSA is deemed to be constant from year to year regardless of streamflow conditions. This is because much of the return flow as measured at Parker gage is from several large irrigation districts located upstream. Many of the upstream stakeholders have significant nonproratable entitlements that generate a constant return flow. Finally, the storage portion of the TWSA is calculated using current storage of the reservoirs.

The hydrological simulation model of the basin uses the ESP traces to run scenarios of possible daily runoff. The traces are generated for seventy-eight locations in the basin (according to the 2001 traces). The model is energy driven, based on the preceding day's snow pack plus historical meteorological conditions. The meteorological inputs are minimum or maximum temperature and daily precipitation. Each trace is generated using the historical data from one year, say 1950, plus antecedent snow accumulation (Warren Sharp, personal communication). The next trace is created using snow pack information and 1951 meteorological conditions, and so on. The model uses historical data from 1950-2003 and generates 55 ensembles. The assumption is that events of the past are representative of events that will happen in the future. The forecaster then must choose which years he or she believes to have similar conditions to the year that is being forecast. The forecast is updated each day as the previous day's values are added to the historic record. Thus, the skill of the traces improves as the season progresses. The chosen years are then input into a hydrological support model. The number of years in the data set is not significantly long or necessarily representative of the entire range of data variability. It is not capable of generating flow scenarios not seen in the historical record.

3.2 Yakima Statistical Seasonal Forecasts: Modified K-NN approach

As noted in the second chapter, the predictands for the Yakima forecasts are the principal components. The beginning of this chapter noted that Mt. Rainier perturbs the local atmospheric circulation patterns in the basin, causing a marked difference in the precipitation distribution between the upper and lower portions. Therefore, a separate PC analysis was performed for the three sites in the upper (northen) part of the basin (Keechelus, Kachess, and Cle Elum) and for the lower (middle-western) region (American, Bumping, and Rimrock). The climate correlation and composite maps were not noticeably different from those generated using a PC comprising all six sites. However, because of the physical

characteristics of the basin, the actual streamflow forecasts employ the first components garnered from the separate analyses of the upper and lower regions. Figure 3.1 shows the component distribution. For the upper portion, the first PC explains 98% of the variance. The first PC of the lower region encompasses 97% of the variance. These PCs form the predictand variables that are used indigenously within the forecast model.

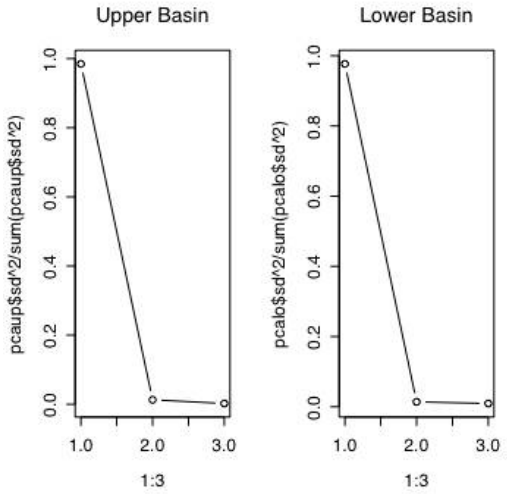


Figure 3.1: PCA of the upper and lower subbasins of the Yakima Basin. First PCs are the predictands to be forecast

The forecast model described below is a variation developed by Grantz (2003) of the original nonparametric modified k-nearest neighbor (K-NN) algorithm implemented by Prairie (2002). The LOCFIT algorithm developed by Loader (1999) was used for fitting the local regression. Grantz (2003) used the K-NN approach to model seasonal streamflow in the Truckee and the Carson basins. This method was also implemented by Singhrattna et al. (2004) for forecasting the Thailand monsoon. The model generates ensemble forecasts of the PCs described above from the noted physical relationships between winter (or fall) teleconnection patterns and spring runoff. The predictors used are: 1) the 500mb geopotential height difference between the Canadian ridge and the Aleutian low, 2) the "warm" pool of SST's off the PNW coast, and 3) the various regionally averaged SWE.

The model generates ensemble forecasts by retaining residuals derived from the local regression between the PC predictand and the predictors. The best choice of neighborhood size (K) and the order of the polynomial were obtained using Generalized Cross Validation (GCV). In this technique, the data are divided into two sets: one containing all the data points and the other consisting of all the points except the one being estimated. The dropped points form the testing set against which the goodness of the fit of the estimated points are measured. The weighted prediction error (residuals from the comparison the estimate and the actual value) is averaged over the neighborhood about the point being estimated (Perry and Stanley, 1995). This method ensures that the influence of any single extreme outlier point on the model is minimal.

The residuals are weighted based on their Euclidean distance to the predicted point, with the closest neighbors receiving more weight than the farthest. The weighted residuals are then added to the mean forecast generated using LOCFIT to create the ensemble members. This ensemble approach is capable of producing values not seen historically, allowing water managers multiple planning scenarios. The modified K-NN algorithm is outlined below:

1. The size of the neighborhood (K) and the polynomial order (p) are determined using GCV.
2. For each observed data point, x_{new}, the K-nearest neighbors are identified and a polynomial of order *p* is fitted. The mean value of the predictand, Y_{new}, is estimated from the fit at each observed data point. The residual ε_j is computed and retained for each data point. This step is the LOCFIT process.
3. The neighbors are weighted using the weight function (Grantz, 2004):

$$W(j) = \frac{1}{j \sum_{i=1}^{k} \frac{1}{i}} \qquad \text{Eq. 3.3}$$

Neighbors closest to x_{new} are given more weight than those farther away. The number of neighbors whose residuals are resampled may not be the same as the number of neighbors used in the local polynomial fit of step 1. Our forecasts used a neighborhood size around the order of the square root of N, the number of years in the data set.

4. The selected neighbor's residual is then added to the mean local regression ($Y_{new} + \varepsilon_j$) to obtain an ensemble member.
5. Repeat the entire process as many times as necessary to produce the desired number of ensemble members (100 in this case) and to produce a probability distribution function that does not change with more sampling.
6. Once the entire ensemble set has been generated, the forecasted PCs must be transformed back to the original space and unscaled to recapture the flows at each of the six sites. This is done by multiplying the forecasted PC (α hat) by the corresponding eigenvector and the site standard deviation, and then adding the site mean.

$$\hat{\mathbf{x}}_{site1} = (\hat{\alpha}_1 * e_1^1 * \sigma_{site1}) + \mu_{site1} \qquad \text{Eq. 3.4}$$

The entire process is more easily visualized in two-dimensional space. Figure 3.2 displays the local regression between the January-March geopotential height difference and the first PC of the upper portion of the basin. The local regression fits the relationship between the predictor and the predictand and is the expected mean value for the forecast. The local regression is denoted by the solid, curvy line through the scatter plot of the points. The predictand is estimated (Y_i*) from the all the predictor values (x_i* = 500 mb geopotential height difference) in the neighborhood except for the predictors belonging to the point we are attempting to predict. A hypothetical neighborhood is outlined by the dashed rectangle. The size of the neighborhood "K" was defined for this model as being approximately:

$$K = \sqrt{n} \qquad \text{Eq. 3.5}$$

where “n” is the length of the data set, 54 years. The size of the neighborhood was tuned about this value of k to determine how many neighbors’ residuals to keep. In practice the size of the neighborhood ranged from 7 to 11 neighbors.

Points within the local neighborhood are resampled according to the weighting scheme described earlier. The residual ε_l* from the selected resampled neighbor, circled in red, is added to the mean estimate of the predictand point to perturb the model. This forms the first member of the ensemble forecast. The whole process is repeated an arbitrary number of times (the choice was 100 times for this study) to obtain an ensemble forecast. Each year of the ensemble forecast contains 100 members such that the full dimensions of the ensemble forecast is 100 members by 54 years. The eigenvectors (Eofs) were retained from the PCA. The first set of Eofs multiplied by the ensemble of predicted first PCs returns the model to the original streamflow space. In other words, this step recaptures the spatial distribution of volumetric flows at each of the six sites. The resulting vectors at each site are then unscaled by multiplying by the standard deviation of the streamflow time series from each site and then adding the mean. In this way, streamflows are reconstituted from the forecasted PCs. Analysis testing the skill of the forecasts is performed by comparing these reconstituted flows with the observed.

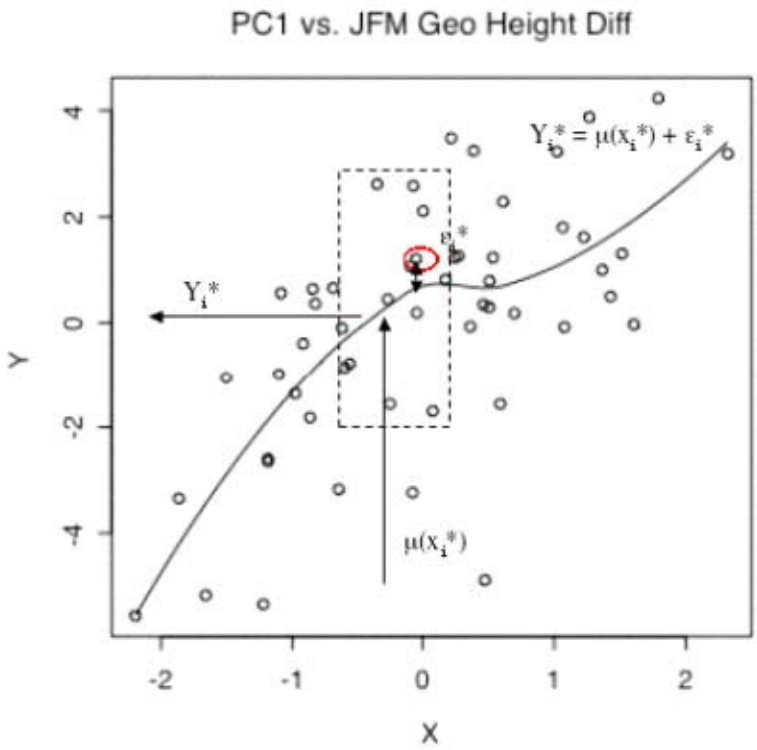


Figure 3.2: 2D local regression and residual resampling. Y* is the mean local polynomial fit, x* is the predictor (500 mb geopotential height difference between Aleutian low and Canadian ridge), and ε* is the residual to be added to Y* to generate an ensemble member.

3.2.1 Model Verification

A forecast is not useful unless a quantification of its performance can be made. As the forecasts made are of the statistical type, it makes sense to measure the forecast's performance in terms of probability. A water manager would like to know how close to the observed value the forecast fell and how probable it is that, say, the 25th percentile of the forecast will occur. Several measures of skill are widely used by the climatology community:

1. The Ranked Probability Skill Score (RPSS)
2. Likelihood Function
3. Correlation between the median ensemble member and the observed value.

The ranked probability score (RPS), developed by Epstein (1969, as cited in Murphy, 1972) is a measure of the sum of the squared Euclidean distance between the forecast

cumulative probabilities and the observed's cumulative probabilities that are divided into certain categories. Often, three categories are chosen using the arbitrary divisions of above normal, normal, and below normal. It is up to the climatologist to quantitatively set the categories according to chosen percentiles of the observed historical data. For this study, the 80th and 20th percentiles marked the divisions between the three categories. The RPS is defined as the following for n equally probable categories (Kumar *et al.*, 2000):

$$\mathrm{RPS} = \sum_{m=1}^{n} (Y_m - O_m)^2 \qquad \text{Eq. 3.6}$$

Y_m is the cumulative probability that a particular ensemble member falls into one category and O_m represents the cumulative probability of the observed values. RPS has a negative orientation that means that smaller scores are better. The RPS values range within the unit interval [0,1] (Murphy, 1972). The probability distribution of the observation is 100% for the category that was observed and is zero for the other two categories. A perfect forecast has a RPS of zero. A separate RPS is then calculated for a forecast that is based on climatology. Climatology is the value one would expect based only on the long-term historical data statistics that has an assumed normal distribution with a mean of 0. It can be likened to generating a forecast by flipping a coin, sometimes the observed value will be forecast and sometimes it will not. The RPS of the ensemble forecast generated with our selected set of predictors is then compared with the RPS of a hypothetical forecast based on climatology.

The comparison of the two RPS is the RPSS. It is a measure of how often an ensemble member falls into the category of the observed value and compares that to climatology's performance (Lea and Saunders, 2004). The RPSS is defined as:

$$\mathrm{RPSS} = 1 - \frac{\mathrm{RPS}_{\mathrm{forcast}}}{\mathrm{RPS}_{\mathrm{climatology}}} \qquad \text{Eq. 3.7}$$

The RPSS ranges from -∞ to +1 where +1 indicates an ensemble forecast with perfect skill, i.e., one that perfectly matched the observed. A negative RPSS or one of 0 implies that the ensemble forecast did not beat climatology.

Another measure of the skill of a forecast is the likelihood function. Likelihood is the hypothetical probability that an event that has occurred would have a specific outcome (Weissten, 1999). In our case, it is the likelihood that an ensemble member falls into one of the three categories that were defined earlier. Each member of the ensemble forecast is assumed to be independent of the others and so their individual likelihoods are independent. The likelihood function for the entire ensemble forecast is defined as the sum of the individual likelihoods of each ensemble member (Kooijman, 1999):

$$L = \prod_{i=1}^{N} \hat{P}_{ji} \qquad \text{Eq. 3.9}$$

The likelihood skill score is formed by comparing the likelihood function of the ensemble forecast with that of the climatology forecast:

$$L = \left[\frac{\prod_{i=1}^{N} \hat{P}_{ji}}{\prod_{i=1}^{N} \hat{P}_{cl\,ji}} \right]^{\frac{1}{N}} \qquad \text{Eq. 3.10}$$

N is the number of categories. The range of the likelihood skill score is from 0 to the total number of categories, which is three for our forecasts. The score of the forecast decreases with the decreasing forecast skill. A skill of one indicates that the ensembles performed no better than climatology. If the score is worse than one, the climatologist had better find another job.

3.3 Ensemble Forecasts

The following sections present a select set of the ensemble forecasts that have the best skill in predicting the observed flows. While the forecasts for the upper section and the lower section of the basin were performed separately, they are presented together. The performance of the predictions is measured using the skill metrics described in the section, Model Verification. These forecasts have the potential to be used in determining unregulated flows for six points in the basin in the TWSA forecasts. Discussion of the implications of the forecasts in certain scenarios, such as droughts, to water management and salmon survival is in successive chapters.

3.3.1 Ensemble forecasts using Dec-Feb climate predictors and April 1 SWE

The two forecasts for the upper and lower regions of the Yakima basin were formulated using the predictor set of the December through February climate indices of the pressure difference between the 500mb geopotential heights of the Aleutian low and the Canadian ridge seen in Chapter 2. While the warm SSTs seen off the PNW coast are mostly a response to the pressure systems, some of the SST information is independent of the atmospheric forcing and adds skill to the forecasts. The ensemble forecasts for this period also include the April 1st SWE data of the respective upper and lower portions of the basin. For the historical period, the majority of snow accumulation is by April 1st. Because of this, a forecast using this predictor set would not be available to water managers until the first week of April. The USBR issues its official forecast to farmers and the subsequent proration levels during the first week of April. Thus, generating a set of reliable forecasts for this time frame is of the utmost importance. The best fit for the upper region ensembles was found using 11 neighbors. The best fit for the lower also utilized 11 neighbors.

The skill of the ensemble is different for each of the six sites we are attempting to forecast. In general, the predicted streamflows for the upper portion of the basin were closer to the observed than the predictions made for the lower region. The correlations between the

median ensemble values for each site and the observed streamflow at that site are listed in the following table.

Kachess	Keechelus	Cle Elum	American	Bumping	Rimrock
0.89	0.90	0.88	0.84	0.84	0.82

Table 3.1 Correlations between observed values and median DJFA predictions. The highest correlations are seen at the three upper sites.

The differences in correlations between the upper and lower regions are most likely due to the fact that what the ensembles actually sought to recreate were the first PCs for the upper and lower sections. The first PC for the upper reservoirs explained 98% of the variance, whereas the first PC for the lower sites only captured 97% of the variance. In recapturing the spatial distribution of the flows for each site by multiplying by the relevant Eofs, further information is lost. Unscaling the flows also adds a small amount of error to the predictions.

The following figures (Figure 3.3) display the ensemble forecasts with the observed streamflow values (marked by the thick black line). The ensemble forecast for each year is presented as a box plot. The box plots provide a visualization of the data spread and other relevant statistics. The line in the middle of the box represents the median ensemble value. The width of the box plot gives the interquartile spread of the data. That the median line does not fall directly in the middle of the box indicates that data distribution is skewed. Finally the whiskers extend to the extreme values of the data, which is approximately 1.5x the interquartile distance. Data points falling outside the whiskers are outliers and are noted as open circles. The horizontal lines correspond to the 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} percentiles of the historical observed data.

In looking at the ensemble forecasts (Figure 3.3), it becomes apparent that prediction is better for the dry years than for the wet years. There is a tendency to overpredict values in the lower (southern) region (RIM, AMER, BUMP). The box plots of the ensembles of the

upper (northern) region (KACH, KEE, CLE) are tighter around the historical data. The most extreme years (+/- 10% of normal) are of great interest to water managers. During high flow years, managers must operate the reservoirs on flood control, but are not necessarily concerned with the ability to meet agricultural and biological flow demands. During drought years, managers are especially concerned with accurate forecasts because of the need to determine proration levels for various water users as well as meet a minimum instream flow early in the year. A general downward trend is also noted in the natural streamflows. This trend might be due to the PDO regime shift from the cool phase to the warm phase in 1977. This year marked the most severe drought in the data record and is also the turning point at which the downward trend begins. The ensemble predictions manage to capture the shift in the streamflows to a good degree.

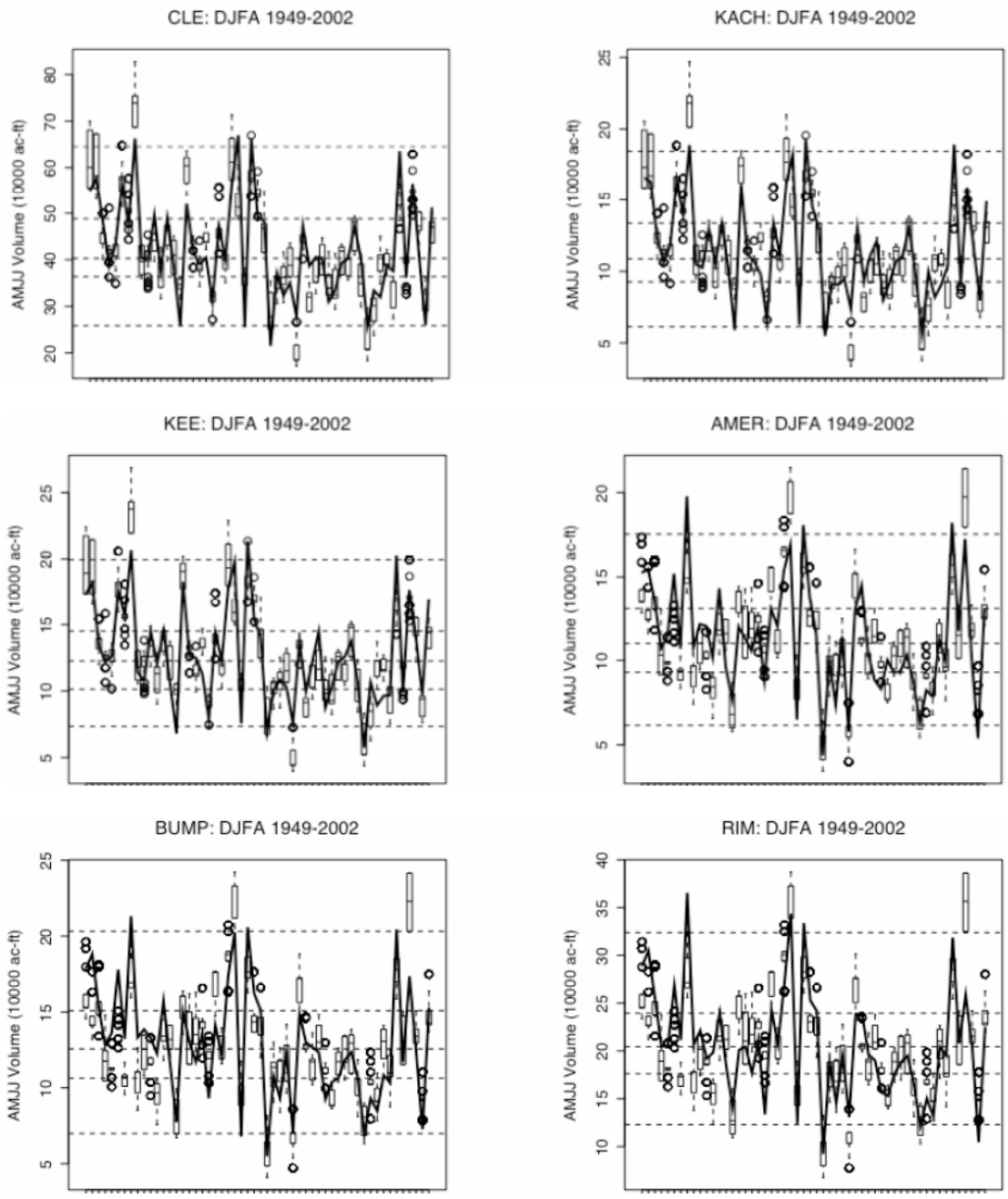


Figure 3.3: Ensembles for DJFA predictor set. The dashed horizontal lines, from the bottom, correspond to the 5th, 25th, 50th, 75th, and 95th percentiles. The solid black line corresponds with the historical spring runoff at each site. Box plots represent the ensemble forecasts.

The following set of plots (Figure 3.4) presents a snapshot of the ability to capture extreme flow events for each of the six sites. The six sites share similar extreme years, but there are differences. The top 10^{th} percentile years and bottom 10^{th} percentile years are different at each of the six sites. The disparity seen between the years is indicative of the amount of variation existing in the basin, even though the basin is small. Differences could be due to distribution of snow pack that contributes to the sites' individual watersheds, temperature sensitivity, and wind, for example.

At most of the sites, five above average years out of six fall before the PDO regime shift in 1977 from a cool mode to the warm phase. The drought years are not so easily explained. The drought in 1981 might have been influenced by the eruption of Mt. St. Helen in 1980. 1992 was part of a multiyear drought that lasted through 1994. The drought period 1992-1994 was ENSO neutral (COAPS ENSO Index), while 2001 was actually a La Niña year. These two cases exemplify why the ENSO index should not be used as a predictor variable.

The 2001 drought year presented a challenge to forecasters. Both the USBR's ESP forecasts and the ensemble forecasts presented here overpredicted the seasonal streamflow volume. The 2001 predicted and observed values are the farthest right box plot and line in the dry years plots for every site except Kachess and Keechelus. The ensembles for the upper region of the basin have better skill than the lower basin's forecasts. Refer to Figure 3.4. Previous La Niña years generally weakened the Aleutian low and the Canadian ridge, however 2001's predictor set shows the relative pressure difference between the two centers to be anomalously high. The PNW coastal SSTs were anomalously warm, but the strong relative ridge mitigated the effects of any increased convection. As noted in Chapter 2, SST relationships were absent in the composite plots for extreme years. However, including the SST indices as a predictor did lead to a slight improvement in the overall skill of the forecasts.

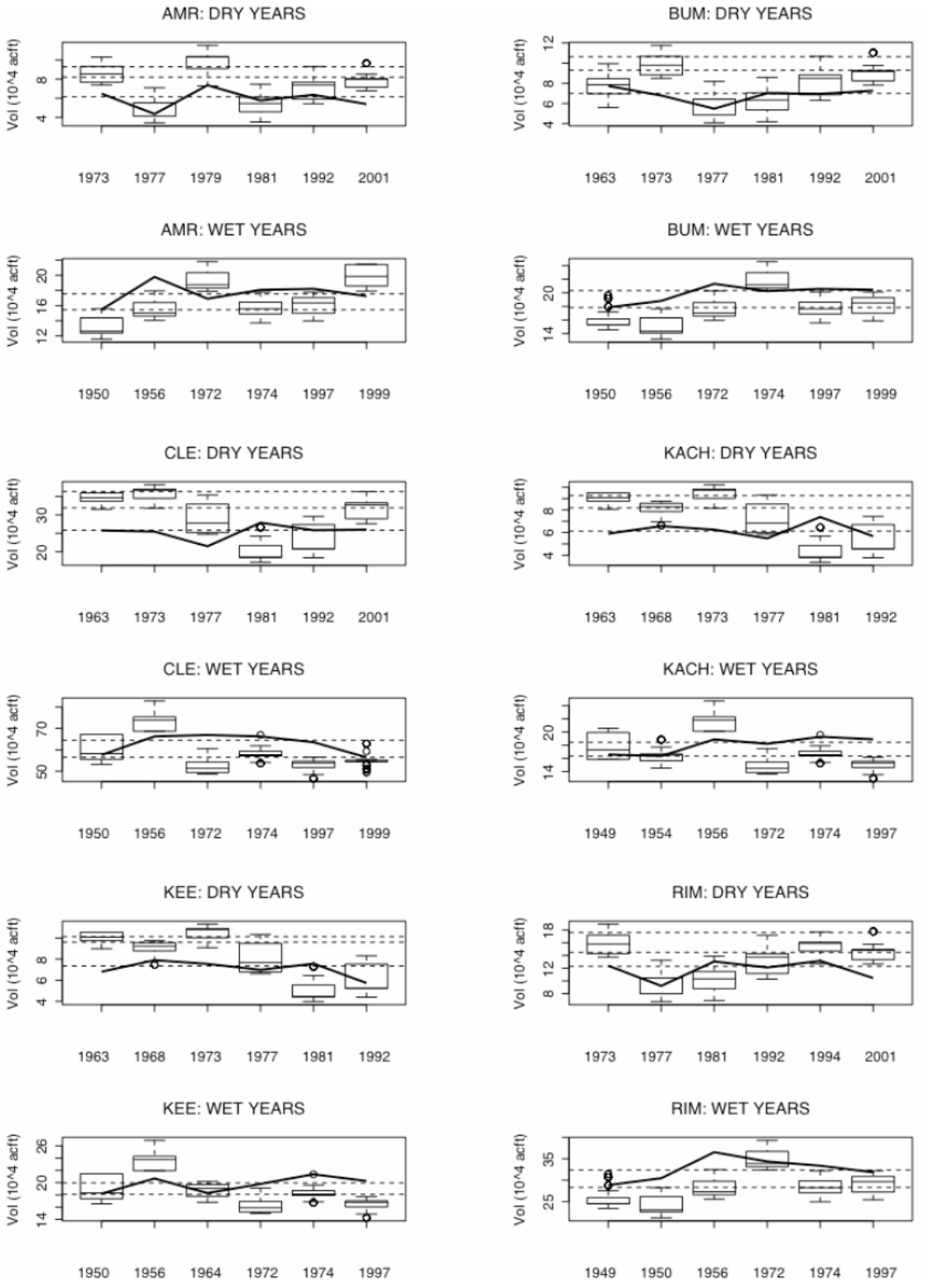


Figure 3.4: DJFA predictions (box plots) vs. observed (solid black line) during extreme years. On Dry Year plots, the dotted lines from bottom correspond to the 5th, 15th, and 25th percentiles of the historical streamflow. On Wet Year plots, the dotted lines from bottom correspond to the 90th and 95th percentiles of the historical streamflow.

The RPSS for each ensemble year is presented below in Figure 3.5. A different denomination for testing the skill of the ensembles in wet years and dry was used than looking at the extreme scenarios above. The definition used in RPSS for a wet year based on the historical data includes those years falling above the 80^{th} percentile. Dry years are defined as falling below the 20^{th} percentile. The farthest left box plot shows the model performance over all years. The middle box plot corresponds with wet years and the far right box plot demonstrates the skill of the dry years.

Overall, the model performs fairly well. For all years and all the sites, the interquartile range of the score is well above 0. The median value is difficult to see as it is skewed toward the upper portion of the interquartile range. This is indicative of the model's tendency to overpredict. As noted earlier, it is evident from the score that the upper sites forecasts perform better than the lower sites. In general, prediction skill of the dry years is higher than that of the wet years. The interquartile range of the dry year box plots is much tighter and closer to 1 than those of the wet years.

The likelihood skill scores reinforce the forecast skill seen in the RPSS. Refer to Figure 3.6. The likelihood that the ensembles perform better than climatology is synonymous across all the sites, with better skill reserved by the forecasts for the upper reservoirs. The interquartile range of the box plots for all years and all sites is above one. The same skew visible in the box plots of the RPSS is present in the likelihood skill scores. The 5^{th} percentile whisker for many of the sites extends to or below 1, indicating that the lowest extreme members of the ensemble plot have no skill. The median ensemble values for all years and all sites are grouped toward the third category, indicating that it has a high likelihood of capturing the observed.

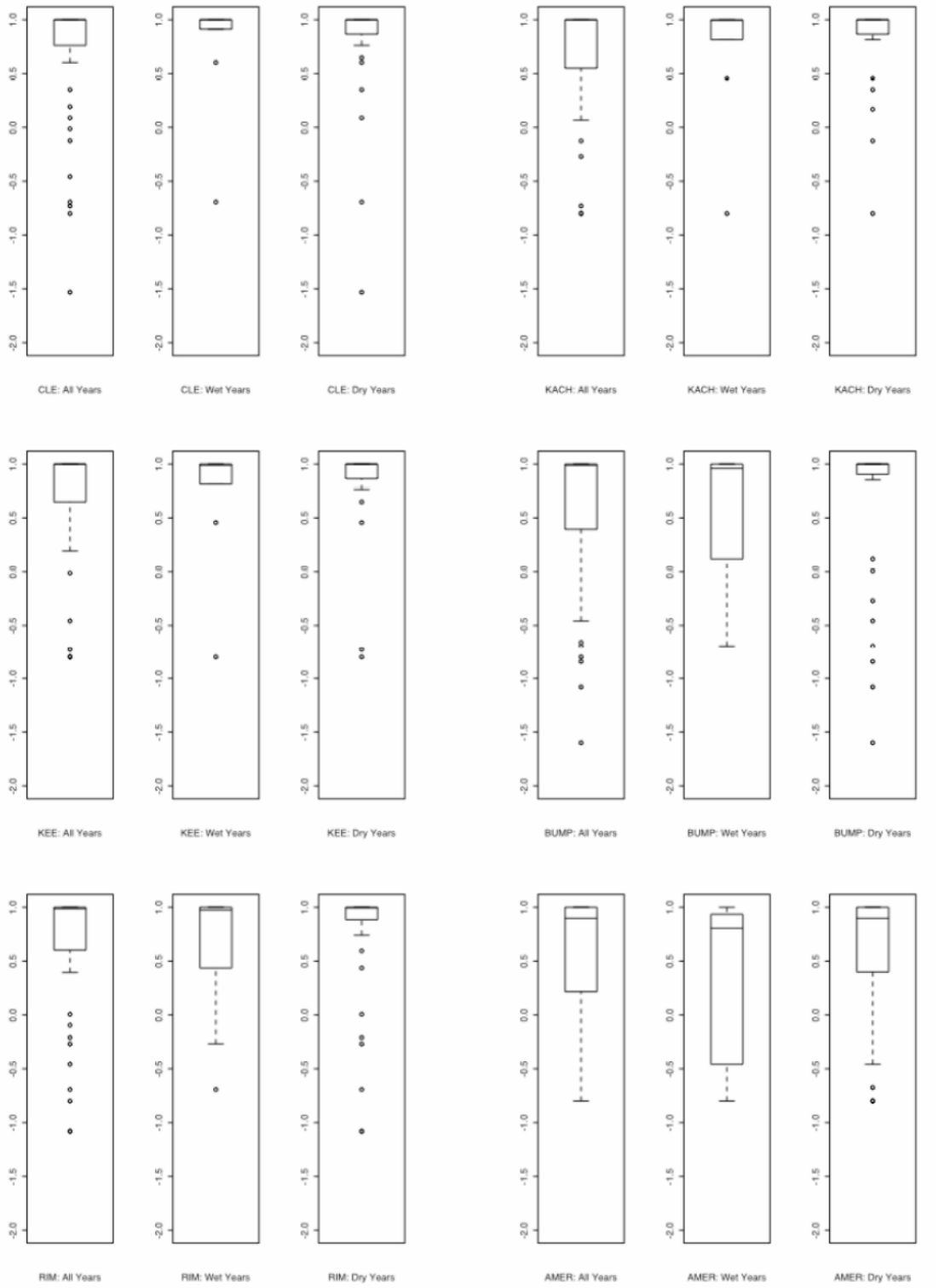


Figure 3.5: Rank Probability Skill Score (RPSS) of DJFA ensembles. The far left plot displays the scores over all years of the forecast, the middle plot displays scores for years falling in the 80th%, and the right plot displays scores for years falling in the 20th%. Tighter box plots closer to 1, indicate good skill. RPSS ranges from -∞ to 1.

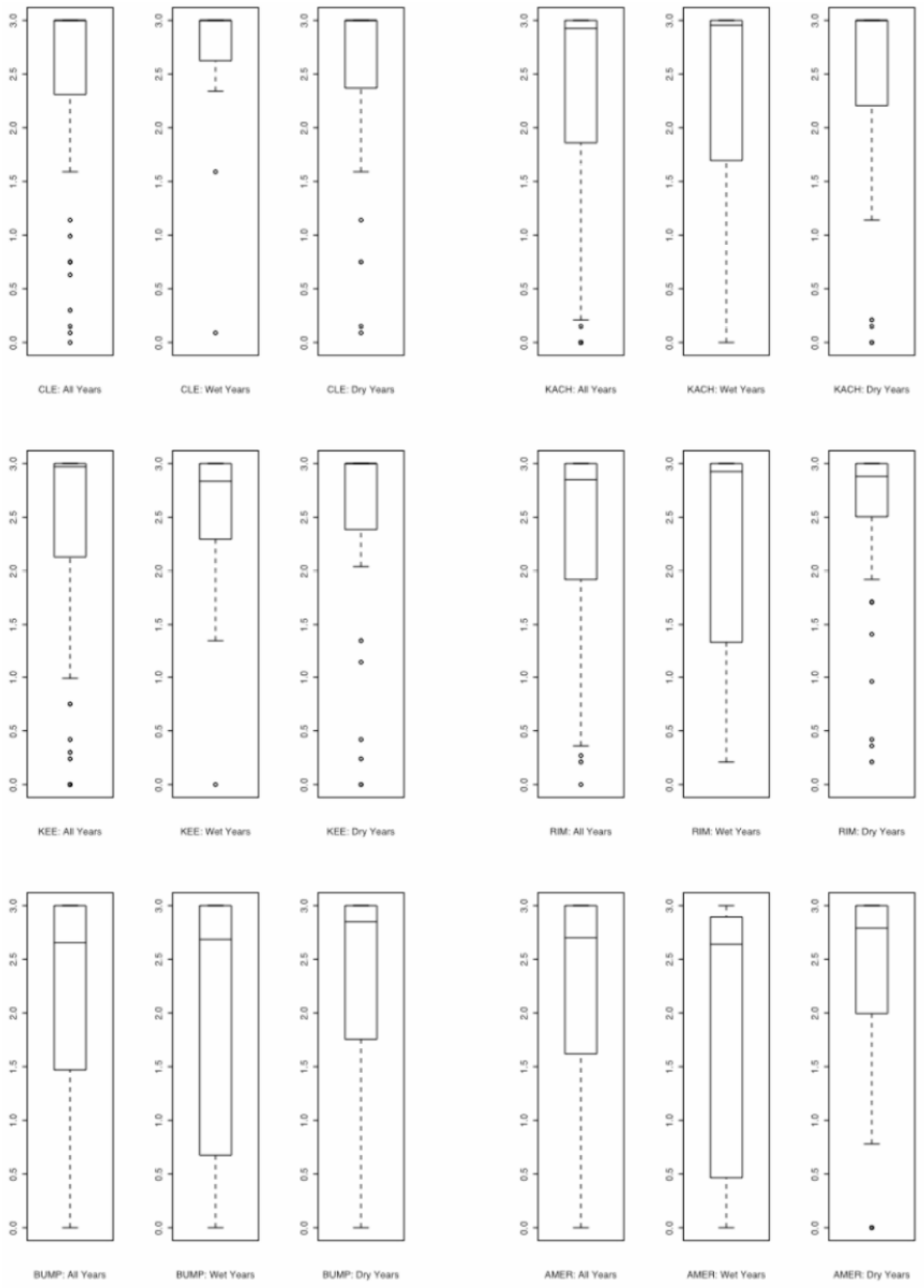


Figure 3.6: Likelihood skill scores of DJFA ensembles. The far left plot displays the scores over all years of the forecast, the middle plot displays scores for years falling in the 80th%, and the right plot displays scores for years falling in the 20th%. Tighter box plots closer to 3, indicate good skill. Likelihood scores range from 0 to 3.

The following table presents the median ensemble member's RPSS and likelihood skill score. It also demonstrates that the likelihood skill is a little more sensitive to nuances than the RPSS.

	Kachess	Keechelus	Cle Elum	American	Bumping	Rimrock
RPSS	1	1	1	0.97	0.97	1
Likelihood	2.97	2.98	2.99	2.45	2.90	2.82

Table 3.2: Skill scores of median DJFA ensemble members in all years. An RPSS value 1 indicates high skill, as does a score close to 3 for the Likelihood score.

The ensembles using Dec-Feb climate predictors with April1 SWE provide water managers with a skilled outlook of spring runoff volume. The proximity of the Yakima basin to the Pacific Ocean and the fact that the majority of precipitation falls as snow lends further credibility to a forecast based on this predictor set. Furthermore, the K-NN algorithm is easily adapted to different predictor sets, allowing water managers to test multiple scenarios.

3.3.2 Ensemble forecasts using Dec-Feb climate predictors and March 1 SWE

The previous set of forecasts was generated using the December through February climate predictors and April 1st SWE. While this is the forecast that would be officially released to farmers and other water users, the USBR begins issuing forecasts internally around the first of each month beginning in January. It would be beneficial to the agricultural industry to be able to issue a forecast earlier than the beginning of April. This way, farmers and fish and wildlife managers would have a longer planning period for adapting to anticipated water supply. A preliminary forecast using the December through February climate predictors and March 1st SWE could increase the lead time of planning by a month. The following forecasts use March 1st SWE information. The change in predictors necessitates selecting a different number of nearest neighbors to get the best fit. The

ensemble generated for the upper reservoirs was found to have greatest skill using eight neighbors, while the forecasts for the lower sites utilized ten neighbors.

There is a slight loss of skill between issuing a forecast using March 1st SWE information and April 1st information. This can be expected as skill generally decreases as the lead time at which the forecast is issued, increases. A forecast issued at the beginning of March has increased the lead time by two months. The following table presents the correlation between the median ensemble forecast volumes with the observed streamflow volumes. Also included are the skill metrics, RPSS and Likelihood, for the same scenarios discussed in the previous forecast. The RPSS and Likelihood scores are for the median ensemble member. As noted, a wet year is greater than the 80th percentile of the historical data and dry are years are below the 20th percentile.

	Kachess	**Keechelus**	**Cle Elum**	**American**	**Bumping**	**Rimrock**
Correlation	0.85	0.83	0.85	0.76	0.75	0.78
RPSS:						
All Years	0.95	0.97	0.95	0.95	0.95	0.96
Wet Years	0.94	0.96	0.95	0.52	0.97	0.92
Dry Years	0.88	0.97	0.98	0.90	0.95	0.95
Likelihood:						
All Years	2.25	2.48	2.45	2.52	2.47	2.35
Wet Years	2.50	2.48	2.60	2.35	1.80	2.17
Dry Years	2.40	2.50	2.50	2.50	2.48	2.47

Table 3.3: Comparison of DJFM ensemble predictions with observed. Correlation is between the median ensemble member and the observed value. Skill scores are for all ensemble members, using the previous convention.

Again, the skill of the ensembles produced for the upper reservoirs is greater than that of the lower sites. Of notable importance for water managers is the fact that the skill of predicting dry years is still quite good. There are two physical explanations for why such information is available: 1) the time frame of largest snow accumulation is typically February for the basin, and 2) the PNA is a long persisting teleconnection whose influence on the stormtracks affecting the PNW region extends back to the fall of the proceeding year. The

PNA is strongest in the mid to late winter periods (December through March) and can provide valuable information to water forecasters.

Presented on the following page (Figure 3.7) are the extreme year plots for all locations using the years described in the previous section. The plots of the ensembles, RPSS, and Likelihood Score are included in the Appendix. The ability of capturing a dry year is stronger in March than in April. In the drought years presented here, it is possible that all the information from the SWE is present by March 1st. In some of the drought years, decay of the snow pack might begin in March due to the temperature sensitivity of the basin. Additional snow throughout March is unlikely to significantly bolster the SWE if warmer temperatures are occurring. In some drought years, such as 2001, the median ensemble members were closer to the observed values using the March 1st SWE than the median ensemble values from the DJFA forecasts.

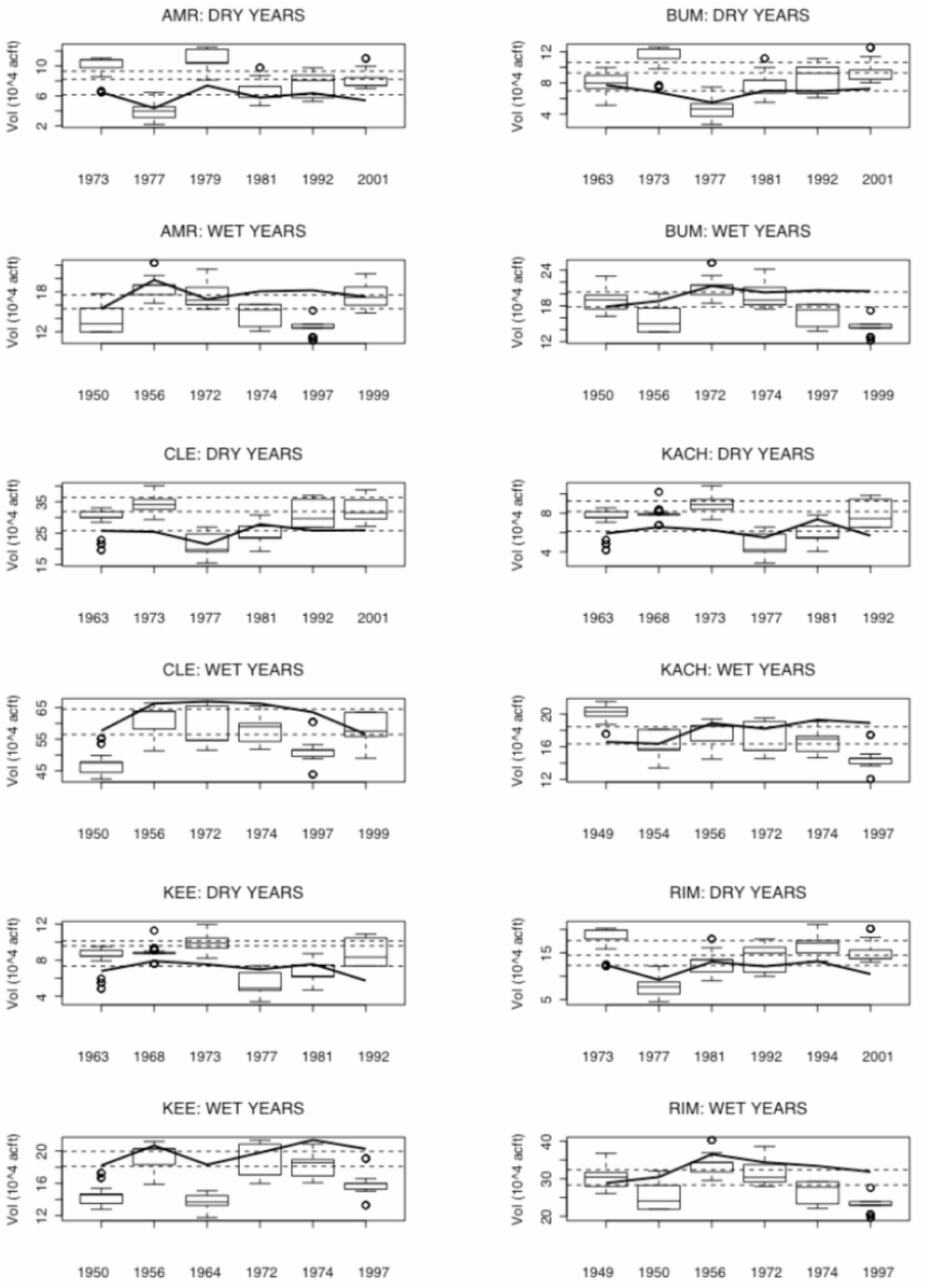


Figure 3.7: DJFM predictions (box plots) vs. observed (solid black line) during extreme years. On Dry Year plots, the dotted lines from bottom correspond to the 5th, 15th, and 25th percentiles of the historical streamflow. On Wet Year plots, the dotted lines from bottom correspond to the 90th and 95th percentiles of the historical streamflow.

3.3.3 Ensemble forecasts using Nov-Jan climate predictors and Feb 1st SWE

This ensemble forecast gives water managers an increase in lead time of three months. The climate predictor for the November through January time period only includes the relative pressure height difference between the Aleutian low and the Canadian ridge. The correlation between the PNW coastal SST and the first PC of the streamflow was not significant enough to be used in the forecast, even though a weak relationship (r = 0.20) did exist. The correlation between the geopotential height difference and the PC is not as strong as that of the December to February climate window. However, it is significant enough to provide water managers insight into future streamflow volumes. The February 1st SWE values also enhance the skill of the ensemble forecasts. The predictor distribution is noticeably different between the upper and lower forecasts. The upper forecast had the best fit using nine neighbors while the lower ensembles required only seven neighbors.

The figures presented at the end of this section show the extreme year plots with the observed values for those years, using the same set of years as in section 3.3.1. The following table presents the correlations between the median ensemble member and the observed streamflow volume. Also included are the various scenarios for the RPSS and Likelihood score that have been consistently employed. Actual plots of the ensembles, the RPSS, and the Likelihood score are included in the Appendix.

	Kachess	**Keechelus**	**Cle Elum**	**American**	**Bumping**	**Rimrock**
Correlation	0.69	0.66	0.71	0.75	0.68	0.75
RPSS:						
All Years	0.51	0.73	0.65	0.65	0.51	0.80
Wet Years	0.52	0.42	0.60	0.70	0.75	0.90
Dry Years	0.70	0.55	0.77	0.50	0.75	0.96
Likelihood:						
All Years	1.80	2.10	2.00	1.85	1.80	2.10
Wet Years	1.65	1.55	2.20	1.65	1.90	2.20
Dry Years	2.00	2.50	2.20	1.95	2.00	2.40

Table 3.4: Comparison of NDJF ensemble predictions with observed. Correlation is between the median ensemble member and the historic value. Skill scores are measurements of all ensemble members.

Surprisingly, there is a shift in the ability of the ensemble forecasts of the upper sections to predict the observed streamflow. The lower site forecasts actually perform as well as the upper model. The influence of climate (the 500 mb geopotential height in November to January) plays a pivotal role in drought years. As the climate predictor set is the same for the two models, but the SWE information is not, this suggests that more information about the snow pack is available around February 1 for the lower region. The skill of the median ensemble members is greater in dry years than in wet years. This suggests that much of the information to be garnered from the snow distribution of the lower region in a drought year is present by the beginning of February.

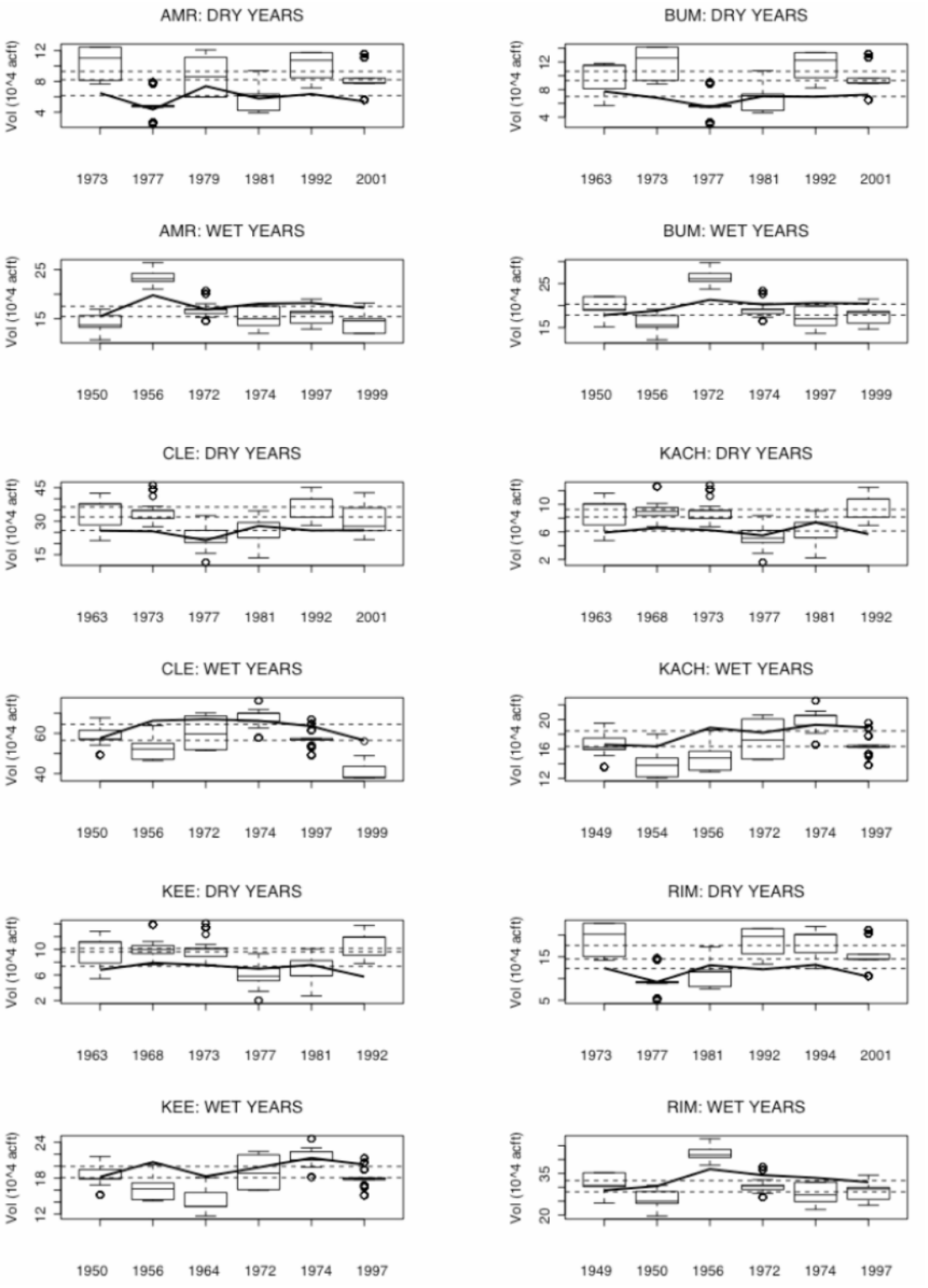
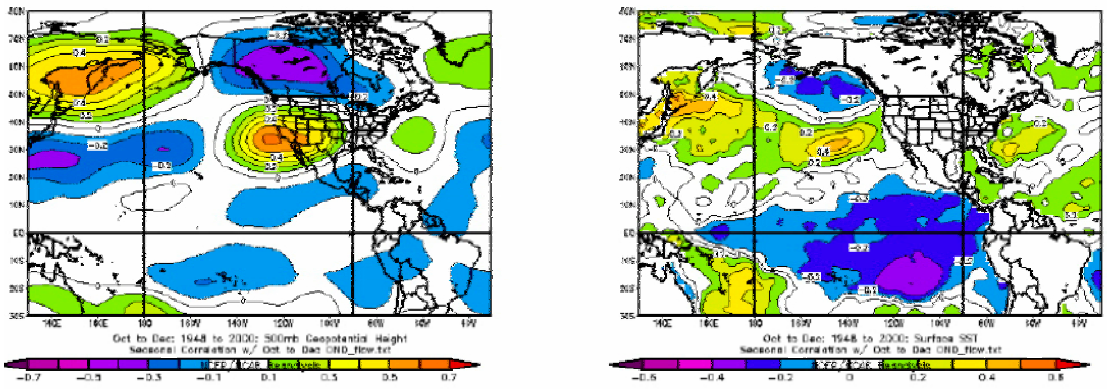


Figure 3.8: DJFM predictions (box plots) vs. observed (solid black line) during extreme years. On Dry Year plots, the dotted lines from bottom correspond to the 5th, 15th, and 25th percentiles of the historical streamflow. On Wet Year plots, the dotted lines from bottom correspond to the 90th and 95th percentiles of the historical streamflow.

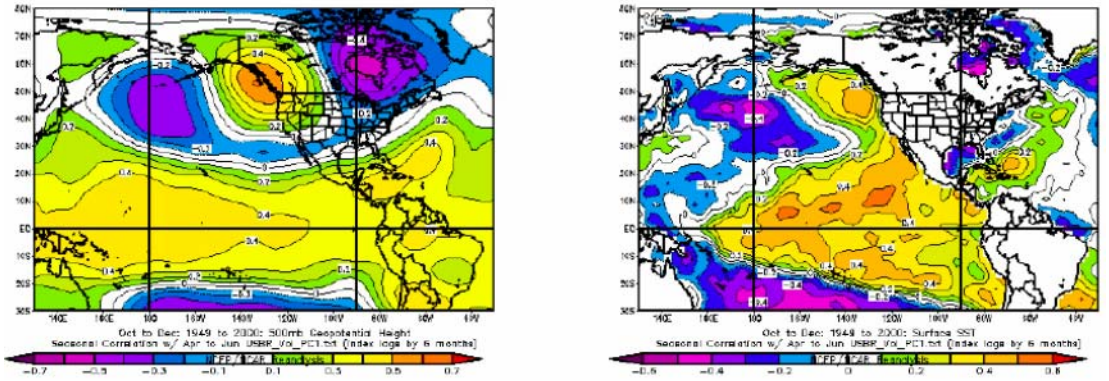
3.3.4 Ensemble forecasts using Oct-Dec climate predictors and comparison

One hope of the USBR is to increase the lead time of the forecasts so that advance warning of adverse conditions might aid planning. To this end, ensemble forecasts were generated using climate information from the OND time frame. No SWE information is available for these forecasts. The dominant climate features at this time are the Canadian ridge and the PNW coastal SST. The Aleutian low did not show any significant correlation with the spring runoff PCs. As noted in the second chapter, there is a small streamflow peak in the OND period. This peak corresponds with the onset of winter precipitation, but temperatures are not yet consistently cold enough to cause the majority of the precipitation to fall as snow. The relationship between the spring runoff peak and the immediate climate conditions is weaker than the relationship between the winter runoff peak and the immediate conditions. This is evident in the correlation maps presented below. While the PNA pattern is evident in the climate correlations with the spring runoff, it is not the main climate feature dominating OND weather. It is for this reason that we feel the OND forecast is more a proxy for soil moisture conditions prior to the spring runoff period. A separate forecast could be made of the OND streamflow to determine if the soil will be saturated before colder temperatures set in. The forecaster would then have an idea that spring runoff might be higher if less infiltration is necessary or that drought conditions might be forming. However, the forecast do present some skill, which is examined below. The best forecasts for all sites incorporated nine neighbors. This OND forecast could be issued at the beginning of January.



(a) **(b)**

Figure 3.9: OND streamflow correlations with OND climate: 500mb height field (a) and SSTs (b)



(b) **(d)**

Figure 3.10: OND correlations with PC1: 500mb height field (c) and SSTs (d)

Finally, the question must be addressed as to whether climate information improves the skill of the forecasts and how the OND forecasts perform. Only the RPSS was calculated for the OND ensembles. In order to test the skill of the forecasts that incorporated climate information, ensembles were generated that used only SWE. The following plot compares the median RPSS scores of all years of the forecasts at all sites for four months. The median RPSS scores were averaged across all sites. The scores for the SWE only forecasts are available only from February onward. The climate forecasts were generated beginning in January, using the OND climate information.

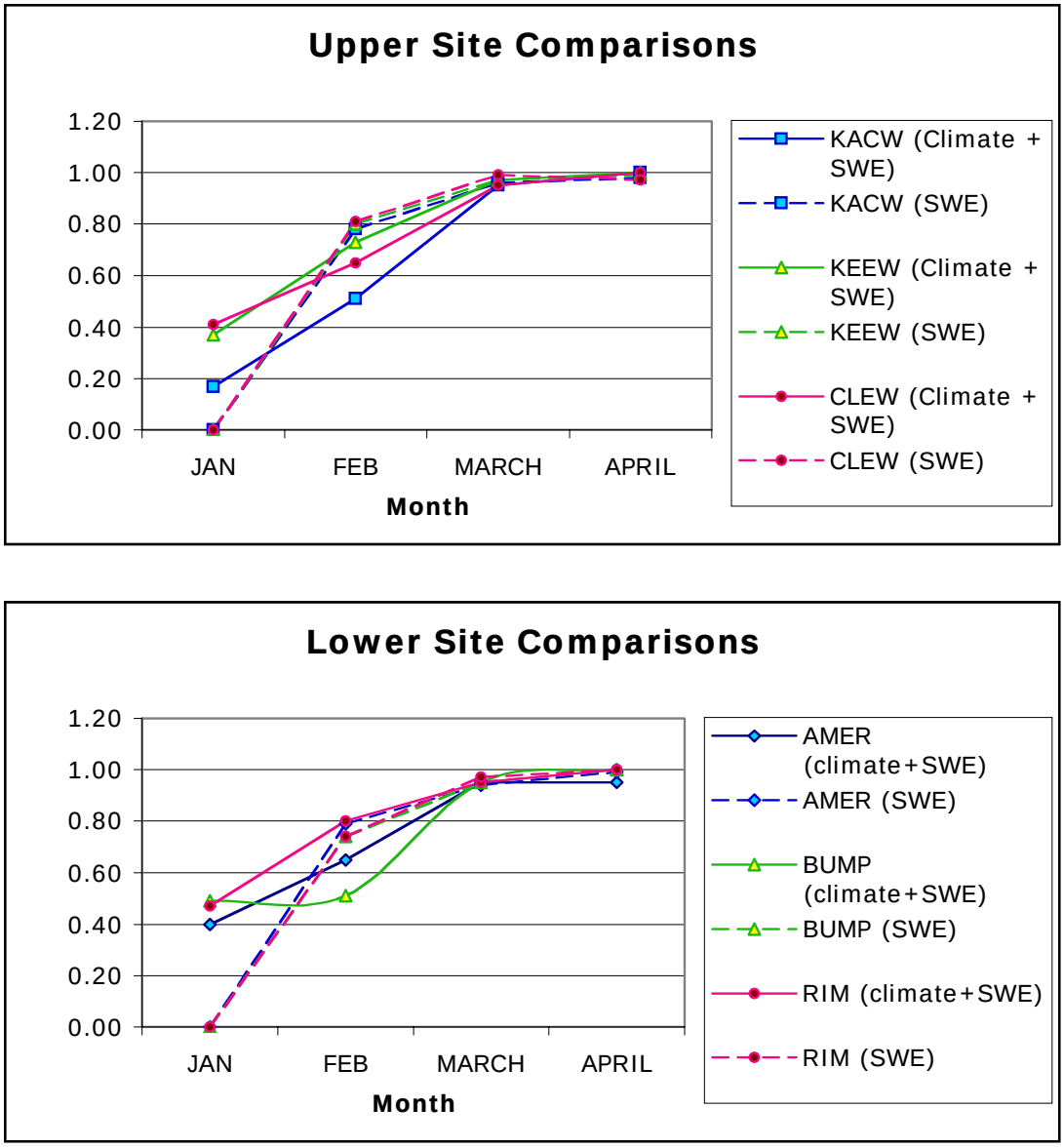


Figure 3.11: Median RPSS values over all years for each site comparing forecasts generated using climate + SWE (solid lines) and forecasts using just SWE (dashed lines)

At first glance, it appears that including climate information in the forecasts is no improvement over forecasts generated using just SWE. The median RPSS score is for all years. In fact, the forecasts using just SWE appear to outperform the forecasts using climate in the NDJ time frame. However, we argue that including climate information provides an advantage to water managers. First, the large scale climate features such as the Aleutian low and Canadian ridge, are responsible for the stormtracks that provide winter precipitation to

the PNW. Thus SWE is related to the PNA pattern. Secondly, in the extreme years, including climate information improves the forecasts over these generated using just SWE. The next set of figures show the forecasts generated using Feb 1st SWE only for extreme years. While the ensembles are pushed in the direction of the actual streamflows, they do not perform as well as the forecasts that used Nov-Jan. climate and Feb 1st SWE. Refer to Figure 3.8 to compare. Extreme year plots for April 1st and March 1st SWE are in the Appendix.

Forecasts using SWE only have a tendency to overpredict the spring runoff in drought years, which could be disastrous for planning and management. Figure 3.8, which displays ensembles using NDJF information, shows that the addition of climate information pushes the forecasts closer to the observed flows. The degree of overprediction is significantly reduced at all sites during drought years when climate information is included. Nor do the SWE-only ensembles provide as large a range of streamflow scenarios as the forecasts using climate.

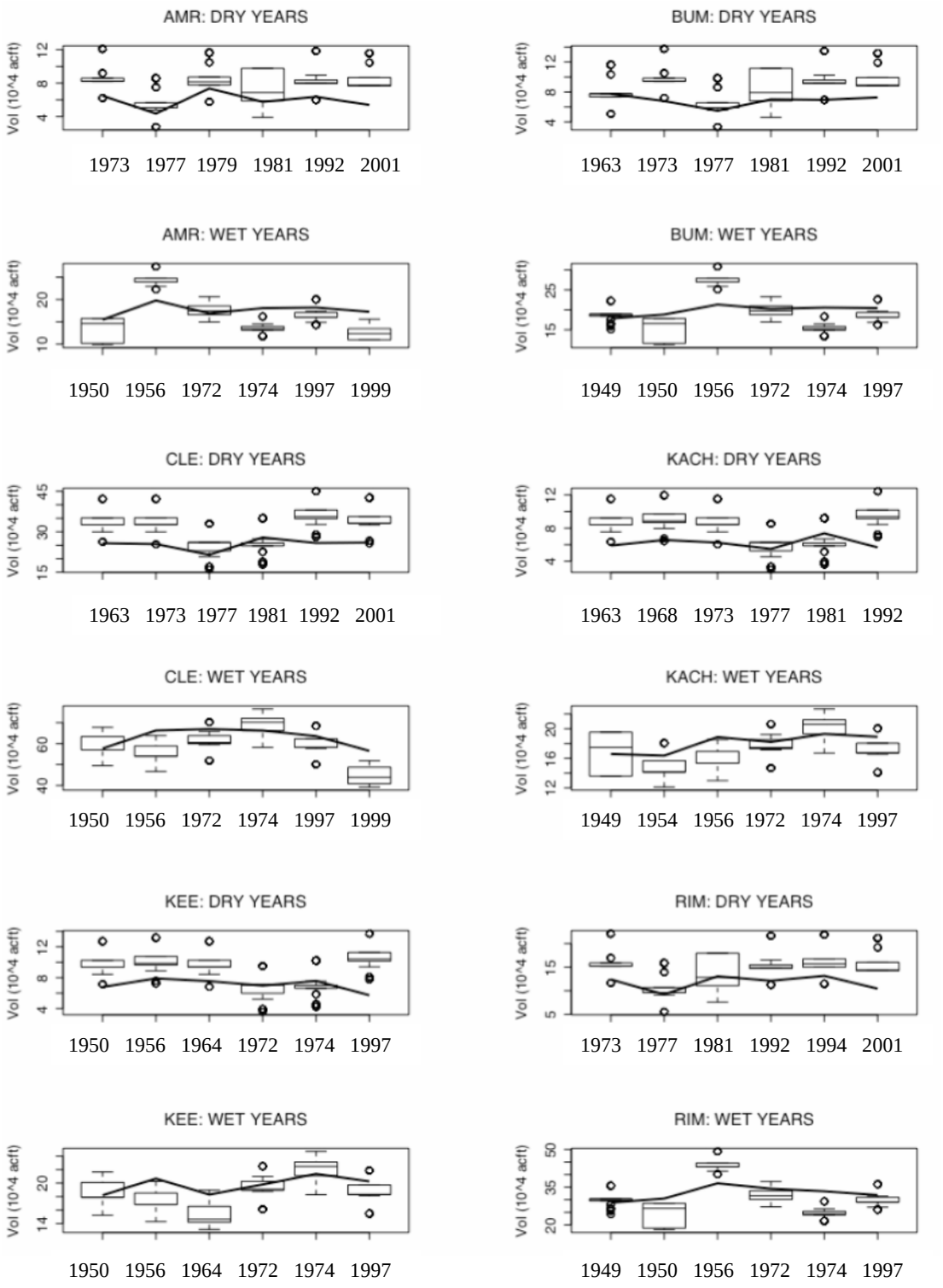


Figure 3.12: Ensemble forecasts (box plots) using Feb 1st SWE only, observed is black line.

3.4 Forecast Discussion and Conclusion

From the previous discussion of the skill of the ensemble forecasts in predicting the historical streamflow record, it is apparent that the modified K-NN approach is capable of providing water managers with reliable forecasts. However, some improvement to the model could be made. There are some drawbacks to the K-NN approach that were used in this study. The modified K-NN algorithm is capable of producing negative residuals, which when resampled and added to the mean forecast, can produce negative streamflow values (Grantz, 2003). This is generally only a problem if the lowest historical streamflow values are close to zero. This problem can be overcome by predicting seasonal streamflow volumes which are unlikely to be close to zero. None of the six sites in the Yakima basin suffered from this potential problem.

The method of selecting the number of residuals to retain was not very robust in reflecting the underlying data distribution for any given set of predictors. In this model, K was found to be within the range of the square of the number of years (54) in the dataset. The choice of K had to be altered by hand and multiple ensemble runs made with different numbers of nearest neighbors before the skill of the forecast could be tested. This method is time consuming and prone to error as the modeler must subjectively decide from the RPSS and Likelihood score which combination provided the greatest skill. The RPSS and Likelihood provide a similar measure of an ensemble's skill, but are sometimes different. A K that is capable of adapting to the underlying data distribution might improve the skill of the forecast and the computational time. Baoli *et al. (*2003) propose selecting K by taking the proportion between the similarity function from neighbors belonging to a certain class to the total sum of the similarity function of all the selected neighbors.

The second problem with the model lies within the choice of predictors and predictands. When principal component analysis is performed, some information about the original data set is invariably lost. While the first PC of the upper sites captured 98% of the

data variance, the first PC of the lower sites captured approximately 97% of the variance. The second PC of the lower portion constituted nearly 3% of the variance, implying that a small but significant amount of the information was lost in the choosing the principal components. The second PC was not included in the ensembles for the lower region because no physical relationships were noted with climate patterns or with the SWE data. Adding the second PC did not improve the forecast skill any. PCA was performed in the interest of finding robust basin wide climate relationships, and worked remarkably well to this end. However, were the forecasts to be regenerated, finding the correlation between the individual sites' seasonal streamflow volume and climate might provide better forecasts. Directly producing ensembles with the streamflow volumes might also lead to increased skill as unique information from each site would not be lost.

Finally, the basin is located in the lower elevations of the Cascades. The area is highly sensitive to temperature changes. As the sites are not at the same elevation, each site might have a unique temperature response. The sites in the lower portion of the basin are at higher elevations than the upper three reservoirs and might be less sensitive to temperature changes. Monthly average temperature for each site might be included in the predictor set used to generate the individual ensemble forecasts for each site. The information from years that experienced sudden late winter warming spells could be captured in the model. The snowcourse/ snotel sites are also unevenly distributed and not all at the same elevation. Some of these sites are close to the reservoirs, but others are not. This might obscure the true SWE distribution contributing directly to runoff at each of our forecast sites.

The stochastic ensemble model based on the modified K-NN algorithm presents water manager with a reliable forecast that makes no assumptions about the underlying data distribution. The model is simple to employ and can be easily modified to accept any predictor set, not just the sets used here. The model can also be modified to generated ensembles at any time step necessary. Prairie (2002) implemented the model for a monthly

time step, while Grantz (2003) changed the time step to yearly. Theoretically the model could be adapted for a daily time step. The model could be used in conjunction with current USBR forecasts to enhance the overall skill of the forecast set.

Chapter 4

Drought Years

Water managers are interested in skillful streamflow forecasts with increased lead-time. Equally important to water managers is the ability to apply the forecasts in a planning model. The USBR is responsible for managing the water supply for over seventy percent of the rights holders in the Yakima basin (IOP, 2002). The official forecast is released to the public on April 1st, along with the initial proration fraction, if applicable. Intensive testing of the forecast scenarios is performed before the public issuance in order to estimate the basin response to projected streamflow regimes. The various ensemble traces are passed into a hydrological model of the Yakima basin that has been rendered in a hydrological modeling tool such as RiverWare.

Drought years present a particular challenge to the water managers of the Yakima basin. Because the water supply they oversee is limited to storage for one operational water year only, the USBR needs to be able to determine the proration levels such that the minimum instream flows are met and that there is enough water left in storage at the end of the irrigation season to provide flows to keep salmon redds covered. Successive years of drought place a large strain on storage and management. This is further exacerbated because Keechelus Reservoir cannot be filled to capacity due to structural problems and Kachess Reservoir has difficulty filling even in years of normal flow. The storage capacity of Kachess Reservoir exceeds its inflow capacity in all years except those with above normal flow (IOP, 2002). The 2001 drought year underscored the importance of an accurate forecast. On April 1st, the total water supply available (TWSA) estimated by the 100% of normal subsequent conditions issued for 2001 was 1.68 million acre-feet (USBR, 2001). In a non-drought year, the TWSA can balance the average consumptive use of about 2.5 million acre-feet.

4.1 RiverWare Modeling

RiverWare is a generalized hydrological modeling tool. It was created under the auspicious of providing the Tennessee Valley Authority and the USBR with the ability to model complex policies and hydrological scenarios in the various river basins and systems these agencies operate. It has a user friendly, object oriented graphical interface that allows decision makers and stakeholders to visualize the hydrological processes occurring in a basin, as well as the results of different policy scenarios. The objects represent physical points in a basin such as reservoirs, river reaches, and diversions. RiverWare allows managers the flexibility of developing a model that suits the particular operating policy and objectives of an individual basin (Zagona *et al.*, 2001). A basin's policy is not hardwired into the code; instead it is implemented by the modeler using one of the software's operating modes. RiverWare has three capabilities: simulation based, rule-based simulation, and optimization modes. The model solves the system according to the constraints and processes specified for each object. The user is free to decide which mode is best suited to the peculiarities of a basin.

4.1.1 The Yakima Model

The Yakima model is a complex model that contains nearly all the operational policies governing the basin. Movement of water through the reservoirs and the irrgation canals is simulated by the model, which employs graphical representation of the reservoirs and water users as objects. The model, implemented in RiverWare, is a rule-based simulation with the rules representing the operational policy. The rules govern the mass balance of the system and determine how the model solves. A schematic of the Yakima basin as implemented in RiverWare is presented below courtesy of the USBR (Roger Sonnichsen, Warren Sharp, personal communication). The Yakima model was designed such that it is visually close to the actual geographic distribution of the reservoirs, irrigation districts, and the tributaries. The model has several key points, the Parker gage, highlighted in the grey

box, and the reservoirs. The reservoirs (highlighted by red circles) are located on the outermost left edge of the model. Major tributaries, such as the Teanaway, Naches, and the American River also originate along the western rim.

Many of the irrigation districts and individual water stakeholders have complex water accounts. Some supplement their irrigation with groundwater withdrawals and many have a mix of proratable and nonproratable rights. The timing of each diversion, conveyance losses, and consumptive use is different for each user and can vary with the streamflow conditions. This in turn affects the lag time at which return flows trickle back to the Yakima River. The model attempts to account for all of these factors by implementing them through rules and data objects that store the water user's information.

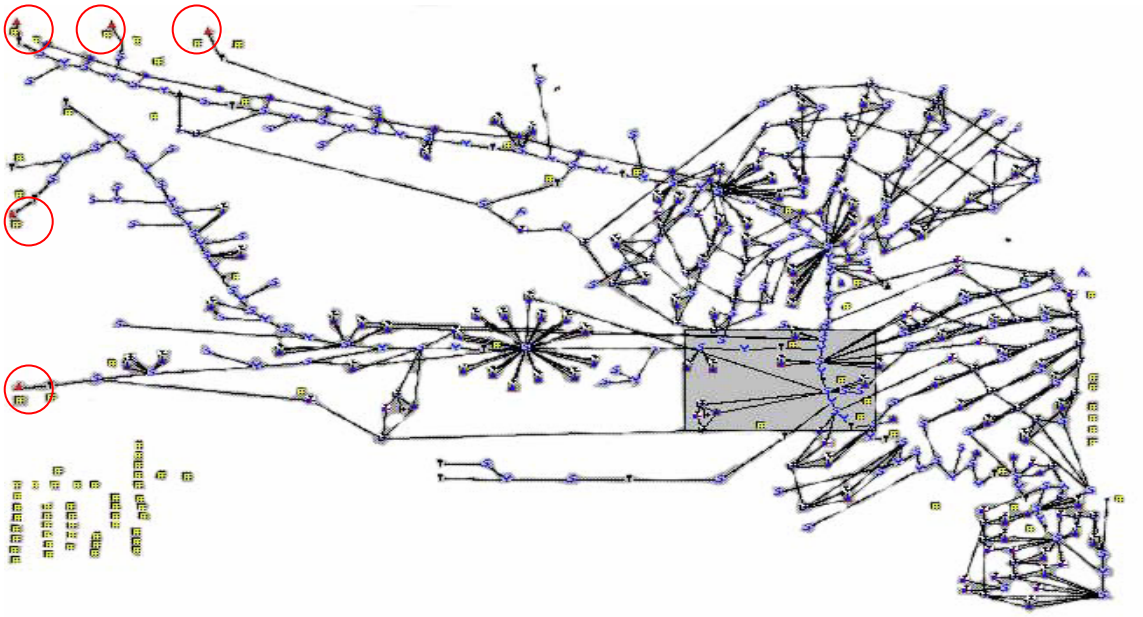


Figure 4.1: Screenshot of the USBR Yakima Model. Reservoirs are circles in red. The rest of the model is devoted primarily to aggregate water users.

The model runs on a daily time step, generally for one year at a time. From Nov. 1st to March 31st of the next year, the model uses historical data. This historical data includes such things as the previous day's reservoir storage and the current day's outflow. On April

1st, the model operates in forecast mode. From April 1st to an average date of June 15th, the reservoirs are being filled during the spring runoff. Irrigation needs are generally met from the runoff during this time. In years of shortage when the reservoirs are having difficulty filling, proration will be implemented during storage control. Storage control marks the period when reservoir filling has generally peaked and demands must be met by reservoir releases. The inflows to the reservoir, Parker gage, and several of the tributaries are input from the USBR's forecasted MMS traces (generated using the deterministic model, not the linear regression model). The model is then allowed to run through the end of the water year (October 31st) to determine the basin response to the projected streamflows.

The Parker gage is the central deciding point of the model. The flow amount (TWSA) that is estimated at this point determines the proration level and the minimum instream target flow for the fish, as outlined in the first chapter. The target flows, proration levels, and date of proration initiation are based on information from the Parker gage and set by rules. Return flows are also estimated at Parker using a RiverWare simulation. Most of the major irrigation districts are located upstream of the Parker gage.

Flow scenarios are closely watched when the model reaches August. August operations establish the demands, constraints and operational criteria for the next season (IOP, 2002). The reservoirs are drawn down in order to meet the last of the irrigation needs. The model also dictates the gradual flip-flop of withdrawal between the upper three reservoirs to the lower two over late August through September 10th. After September 10th the remaining irrigation demands are covered primarily through releases from Bumping and Rimrock reservoirs. This is done to provide enough water to cover salmon migration and spawning and keep the salmon from traveling too far upstream to reaches that have the potential to be dewatered during the winter. A mini flip-flop is sometimes implemented between Cle Elum and Kachess reservoirs.

4.1.2 Disaggregating Ensemble Forecasts

As stated, the Yakima model runs on a daily time step. Therefore, any streamflow values passed into the model must be disaggregated to fit the model step size. The generated ensembles are predictions of the spring runoff seasonal volume at each site. Disaggregating the volumes into daily flow values is not particularly difficult if the forecast is for a historic year. To find the proportion of the seasonal runoff that each day contributes, it is merely the matter of dividing that day's streamflow by the seasonal total. This is repeated for each day of the April-July period in order to calculate daily flow percentages at each site. The timing of the runoff is thus captured quite accurately. The daily proportion vectors for each site are then multiplied by the corresponding ensemble forecast values to recapture the timing of the distribution of the spring runoff.

However, divining the disaggregation proportions for a future runoff season presents a unique problem, especially if one is attempting to disaggregate to a daily time scale. The timing of runoff is unknown and dependent on numerous factors such as temperature, subsequent precipitation, or snowpack density. It is possible to break the forecast into monthly time steps by calculating how the long-term mean proportion of each month contributes to the seasonal runoff. The long term mean of the daily data, however, provides too much temporal smoothing and erases the information about melt due to temperature and other factors. One proposed solution would be to use the nearest neighbors from the modified K-NN algorithm and use their daily percentages. The model could also be perturbed slightly to produce disaggregation proportions not seen in the historical record. Years that share similar climate features would seem more likely to have similar spring runoff patterns. This is due to the fact that temperature, wind, and cloud cover are strongly influenced by large-scale climate behavior. Snowpack densities might also be similar among years sharing similar climate characteristics. Differences between nearest neighbor years are due to the local vagrancies of the basin that affect daily weather and to land use changes over the historical record. Thus, the modeler could select several nearest neighbor years and

simulated years from which to disaggregate the ensemble forecasts. This approach allows testing of several runoff scenarios in the RiverWare model.

4.2 Comparison of the Ensemble Forecasts with USBR forecasts

The focus of this chapter is on drought years. 2001 represents the first year in which the USBR was faced with meeting competing water interests (minimum instream flows and agricultural diversions) in a water shortage. No significant droughts had occurred in the basin since 1994 when the Title XII minimum instream flows were initiated. The average consumptive use for irrigation and regulation flows in a normal year is 2.5 million acre-feet (USBR). This demand is met through a combination of reservoir releases and instream flows (return flows + existing flows). In 2001, the TWSA could not balance demand and significant prorationing occurred. Title XII target flows at Prosser and Sunnyside (Parker) dams were 300 cfs per day for 2001.

We are interested in comparing our ensemble forecasts to the TWSA forecasts issued by the USBR at each reservoir. Our ensemble set used in the comparison was generated using the December through February climate predictors with March 1st SWE. This forecast had better skill in predicting the actual 2001 runoff than the ensembles created using April 1st SWE information. During drought years, the PNA pattern is anomalously strong as early as the November through January time frame. These differences might also be caused by degradation of the snowpack by April 1st, leading to faulty forecasting. Another reason might be that the climate predictors carry more weight in drought years than the SWE and including April 1st SWE degrades the quality of the forecasts. The USBR's median forecasts are compared with ours. A comparison between the American River streamflow ensembles is presented here with the observed flow, but no USBR forecasts. This site is not included in their TWSA forecasts, although it is used in the RiverWare model.

The plots at the end of this section display the probability distribution function (PDF) of our ensembles for the five reservoirs with the historically observed value (solid black line) and with the USBR's median forecast (dotted line). A probability distribution function of the ensembles displays the probability that a forecast member lies between two streamflow volume points. A PDF can be thought of as a continuous, smoothed histogram of the probability of the data distribution. The PDFs are not purely normal distributions and reflect a large amount of skew in the ensemble forecasts. Encouragingly, the median ensembles are skewed toward the lower streamflow volumes.

The median USBR forecast outperformed our median DJFM forecasts for 2001. While their median forecast is short of the observed by nearly 48,000 acre-feet, our median forecast overestimates by roughly 40,000 acre-feet. At Kachess and Keechelus, our median ensemble members under-predicted the observed flow. However, at Cle Elum, Rimrock, and Bumping, the PDF's of our forecasts are shifted largely away from the observed streamflow values. The lowest 15th percentile members of our ensembles capture the historical flows at Cle Elum and Bumping, but still over-shoot Rimrock Reservoir's. Our 15th percentile forecasts under-predict the observed by approximately 3,400 acre-feet. This result is quite good. Our 75th percentile ensemble members generate 140,000 acre-feet more than the observed. The following table displays the flow volume of all five reservoirs for our ensemble forecasts at the 15th, median, and 75th percentiles. It also shows the USBR's official forecasts which are generated at 50%, 100%, and 150% of subsequent normal conditions (provided by Christopher Lynch, personal communication). All values are in acre-feet.

Our forecasts generally overpredicted the runoff volume that would have been generated at each of the reservoirs. Based upon this result, a TWSA forecast using our predicted runoff scenarios would have set the proration levels slightly higher than what levels should have been to meet targets. The official proration level set by the April 1st TWSA was

29% (Christopher Lynch, personal communication) and began on May 1st. This proration level was later revised at the beginning of June to 30%. Additional precipitation in May and June helped to ease the proration levels slightly so that by August, the proration level was 37%. Had our median ensemble forecasts been used as an alternative to the forecasts at the reservoirs, the proration levels would have been slightly higher at the initial April issuing. We estimate that the proration levels using our ensembles would have ranged from about 27% to 35%, depending on the ensemble scenario chosen. Our forecasts would have provided water managers a wider range of scenarios to test in RiverWare.

	Kachess	**Keechelus**	**Cle Elum**	**Bumping**	**Rimrock**	**Total**
Ensembles						
15th %	74151	83024	295592[t]	80827	130949	664542
Median	80574	89925	315362	84951	137494	708305
75th %	93711	104041	355797	97377	157214	808139
USBR						
50%	63100	71400	233400	62400	85900	516200
100%	78900	87800	269800	74600	109100	620100
150%	94800	104000	304700	86900	132300	722700
Observed						
	79557	**97321**	**259981**	**72700**	**104528**	**667985**

Table 4.1: 2001 DJFM forecasts v. observed: values in ac-ft. At Cle Elum, the lowest value in the table is in the 10th percentile.

However, as noted in the previous chapter, the skill of predicting a drought year comes from using earlier climate and SWE data. Our forecasts using the November through January climate information with Feb. 1st SWE perform better than those using the DJFM information. The PDFs of our NDJF ensembles are also located at the end of this section and precede the PDFs of the DJFM ensembles. The following table displays the flow volume at all five reservoirs using the NDJF predictor set for the scenarios described in the previous table. The USBR forecasts are not included in the table.

	Kachess	Keechelus	Cle Elum	Bumping	Rimrock	Total
Ensembles						
25th %	61843	69799	257708	89068[t]	144029[t]	622447
Median	68390	76834	277862	89888	145330	658305
75th %	94707	105111	358865	96505	155830	811018
Observed						
	79557	**97321**	**259981**	**72700**	**104528**	**667985**

Table 4.2: 2001 NDJF forecasts v. observed: values in ac-ft. At Bumping and Rimrock, the lowest values presented here are in the 5th percentile

Overall, the site total 15th percentile and median NDJF ensemble volumes are closer to the observed total. The upper reservoirs perform better than the lower reservoirs whose predictions for this set of climate information perform worse than those using DJFM information. The under-predictions of the upper reservoirs serve to offset the over-predictions made at the lower two reservoirs. The 15th percentile forecast falls short of the observed total volume by approximately 46,000 acre-feet. The median NDJF ensemble total is roughly 9,700 acre-feet less than the observed total. Thus, a forecast issued at the beginning of February provides a reliable estimate of the severity of the drought and the range of possible runoff scenarios.

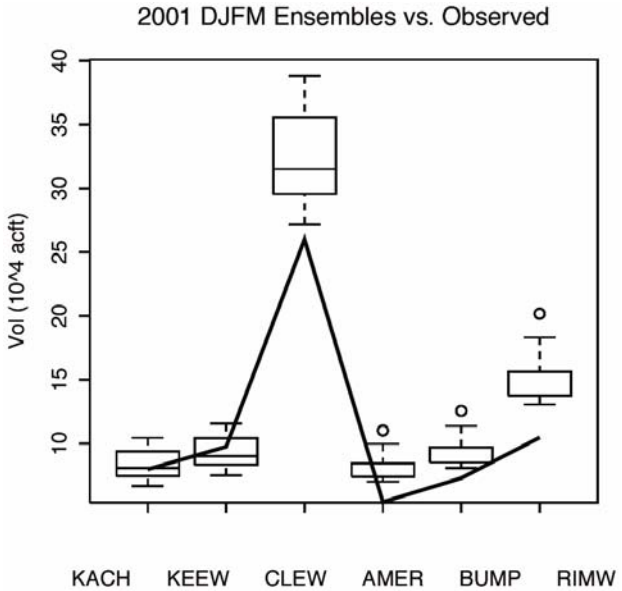


Figure 4.2: 2001 DJFM forecasts (box plots) v. observed (solid black line). The sites are listed on the x-axis.

These two figures (Figures 4.2 and 4.3) show the range of our ensemble forecasts plotted with the observed streamflow volumes at each site. The first figure corresponds to forecasts generated using the DJFM predictor set. The second figure uses the NDJF predictor set. The NDJFM forecasts are better at capturing the observed streamflow at the upper three reservoirs than the DJFM forecasts. However, the NDJF forecasts perform worse at Bumping Reservoir and Rimrock Reservoir than the DJFM forecasts.

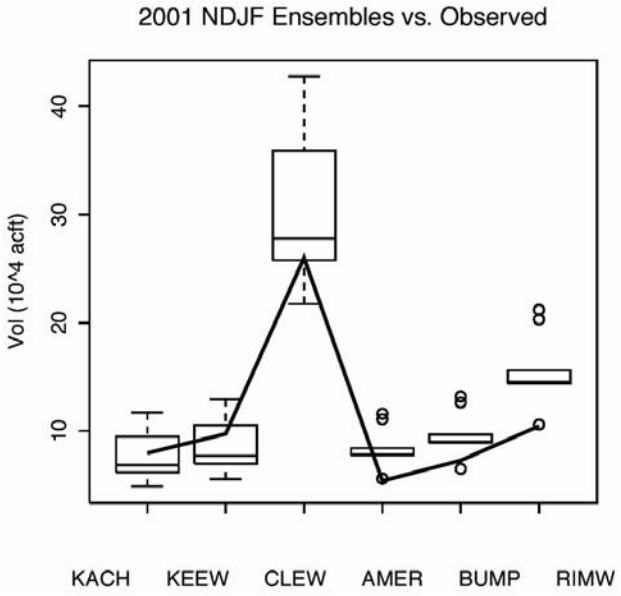


Figure 4.3: 2001 NDJF forecasts (box plots) v. observed (solid black line). The sites are listed on the x-axis.

While our median DJFM forecasts did not perform as well as the USBR's 100% normal subsequent conditions April 1st forecasts, there are some advantages to using our ensembles or employing the technique that we used. As noted, the USBR generates three forecasts based on anticipated precipitation levels of 50%, 100%, and 150% of normal subsequent conditions. The precipitation ranges are arbitrarily decided by USBR staff and anticipate events that have not yet occurred. These ranges are quite large and limit the runoff forecast's range. Second, the USBR model assumes a multiple linear regression relationship over the entire data set. This may or may not accurately reflect true data distribution. Our

ensembles make no assumptions about the historical data's distribution and do not rely on future events to generate streamflow. The advantage to ensembles is that a wider range of streamflows can be generated and the exceedence probabilities of each can be calculated. Yet, this range is still within the statistical distribution of the historical record. The exceedence probability is the probability that a streamflow value equal to or greater than the streamflow volume selected from the ensemble could occur. This would provide water managers with a more robust, statistically based estimate of the probability that a certain runoff scenario could occur.

The final advantage to our forecasts lies in the neighbors selected in the K-NN algorithm and their applicability to disaggregating the streamflows. Our ensembles were generated for 1949-2002 only because the American River data set was truncated at 2002. If American River were dropped from the predictand sites, the ensembles could have been generated through this current year (2004). As all the years in the ensemble model are historic, disaggregation of the volumes was performed using the technique described in section 4.1.2. We then attempted to pass our disaggregated streamflows into the RiverWare model to see how they would have performed in comparison to the USBR's MMS trace forecasts. We replaced the traces at the six forecast sites with our disaggregated ensembles and left the traces in place at the other 72 model points with forecasted flows. Unfortunately, our inputs caused the model to stop working on April 1st, the first time step of the forecast period.

The model did not function largely because of the difference in the timing and distribution of runoff between our disaggregated values and the MMS trace values at the reservoirs and American River. The snapshot of the 2001 model with which we were working was highly pre-determined. For example, the proration levels and previous reservoir storage levels were already set. There were too many values on the reservoirs, water users, and target points that would have had to be reset before our forecasts would work with the

model. A comparison between our forecasts and the MMS traces for two reservoirs, Keechelus and Cle Elum, for the first five time steps of the model forecast period is shown below. All values are in cfs. The differences between our daily flow values and the MMS trace values is an order of magnitude. Our values, being based on the historical record's disaggregation information, are much closer to the observed than the MMS traces. The model then breaks because the MMS unregulated flows at the other 72 points in the model were not summarily adjusted. The proration levels would also need to be adjusted slightly to reflect the new information. Finally, the proration date was instantiated as May 7th in the model. Our proration date would have occurred earlier, possibly toward the middle of April.

	KACHESS		CLE ELUM	
Date	**MMS Trace**	**Our Value**	**MMS Trace**	**Our Value**
April 1, 2001	982.66	350.31	1743.2	784.67
April 2, 2001	880.52	280.53	1444.21	784.48
April 3, 2001	601.8	282.5	1089.12	693.49
April 4, 2001	429.7	178.21	880.93	644.79
April 5, 2001	384.64	278.72	847.72	634.57

Table 4.3: 2001 MMS traces v. our values at two reservoirs: all values in cfs.

The Yakima model is such a complex model that making changes to runs of previous years is difficult. The model behavior and rule set are not readily transparent to stakeholders or others seeking to input different scenarios for previous year runs or the current year. We attempted at first to input our values into a snapshot of the 2001 model run. When this approach was unsuccessful, we attempted to back up a 20-year model that had been initialized for the 2002-2003 water year. The model start time was backed up to Oct. 31, 2001 and it was attempted to initialize this model with values from the 2001 model. This approach also failed, because there were too many points in the model that required initialization values to easily input.

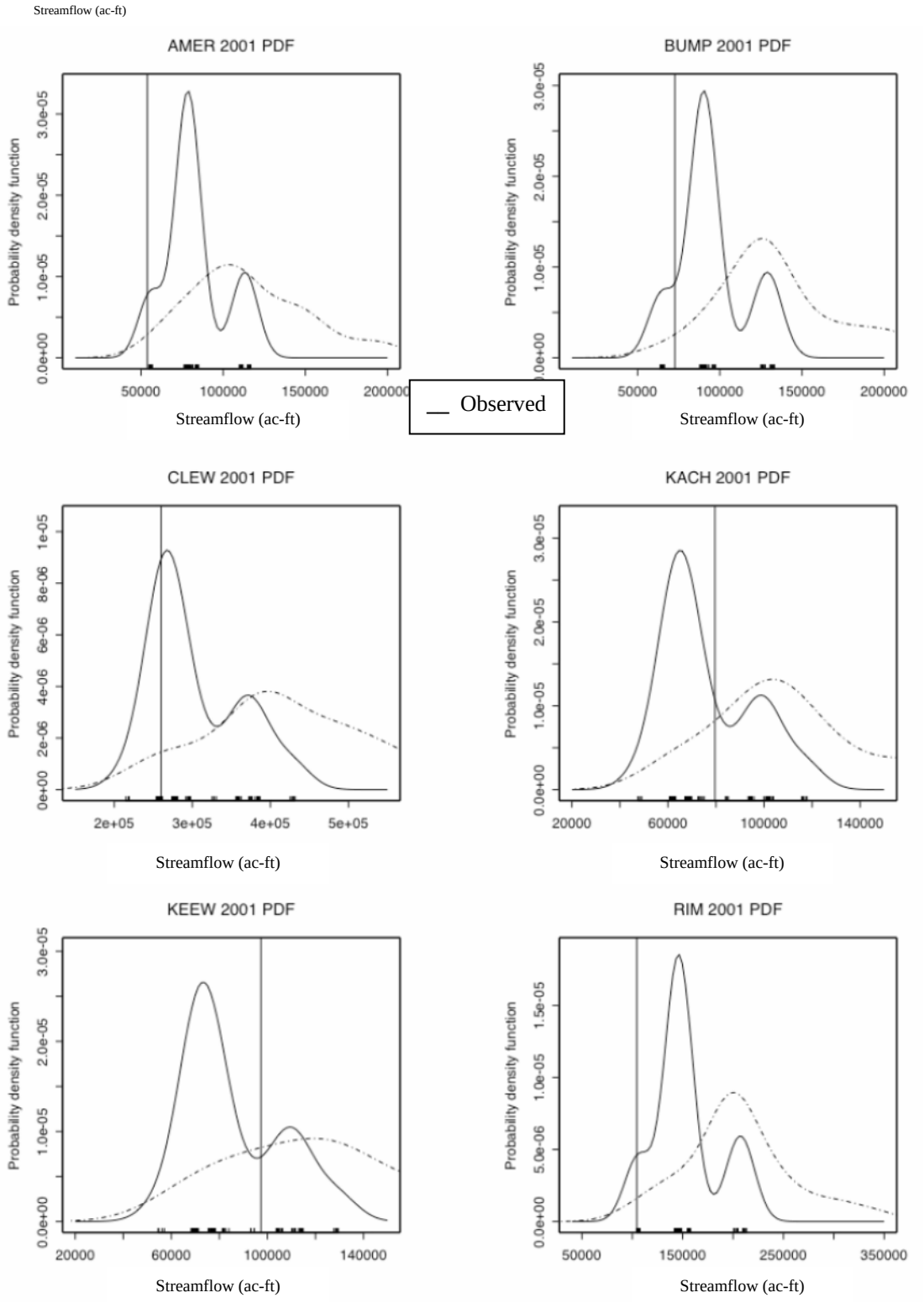


Figure 4.4: 2001 NDJF ensemble PDFs with observed value (solid vertical line). The dotted line is the PDF of the forecast generated using climatology.

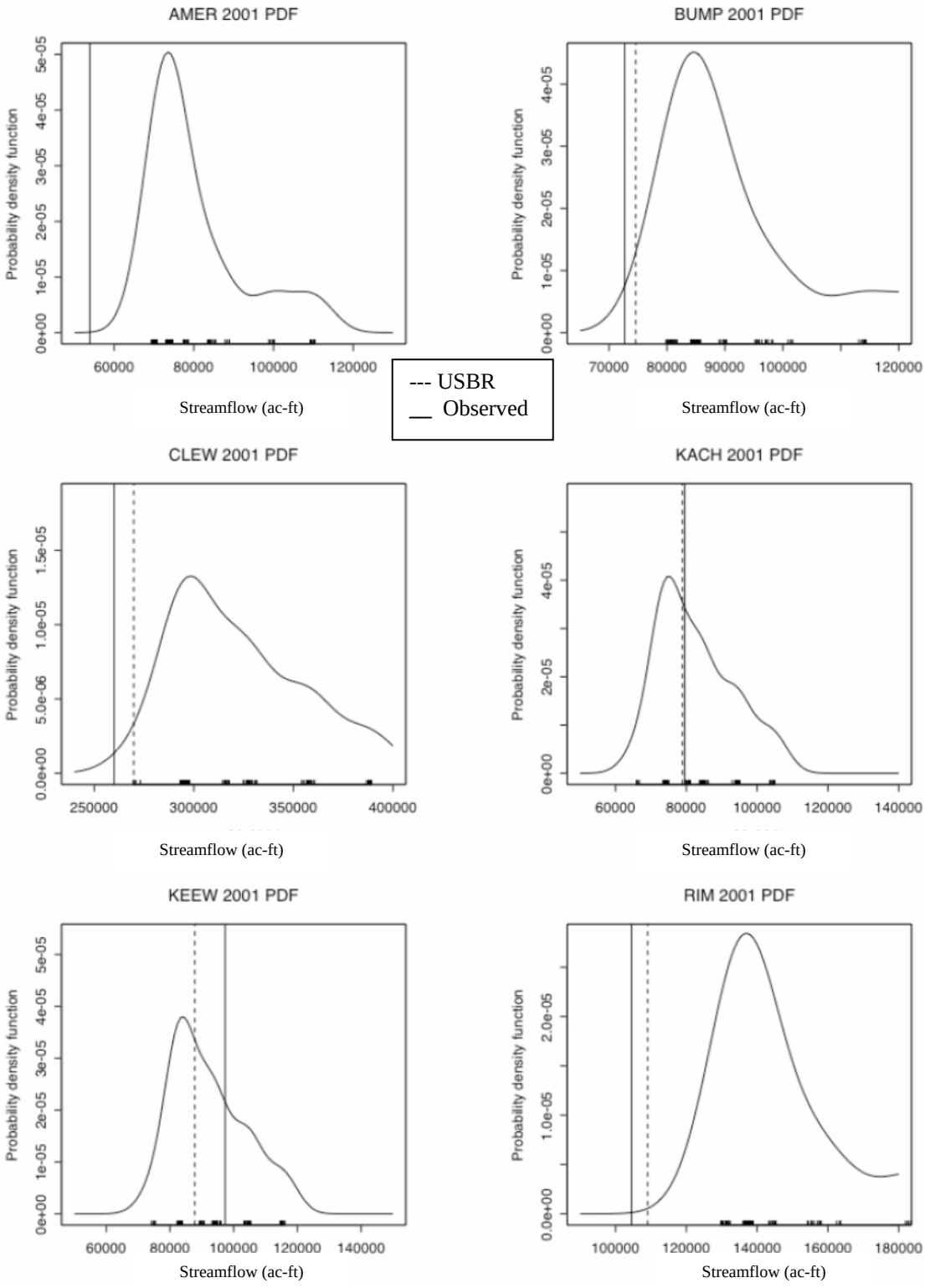


Figure 4.5: 2001 DJFM ensemble PDFs with observed and 100% USBR forecast

4.2.2 The 1977 Drought Year

The 1977 drought represents the most severe drought in the Yakima basin for the period of 1948-2002. It is generally agreed that a regime shift of the PDO occurred in the winter of 1976 to 1977 from a cool phase to a warm phase (Mantua *et al.*, 1997 and Zhang *et al.*, 1996). This led to significantly reduced winter precipitation in the PNW and subsequent lower runoff in the entire basin of the Columbia River. We were interested in our ensembles' performance in this year because of the severity of the drought. The 1977 drought has a recurrence interval of 55 years. This means that the average interval between events equaling or being below the 1977 magnitude is 55 years. However, the historical record cannot be taken as indicative of the full range of streamflow variability in the basin. It is possible that a more extreme drought event could occur or even has occurred in years not included in our data set. A data set comprised of 54 years is not a significantly long record.

Congress had not yet implemented the Title XII target flows. Yet the severity of the drought prompted prorationing levels of 66% (Watershed Assessment, 2003). Had the Title XII target flows been in place in 1977, the proration levels would have been lower than those implemented in 2001. This potential for very low proration levels underscores the importance of skillful forecasts, especially in drought years. Furthermore, increased lead-time will allow stakeholders to decide whether or not to plant or how to manage their business interests. Water managers will be better prepared to ensure that reservoirs begin storing water earlier in the season to be able to meet instream flow needs as well as those of the stakeholders with nonproratable entitlements.

Presented on the following page are the PDFs of our ensembles that were generated using the December through February climate predictors and March 1st SWE for the 1977 drought. While the April 1st forecasts captured the observed flows quite well, we present the earlier forecasts. The forecasts using the November to January climate predictors also performed well and are presented before the DJFM PDFs.

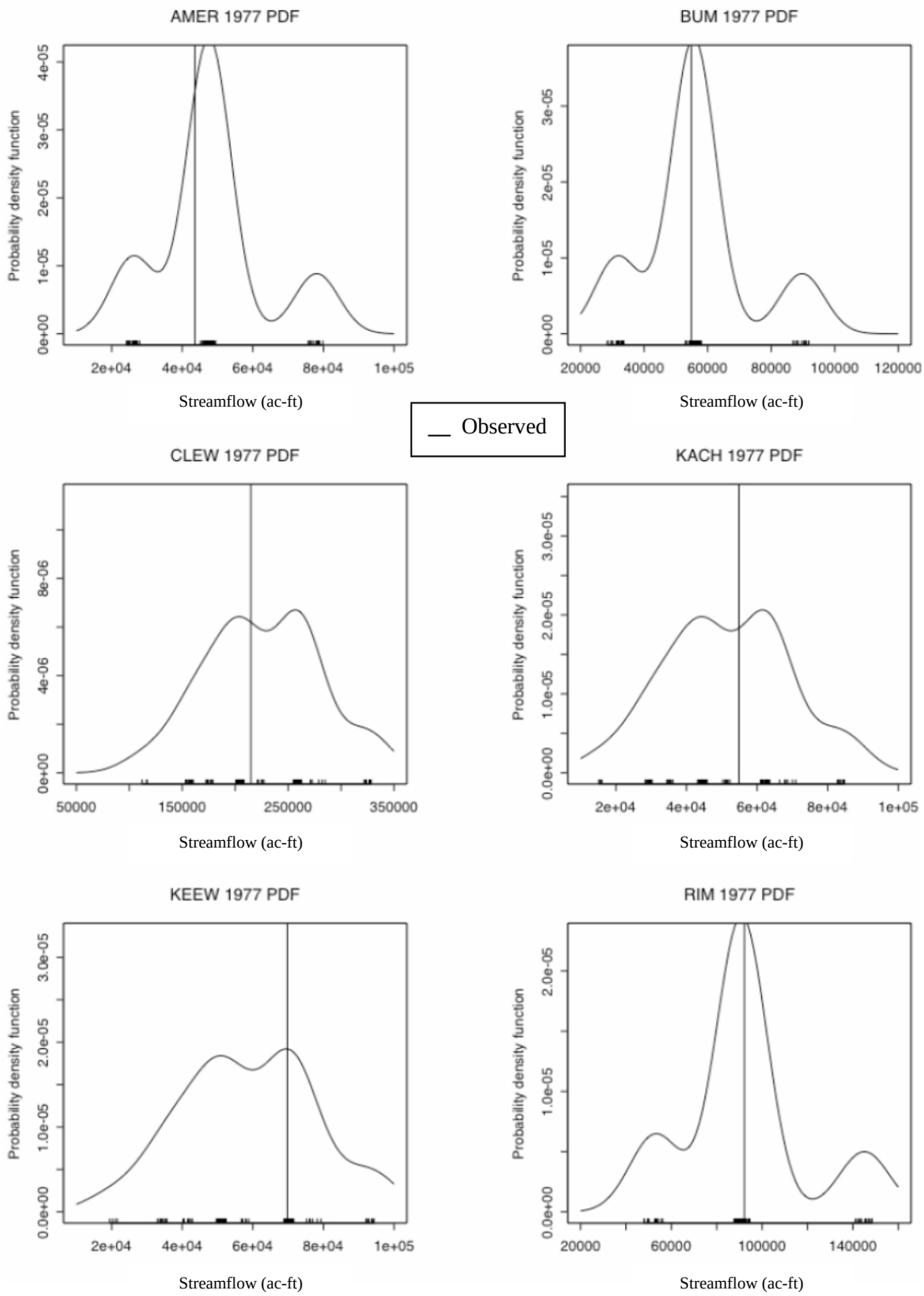


Figure 4.6: 1977 NDJF ensemble PDFs with observed (solid vertical line)

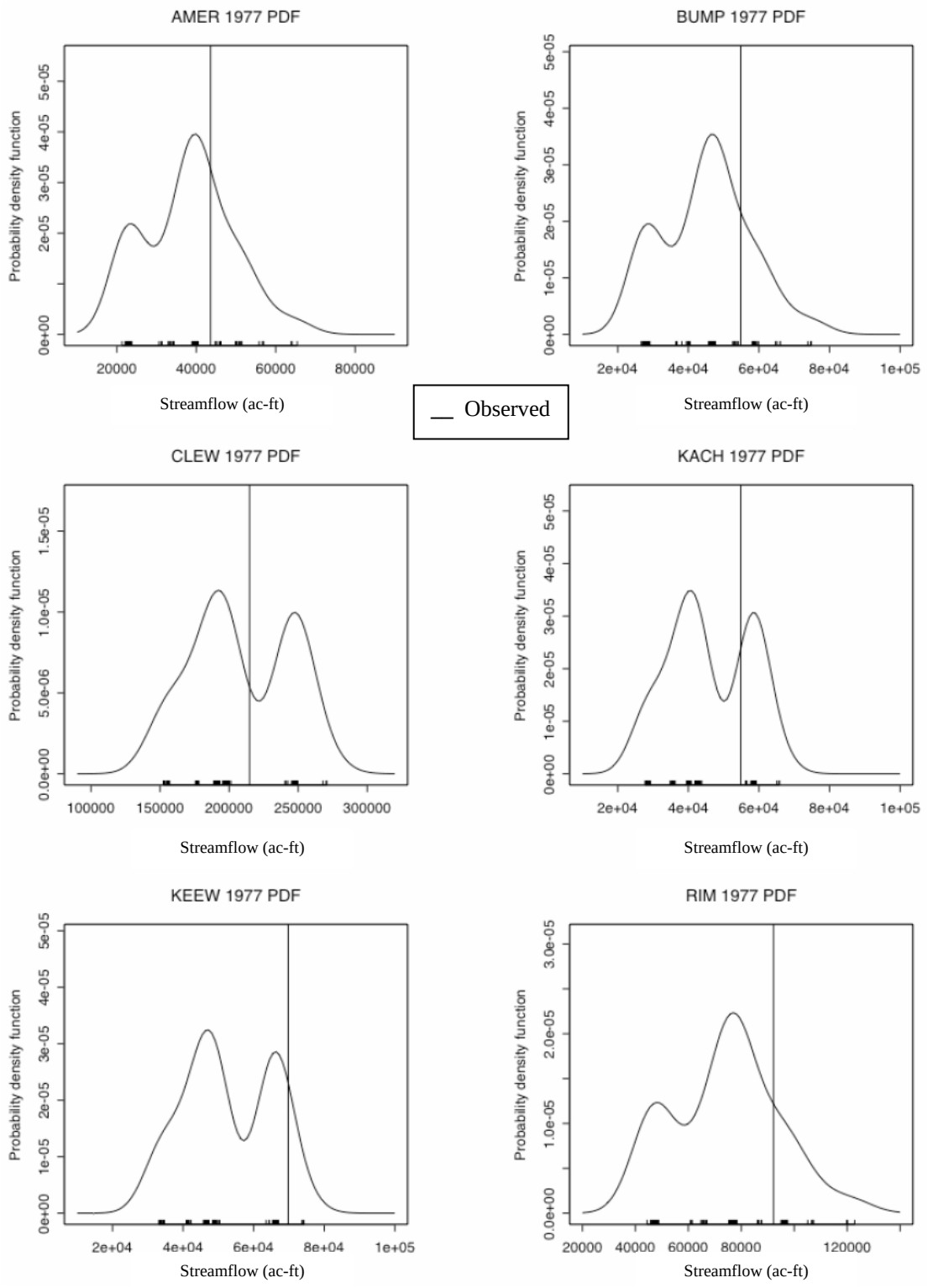


Figure 4.7: 1977 DJFM ensemble PDFs with observed (solid vertical line)

The forecasts generated using the NDJF set of predictors outperform those using the DJFM information, although that set did quite well. The median ensemble members for the three lower sites almost capture the observed value at each site. The three upper reservoirs also perform well, even though there is no distinct median value. The following table compares the performance of the two sets of forecasts to the observed. It is actually the 75th% forecasts that are closest to the observed total. The NDJF 75th% total is greater than the observed value by 11,000 acre-feet. The DJFM 75th% total under-predicts the actual value by 17,000 acre-feet. The range of values generated by both sets of ensembles at all forecast sites covers the historical volumes recorded at those sites.

	Kachess	**Keechelus**	**Cle Elum**	**Bumping**	**Rimrock**	**Total**
NDJF						
25th %	44445	51106	204160	54110	88550	442373
Median	50951	58096	224183	55529	90803	479562
75th %	62212	70196	258846	56849	92896	541000
DJFM						
25th %	40183	46526	191040	37268	61822	376840
Median	42396	48904	197851	46773	76907	412832
75th %	58499	66206	247416	53419	87454	512995
Observed						
	54864	**69784**	**214784**	**54861**	**92189**	**530067**

Table 4.4: 1977 forecasts v. observed: values in ac-ft

The following set of figures displays our forecasts against the observed streamflow volumes at each site. The median and 25th% DJFM forecasts underpredict the 1977 streamflow volumes. There are several possible reasons for the underprediction. The snowpack could have begun degrading by March or the pressure difference between the Canadian ridge and the Aleutian low could have intensified further between the NDJ and DJF periods. This pressure intensification could have biased the forecasts, making the streamflow predictions lower than what conditions actually produced. The 1977 NDJF forecasts are clustered more tightly about the observed values at four of the reservoirs. The observed at

Cle Elum Reservoir falls between the median and 25th% NDJF predictions.

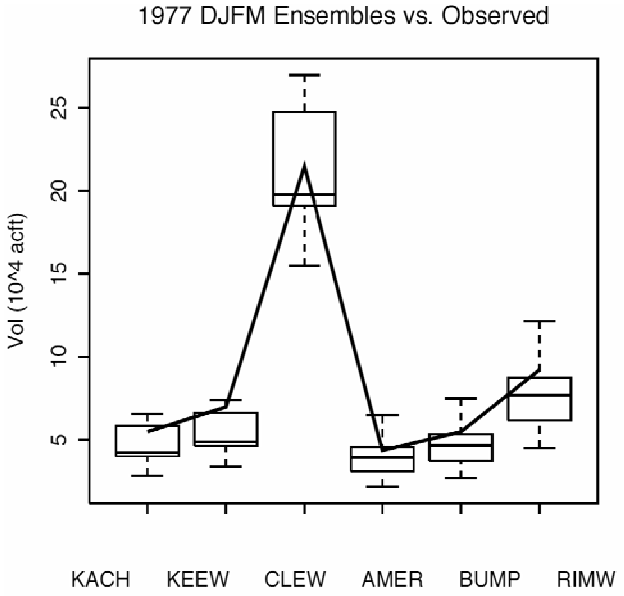


Figure 4.8: 1977 DJFM forecasts (box plots) v. observed (solid black line). The sites are listed on the x-axis

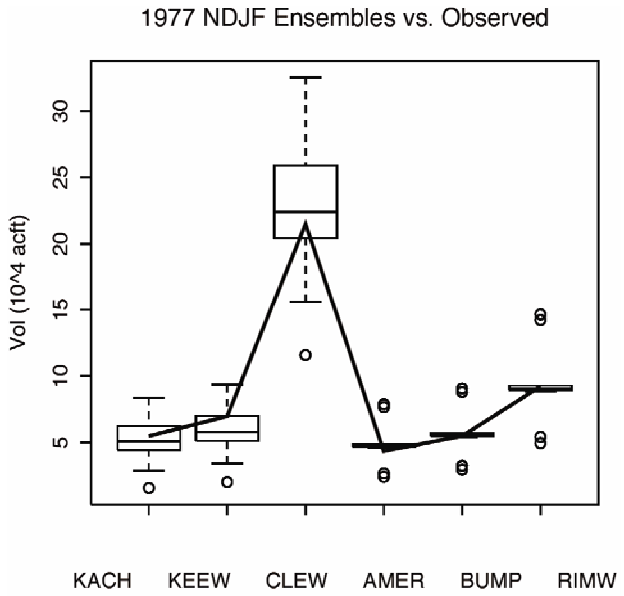


Figure 4.9: 1977 NDJF forecasts (box plots) v. observed (solid black line). The sites are listed on the x-axis

4.3 Discussion and Implications for Stakeholders

One can see that our forecasts perform comparably to the USBR's forecasts and generate a wider range of scenarios that can be tested in the decision support system, RiverWare, if implemented correctly in the Yakima model. Our forecasts perform better at an earlier lead-time using the November through January climate information and February 1st SWE. This would allow tentative issuance of a forecast to interested stakeholders at the beginning of February.

Those stakeholders most vulnerable to serious prorationing levels are implemented are those whose primary economic activity is agriculture. Many of the highest valued crops are grown in irrigation districts with a significant proportion of their water portfolio being proratable (Scott *et al.,* 2004). Three of the irrigation districts, Roza, Kennewick, and Kittitas, have entitlements that are completely proratable (Watershed Assessment, 2003). While the Wapato Irrigation District has nonproratable rights, half of its portfolio is proratable. Roza, in particular, grows primarily high valued crops such as apples, mint, and hops. Thus, these districts, and other proratable users stand to suffer heavy economic losses if a severe drought occurs and drastic prorationing levels are set. As stated earlier, the 1977 drought was the most severe on record, but it cannot necessarily be held as representative of the most extreme dry event possible in the basin. There is no guarantee that the USBR will be able to meet instream target flows and nonproratable entitlements, let alone proratable ones, in a scenario more extreme than that of 1977.

Many of the planting decisions for an upcoming agricultural season begin in January and February of that year (Scott *et al.,* 2004). Our forecasts for drought years had the best ability for predicting severe conditions when the NDJF predictor set was employed. As the PNA pattern is usually strongly developed by November of the proceeding year, the USBR could begin watching the PNA conditions in October or November. This would provide invaluable early insight into the upcoming water supply. If the PNA pattern begins exhibiting

characteristics similar to other notable drought years, the USBR could issue a forecast in the beginning of February using our NDJF information with a high degree of confidence. This would allow stakeholders to implement different plans during the critical decision period at the beginning of the year.

The extended drought of 1992-1994 prompted the State of Washington to pass legislation (Substitute House Bill No. 2537) allowing for temporary water transfers in years of drought. Transfers are facilitated through Boards of Joint Control between irrigation districts (Scott *et al.*, 2004). The boards coordinate the transferring of water between farmers in the districts or between farmers in district. Currently only Roza and Sunnyside irrigation districts have formed a board. Often the transfers are economically driven. Current market prices for individual crops is often a deciding factor as to whether a farmer will pay to transfer water from another farmer or forego planting and lease his or her water entitlement (Voluntary Water Transfers, 2002). Water transfers for municipal or instream needs become financially attractive only if the purchasing agency is willing to pay enough to make not planting attractive.

In late spring of 2001, a drought emergency was declared in the Yakima basin. The transfers were facilitated because the TWSA forecasts are issued in late spring. However, if the TWSA had been tentatively released at the beginning of February, based on our projected runoff scenarios, the water transfers might have begun earlier in the season. More farmers would have been able to decide whether or not it was worth the economic expenditure to plant, limit what crops were planted, or forego planting altogether. We are confident that our earliest (NDJF) forecasts, as well as the DJFM forecasts, can provide reliable estimates of projected runoff for the TWSA in drought years.

Chapter 5

Climate Change, Salmon, and Water Management

One of the greatest unknowns in the future of water management and of salmon recovery in the Yakima basin is the potential impact of climate change. Definitions of climate change vary among scientists and can include the natural variability of the climate system. Others prefer a definition that accepts variations caused solely by anthropogenic activity (IPPC, 2003, United Nations Framework Convention). Regardless of the accepted definition, climate is a dynamic system that does not stay in one state for very long. The rate at which the system changes and the mechanisms of change are not well understood. However, any variation of the climate directly impacts a basin's hydrology and has serious implications for water management. The Pacific Northwest (PNW) seems particularly sensitive to changes in temperature and the teleconnection patterns responsible for the region's precipitation. In this chapter, possible implications of climate change for salmon recovery and water management in the Yakima Basin are discussed.

5.1 PDO and ENSO

Emerging studies indicate that the Pacific Decadal Oscillation (PDO) strongly affects salmon survival at sea and other marine life by altering the ocean food web and conditions (Polovina *et al.,* 1995, and Mantua *et al.,* 1997). Francis et al., (1998) describe three dominant fish production regions in the northeast Pacific that are defined by the Alaskan and Californian Currents. These currents cause upwelling off the coasts of the PNW and downwelling in the Gulf of Alaska. Figure 5.1 shows the production regions of the North Pacific. Primary production by phytoplankton is affected by the degree of upwelling and

vertical mixing that is crucial to the supply of nutrients to coastal surface waters. Higher trophic levels respond to changes in the abundance of phytoplankton. Less upwelling leads to decreased nutrient availability and phytoplankton production.

Anderson (1996) and Polovina (1995) note that the PNW coastal waters were cooler and had enhanced upwelling conditions prior to the PDO regime shift in 1976. When the PDO changed to the warm phase, the Aleutian low (associated with the Pacific North American 'PNA' pattern) intensified and shifted southward. The circulation of the Alaskan Current intensified in response to the change in the surface winds induced by the shifted Aleutian low which had pushed the Subartic Current further north. This induced warmer waters to move up into the coastal PNW upwelling region and reduce the upwelling of nutrients. At the same time, upwelling in the Gulf of Alaska increased (Emery and Hamilton, 1985, as cited in Anderson, 1996).

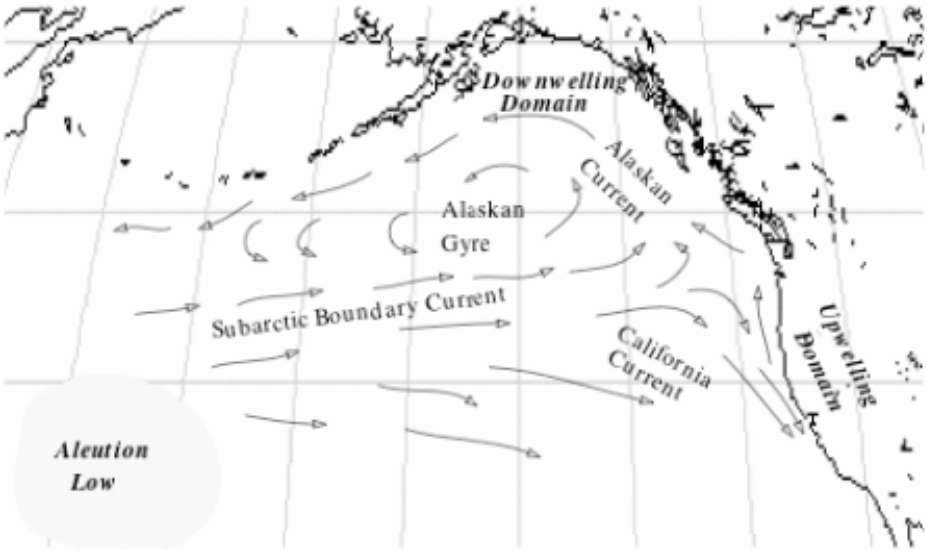


Figure 5.1: Currents and production regimes in Northeast Pacific

Mantua et al. (1997) and Hare et al. (1999) have linked salmon survival at sea to the phases of the PDO. Mantua (1997) compiled the historical catch record of various Alaskan

and PNW salmon and used these values as representations of salmon survival rates at sea. Low numbers of returning salmon are indicative of high mortality rates that year or set of years. The reduction in ocean salmon populations can be attributed to the degree of upwelling in the Gulf of Alaska and the PNW, which is dependent upon the current PDO regime. Nutrient limitation in regions of little upwelling causes a bottom-up reduction across all trophic levels, including salmon. Alaskan salmon catches and PNW catches show an inverse relationship to PDO. During the cool phase of the PDO, upwelling conditions are ideal along the PNW coast and poor in the Gulf of Alaska. The conditions in the Gulf of Alaska and the PNW coast switch when the PDO enters a warm phase.

The species within a riverine-coastal ecosystem, including salmon in the PNW, have evolved and adapted to changing ocean conditions. There is evidence suggesting that the El Nino Southern Oscillation (ENSO) and PDO are not new phenomena, but have existed several hundred years. COAPS (2004) has reconstructed monthly mean SST fields in the tropical Pacific and found ENSO events occurring as far back as 1869. The COAPS index relies on direct observation of SST from 1949 to present and reconstructed events back to 1869. The reconstructed data is compared with historical accounts. PDO events have been directly recorded since 1925, with records extending back to 1900 (Zhang *et al.*, 1997). Furthermore, estimates of salmon populations in Alaska based on sediment deposition from salmon carcasses have shown a cyclical response to climate variations over the past three hundred years (Finney *et al.*, 2000). Salmon catch records dated back to the 1880's in the PNW and Alaska verify this variability in salmon abundance (Francis and Mantua, 1996). However, both records show a marked decline in salmon numbers by the 1920's when heavy commercial fish began to take its toll on the salmon population. The following figure (Figure 5.2) shows the inverse relationship in the response of Alaskan salmon and PNW salmon to PDO over the last century. Thus, the recurrent unfavorable ocean conditions associated with

climate variability, such as the PDO, appear to exacerbate the decline of salmon that are already severely stressed by human activity.

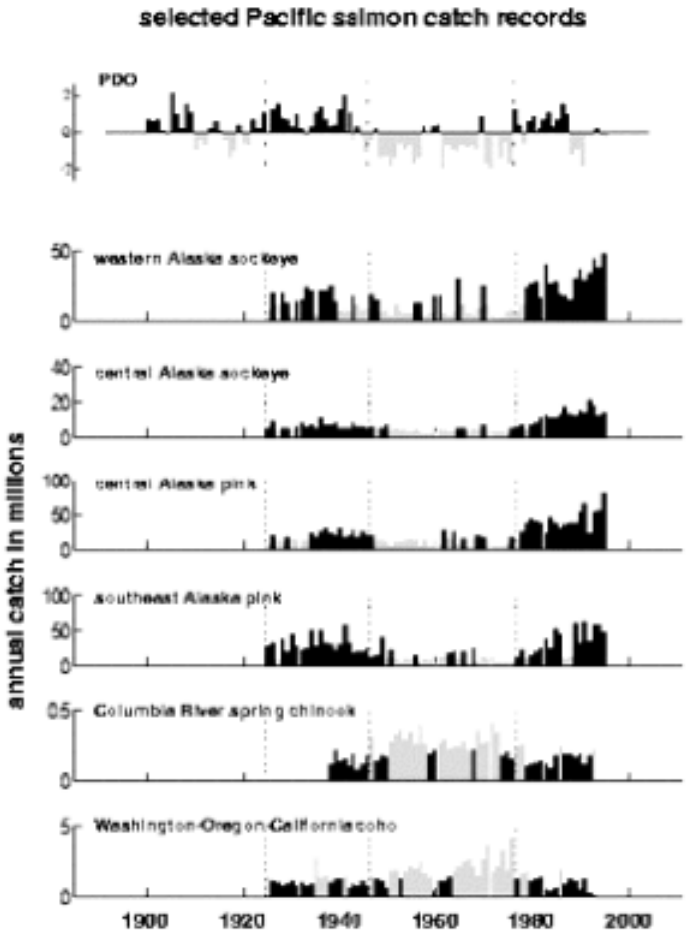


Figure 5.2: Historic salmon catches and relation to PDO. For the PNW (Alaskan) stocks, grey bars (black) correspond to values greater than the long-term median. The dotted lines mark PDO regime shifts in 1925, 1947, and 1977.

5.1.2 PDO and ENSO implications for salmon in the Yakima Basin

Beyond altering ocean conditions, PDO and ENSO also impact streamflow conditions in the Yakima basin, which in turn affects the instream survival rates of salmon. We have noted that the primary teleconnection influencing the winter climate of the PNW is the PNA, which operates on an annual basis. The teleconnection patterns, PDO and ENSO,

serve to modulate the PNA on longer time scales, for example, interdecadally and interannually, as described in the second chapter. The modulating effects of PDO and ENSO on the Yakima's hydroclimatology are described in greater detail below.

Some strong trends in the snow pack and spring runoff in the PNW coincide with the various phases of ENSO and PDO (Lluch-Cota *et al.,* 2003, McCabe and Dettinger, 1999, and Clark *et al.,* 2001). Cayan (1996) notes that SWE and annual runoff in the PNW are positively correlated to the Southern Oscillation Index, which is a measurement of the pressure systems associated with ENSO. La Niña years generally correspond to years with increased SWE and spring runoff in the PNW, while El Niño events have the tendency to bring drier, warmer conditions during the PNW winters. Clark (2001) describes the physical mechanism by which ENSO affects the PNW climate as amplifying the northern branch of the North American jet stream and deflecting storms north toward Alaska. The general effect of ENSO on the spring runoff volumes in the Yakima basin is noticeable. The following figure displays the spring runoff volumes for Cle Elum Reservoir for the period of 1950-2002 with the Multivariate ENSO Index (MEI) for the same time frame (1950-2004). The MEI is a weighted average of six variables such as SST and pressure, that define main ENSO features in the tropical Pacific (Klaus, 2004).

It is apparent that there is a strong relationship between spring runoff in the Yakima basin and ENSO events. The five year intervals of the MEI correspond to the same five year streamflow intervals. Warm ENSO events are noted as the red departures from the 0 MEI scale and cool events are registered in blue. Significantly strong ENSO events are those with departures of 2 or more in either direction. The dashed line through the streamflow time series at 400,000 acre-feet is the 50^{th} percentile runoff volume seen at Cle Elum. While there is not always a direct one to one relationship between an MEI event and the runoff, there is a weak correlation. The strongly positive departure of the MEI in 1973 corresponds to the lower spring runoff conditions seen at Cle Elum and at the other forecast sites. The

preponderance of El Niños in the 1980s and 1990's is reflected in an extended period of below 50th percentile flows at Cle Elum. However, not all years reflect a clear relationship. When correlated with the first principal component (PC) of the spring runoff the correlation value was only 0.11. For instance, the 2001 drought corresponded to a prolonged La Niña that lasted from 1998-2002 (Hoerling and Kumar, 2003). Thus, we must refrain from suggesting the relationship is purely linear. The other forecast sites display similar historical spring runoff trends as those seen at Cle Elum.

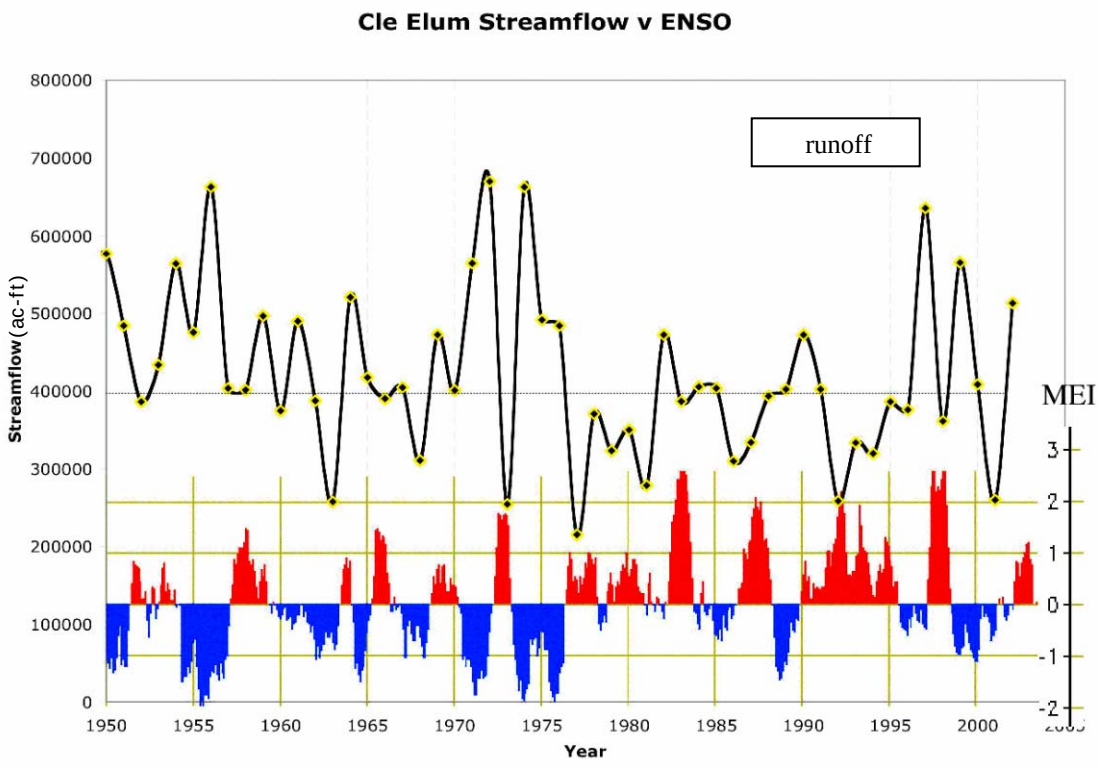


Figure 5.3: Spring runoff vs. MEI. Red departures indicate El Nino conditions, blue corresponds with La Nina conditions.

The Yakima spring runoff time series also shows a strong relationship with the PDO index. While the length of our streamflow time series is such that only two PDO regime shifts are captured, a definite pattern emerges. The correlation value of the first PC with the PDO index was 0.34, which is significant to approximately the 90th%. The zero line of the

PDO index is set on the 50th percentile runoff for the Cle Elum in the next figure. From 1944 to 1975, the PDO was in a predominantly cool phase, as denoted by the grey bars in the figure. During this period, the total spring runoff volume across all six forecast sites was on average, 49,000 acre-feet greater than the spring runoff volume seen in the 1977 to 2002 period.

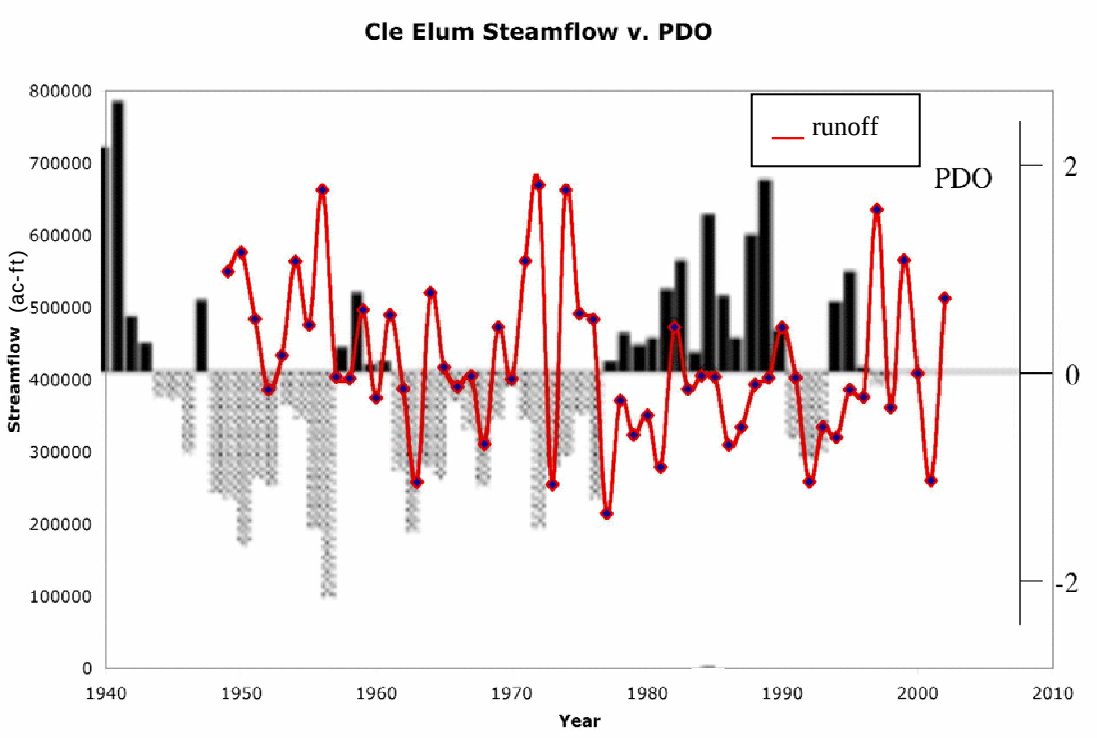


Figure 5.4: Spring runoff vs. PDO

In 1976, an abrupt phase change coincided with one of the worst droughts in the historical Yakima record in 1977. The warm phase of the PDO lasted roughly from 1976 to 1999, although there is some debate as to whether or not a shift toward the cool phase occurred. While the PDO was in the warm phase, the Yakima basin witnessed several years of below normal runoff. As noted with ENSO however, the relationship between streamflow and PDO is not linear. For example, the PDO was briefly cooler in the early 1990s during which the Yakima basin experienced a multiyear (1992-1994) drought. There is still

important information to be gained from watching the phase shifts of the PDO and ENSO however, which could be used as guiding indicators of the PNA's strength and structure. Water managers could make more informed forecasts by watching the PNA's development in conjunction with ENSO indicators. Fish managers could watch ocean and instream conditions.

Anadromous fish and related salmonids, such as the bull trout, have complex life cycles during which habitat requirements change. Salmon require cooler water temperatures than many fish and good water quality (Watershed Assessment, 2003). The surviving salmon species in the Yakima basin have different emergence and migration periods, making it difficult for the USBR to address the needs of all the species throughout the year. The preferred spawning grounds also vary between salmonids. Spawning salmon seek sheltered waters with good pool-riffle sequences, sufficient woody debris, and shaded areas.

The following set of figures portrays the timing of salmon migration (for five salmon species) and the preferred spawning reaches in the Yakima of the spring chinook. The fall chinook has a much shorter spawning run in the fall and generally spends only one year rearing in the freshwater. Fall chinook juveniles tend to migrate during the spring runoff when high flows help facilitate their passage. Spring chinook generally migrate back to their natal stream in the spring. Spring chinook fry emerge in the spring but do not complete outmigration until the next spring. Steelhead return in the fall but overwinter in various locations throughout the basin before spawning. Unlike other salmonids, steelhead can migrate and spawn multiple times before dying. Steelhead juveniles may also spend up to four years rearing in the freshwater environment before beginning outward migration. Steelhead juveniles have been observed to take two years to complete migration out of the Yakima basin, overwintering at several locations (IOP, 2002).

Habitat/Species Associations Spawning, Rearing, & Key Migration	Bull trout	Steelhead Rainbow trout	Spring Chinook	Fall Chinook	Sockeye
Lower Mainstem (Status)				P	S
Middle Mainstem (Roza & Tieton)			P	P	S
Middle Mainstem Side Channel			P	P	
Upper Mainstem (Dams)	S	P	P		S
Upper Side Channel		P	P		
Major Tributaries	P	P	P		P
Headwaters	P	S			
Ephemeral	S	S			
Lake	S				P
Habitat Use P = Primary & S = Secondary	S				P

Figure 5.5: Preferred spawning habitat of Yakima salmon species. P = primary spots, S = secondary spots.

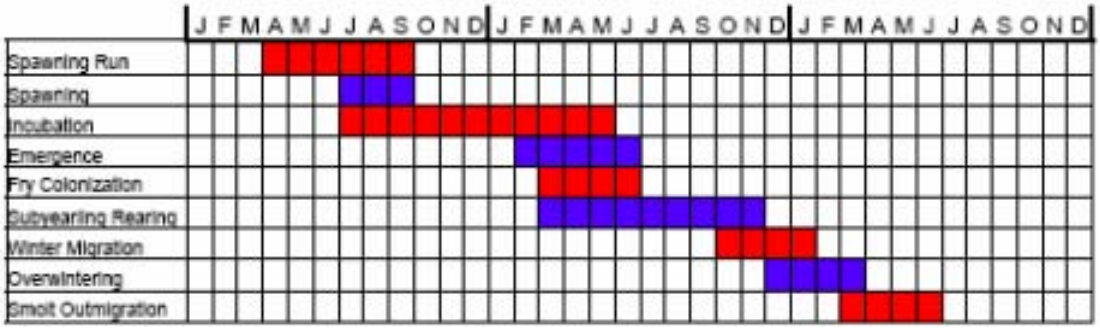


Figure 5.6: Mean timing of freshwater lifestages of spring chinook in the Yakima Basin. Months are on top of figure.

The flow needs of salmon also vary with the instar. During spawning, the salmon prefer reaches where the flow is not too fast, yet fast enough to provide good oxygen flow to the redd and keep it covered. As fry emerge, they are vulnerable to predation, temperature fluctuation, and fast flows against which they cannot swim. They require cool water, sufficient water quality, and abundant food supply (Smith, 2004). Developing fry and parr are especially susceptible to temperatures fluxes and degraded water quality. Turbidity and fluctuating streamflows seen during the irrigation season are impediments to salmon growth and development (Anderson, 2003). During drought years, which coincide strongly with the warm phase of the PDO, water temperatures tend to be warmer and pollution concentrations increase due to low flows. These poor water quality conditions are further exacerbated by on-stream agricultural diversions, the timing of releases, and agricultural return flows.

Juvenile migratory salmonid survival is dependent on instream conditions, while non-migratory salmonids are affected throughout their entire life cycle. Anderson (2003) found that survival rates of migratory juveniles along the Snake River decreased significantly with increasing temperature and turbidity. The survival response across several salmon species varied, but all exhibited a relative downward trend with decreasing water quality. Migrating smolts require sufficiently high flows to speed their passage to the ocean. Many smolts of all five salmon species normally migrate during the spring runoff to take advantage of higher flows. However, spring runoff has been strongly attenuated by river management as reservoirs are operating on flood control or are filling. The seven irrigation divisions also begin withdrawals around April 1st to prime their canal and waterway systems. Entitlements are filled through natural runoff until the reservoirs go on storage control, which is usually around the beginning of June (IOP, 2002). These practices further decrease the available flow, especially in tributaries, during the spring runoff season. During drought years in the Yakima basin, managers may begin filling reservoirs as soon as runoff commences, which further lowers streamflows during the time when smolts are beginning outward migration.

Current studies (Columbia/Snake River Temperature TMDL, 2002, herein referred to as TMDL, Anderson, 2003) generally focus on temperature exceedence resulting from low streamflow or altered flow regimes. Temperature exceedence is a widespread problem during summer months when the various reaches of the Yakima River have been drawn down for irrigation or are receiving warm return flows. The next figure displays temperature fluctuations throughout the year at several locations on the lower reaches of the Yakima River. Earlier it was noted that the majority of water in the lower reaches of the Yakima River during the summer months is comprised of return flows. Temperatures remain elevated through October when several salmon species have completed spawning and eggs are incubating.

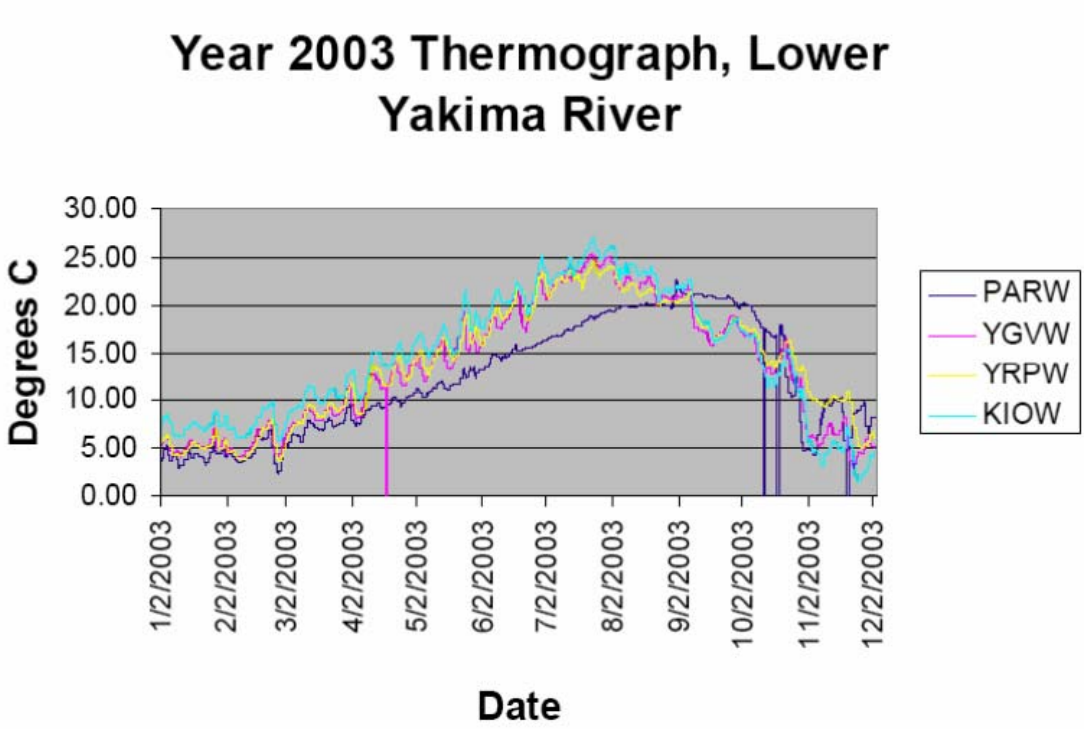


Figure 5.7: Temperatures at select locations in lower reaches of Yakima. Fatal temperatures are routinely recorded from mid-May through October.

The effects of pollutants such as DDT, pesticides, and fertilizers, are more difficult to quantify, as these are non-point sources pollutants. However, the United States Geological Survey (USGS) has been conducting an extensive study of surface water quality in the Yakima basin as part of the larger NAQWA study effort. In the upper subbasin, the primary contaminants of concern were natural inputs such as arsenic and mercury. A basin wide assessment of water quality issues was carried out for numerous locations in 1998 (WA. Dept. of Ecology, 1998). The study indicated that the biggest water quality issues affecting salmon recovery in the headwater reaches are temperature exceedence, DDT, dieldrin, mercury, and cadmium. This study did not monitor DDT or dieldrin sorption from soil into the water. DDT and dieldrin levels were detected by sampling fish tissue. The sample size of concentrations in fish tissues was not as large or robust in some sampled species as in others. A follow-up study was conducted in 1999 to verify the 1998 303(d) listings. While some concentration values were not as high as in the previous study, it was determined that

DDT and dieldrin still pose aquatic and human threats. Unfortunately, continuous monitoring is limited to three stations along the Upper Yakima, with only one being close to the headwater reaches. The Washington State Department of Ecology is in the process of updating and verifying listings for the same sites visited in 1998.

The initial results of the USGS study, released in 1999, indicated that surface water quality in the lower subbasin was severely impacted by irrigation return flows and runoff from livestock lots. Water quality is of extreme concern in the lower subbasin below Union Gap, WA. As noted in the first chapter, much of the irrigated acreage is located in the lower subbasin. Substantial dairy farming has also become an economic force in this area (Natural Resources Defense Council, 1998). Return flows in the lower reaches of the Yakima River constitute nearly 80% of the summer time water in the river (USGS, 1999; IOP, 2002). Return flows from the agricultural drains were found to contain significant levels of nutrients, pesticides, fecal coliform bacteria, and suspended sediment. Ecology listed portions of the lower Yakima as exceeding state water quality standards for fecal coliform bacteria in 1997, when levels were nearly 200 colonies per 100 milliliters (Ecology, 1997). Two dairy farm spills resulted in colony levels of 35,000 per 100 milliliters in the Sunnyside Irrigation District in 1997 (USGS, 1999). Water quality is thus seriously degraded in the lower Yakima River during the summer when natural flows comprise a minority of the total water in the river. Yet, these are the conditions through which the adult spring chinook salmon must travel to reach spawning grounds in the upper Yakima. Resident bull trout are subjected to impaired water quality at all life stages. Steelhead salmon are also subjected to summer low flow water quality conditions throughout several life stages.

It stands to reason then, that salmon survival rates are not only impacted at sea by the PDO and ENSO cyclical variations, but are also impacted by instream conditions that vary with the phases of these teleconnections. The strong relationship between ENSO, PDO, and runoff conditions in the Yakima basin indicates that PNW salmon face poor conditions both

in the ocean and inland during the warm phases of the PDO or ENSO. This underscores the importance ensuring sufficient instream levels and strict water quality measures during years of drought when returning salmon are already stressed due to poor ocean conditions.

5.2 Temperature Change in the PNW

A number of researchers (Mote, 2003, Parson, 2003, Miles, *et al.*, 2000) have noted an average warming of 1 to 3°F over the region in the 20th century and project that warming will continue at an increased rate through the 21st century. Both the winter and summer months seem equally effected by warming. Average regional precipitation has increased approximately 11% (Parson, 2003). As discussed earlier, the Yakima basin derives most of its water for storage and runoff from snow melt, as little precipitation occurs between March/April and November. We also noted that the snow pack in the Yakima basin is highly responsive to slight temperature variations, with severe flooding during rapid, late winter thaws or when rain-on-snow events occur. Mote (2003) found the correlation between SWE and temperature to be very strong in the Washington Cascades when elevations are below 1500m. Much of the mountainous portions of the Yakima basin lie below these altitudes and are therefore sensitive to winter temperature fluctuations.

The implications for the amount and duration of the snow pack in the Yakima basin are important to water managers, streamflows, and instream salmon survival. Over the past 50 years, Mote (2003) has noted a decrease in SWE in many areas of the PNW. Mote performed multiple linear regression using precipitation and temperature to look for trends in April 1st SWE. The Cascades saw significant decreases in SWE from 1950-2000. Concerns have been raised as to whether this is a long term trend or part of a cyclical pattern because the data record only covered 50 years. The following figure displays trends in SWE throughout the region. Open circles indicate decreases in SWE. We have outlined the

Yakima basin in a rectangle. The majority of the snow course sites in the Yakima basin indicate a significant decrease (~40%) in SWE from 1950-2000.

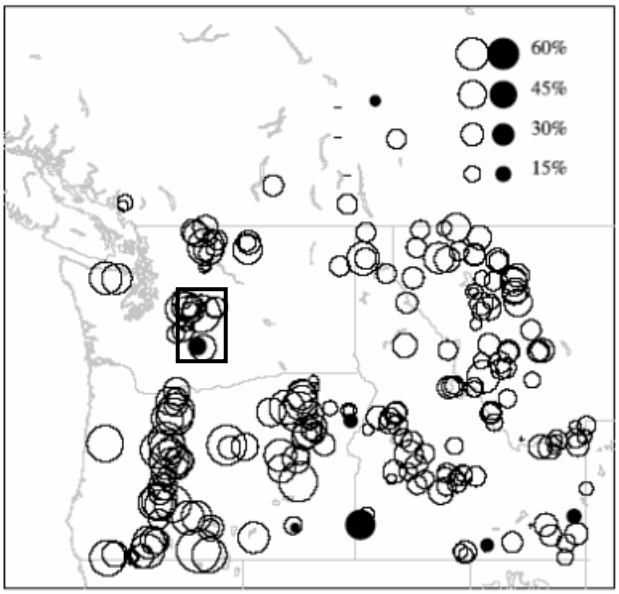


Figure 5.8: Linear trends in April 1st SWE between 1950-2000. Open circles indicate a decrease. The rectangle highlights the Yakima basin.

Regardless, the implications for water managers are bleak. If SWE is truly decreasing and regional temperatures are warming, the Yakima basin faces greater strain on its water supply. Warmer winter temperatures indicate that a fraction of the precipitation is now falling as rain rather than as snow. This increases the likelihood of winter flooding events in which the Yakima's small reservoirs must operate on flood control, rather than attempting to store runoff generated by the rain. Lower snow pack also leads to lower spring runoff volumes, decreasing the TWSA. Reservoirs will have to begin operating on storage control (making releases for Title XII flows and demand) earlier in the season, which effectively extends the period of storage control. Because the reservoirs only store 1 MAF of water, the remainder of irrigation demand must be met through unregulated tributary flow and bypassed reservoir inflow (IOP, 2002).

All these trends indicate a high probability that streamflow levels will be reduced throughout the basin, hampering the efforts of salmon recovery. Title XII flows are in effect at only two points along the lower Yakima mainstem. Minimum instream flows are nonexistent in other reaches of the mainstem or any of the tributaries. During years of normal flow, there are reaches of the Yakima in which stream levels become critically low and smaller tributaries that are dewatered during the summer months (Watershed Assessment, 2003). If reservoir storage is low during a drought year, the irrigation entities are more reliant on the unregulated tributary flow and runoff to meet delivery obligations. Nor is the USBR responsible for fulfilling all the water entitlements in the basin. Some significant diversions account for a 130 cfs per day reduction in streamflows that is not managed by the USBR (IOP, 2002). Thus, tributaries and the upper reaches of the Yakima mainstem are even more stressed during drought years.

We have noted that salmon survival in the Yakima basin is affected by a number of flow related issues such as instream levels, temperature, and agricultural runoff. Survival rates are likely to be affected especially during years of low flows. In drought years, tributaries in the upper region are dewatered, temperatures are raised and sustained, and concentrations of species toxic to fish are increased. Low flow regimes correspond significantly with the phases of PDO and ENSO. Drought years in the Yakima basin have the tendency to coincide with the warm phases of these two teleconnection patterns. Thus, water quality and salmon survival within the basin are modulated to some degree by PDO and ENSO. This is significant because previous studies have only examined the link between salmon survival in the ocean with the PDO regimes. We have shown that this link extends to inland situations as well.

5.3 Thesis Conclusion

The Yakima River Basin exemplifies the range of water management issues facing water managers across the United States. The basin is arid because it is in the rain shadow of Mt. Rainier and receives a basin average of only 27 inches of precipitation per year. The majority of the precipitation, approximately 80%, falls as snow during the winter months. Thus, the basin is entirely dependent upon snowmelt for the water supply. Despite being so arid, the Yakima Basin has become the nation's leading producer of apples, pears, cherries, and hops. All these soft fruit crops require significant amounts of water to grow, which must be delivered through a complex network of irrigation canals and ditches.

Perennial streams in the basin were completely overappropriated by the late 1800s. The unreliable water supply and resulting disputes caused farmers to petition Congress for help. Congress passed the Reclamation Act in 1902 and created the Bureau of Reclamation with the intent of providing a more reliable water supply to farmers in the western United States. The USBR constructed five reservoirs in the basin with a combined capacity of around 1 MAF, which is enough water for the demands of one operational water year. The USBR operated the reservoirs with the primary purposes of meeting irrigation demand, flood control, recreation, and to a lesser degree, power production. Operation priorities began shifting in the 1980s when a precipitous decline in one of the basin's other economic staples, salmon, began to be addressed.

Six salmon species once spawned the Yakima River and its tributaries. The salmon provided the cultural identity and economic livelihood of several Native American tribes, now represented by the Yakama Nation. Massive westward migration severely altered the lifestyle of the Native Americans and disrupted the life cycle of the salmon. Pesticide laden agricultural runoff, impassible reservoirs, and highly altered flow regimes are a few of the anthropogenic factors responsible for the decline of salmon within the basin. Outside the basin, heavy commercial fishing and polluted ocean waters have contributed to unfavorable

conditions. By 1980, three of the six salmon species common to the Yakima Basin, sockeye salmon, summer chinook, and coho salmon were locally extirpated. Bull trout and steelhead are currently listed as threatened under the Endangered Species Act. Such declines prompted Congress to pass the Title XII legislation in 1994, which requires the USBR to maintain target flows at two locations in the lower subbasin.

The target flows are not strictly defined amounts, but rely on the USBR's forecasts of the total water supply available (TWSA) for the upcoming irrigation season. The TWSA is determined for April 1st through a combination of reservoir storage levels at the end of March, anticipated spring runoff at the five reservoirs and three other locations in the basin, and anticipated return flows. The possible target flows are determined from the TWSA. In years of normal or above average streamflow, the TWSA is sufficient to meet demand, which averages 2.5 MAF. During drought years, consumptive demand cannot be met and prorationing occurs. There are two types of water entitlements in the Yakima Basin, nonproratable and proratable. Water rights with a priority date of 1905 or earlier must be fulfilled in entirety in all years and are nonproratable. Water rights with a later priority date are said to be proratable. These rights receive a fraction of the full entitlement in years of drought after the target flows and nonproratable rights have been fulfilled. Many water users hold a mix of proratable and nonproratable rights, but some of the younger irrigation districts such as Roza, have only proratable rights. These water users are most vulnerable in years of scarcity.

The forecasted TWSA is central to the USBR's planning operations for the Yakima Basin. Overestimation of the water supply can lead to faster drawdown of the reservoirs, dewatering of reaches, and a shortened water season. Both salmon and the irrigators are deleteriously affected by an overestimation. Likewise, underestimation means lower target flows for the salmon and tighter proration levels. Both situations have severe biological and economic implications for the basin. Thus, the USBR forecasters are constantly searching for

methods that improve the reliability of the forecasts and increase the lead-time at which the forecasts could be issued.

This thesis attempted to examine the possibility of improving forecasts for the USBR by incorporating large-scale climate information. In the second chapter, three teleconnection patterns were identified as playing a significant role in modulating the winter stormtracks that provide the Yakima Basin with the majority of its precipitation. Through correlation of the spring runoff with climate indices such as the 500 mb geopotential height, the PNA pattern was identified as the most significant teleconnection. Two pressure systems in particular, the Canadian ridge and the Aleutian low, dictate the movement of the winter stormtracks into the PNW. The strength and location of these two pressure centers are modulated at longer time scales, interannually and interdecadally, by ENSO and PDO. When the Canadian ridge and the Aleutian low are anomalously strong, spring runoff decreases because the winter stormtracks have been deflected northward toward Alaska. The opposite affect has been noticed in years when the two pressure centers are weak.

Based on the information gathered in the second chapter, a predictor set consisting of the 500 mb geopotential height, SSTs, and SWE information was utilized in the formation of stochastic streamflow forecasts. These ensemble forecasts were generated using a local regression method that did not make assumptions about the underlying data distribution. A modified k-nearest-neighbor approach, developed by Prairie (2002) and adapted by Grantz (2003), was utilized to perturb the forecasts and generate streamflow scenarios not seen in the historical record. The forecasts were generated for four different time frames beginning with Oct. – Dec. (OND) climate information and ending with Dec. – Feb. (DJF) climate information coupled with April 1st SWE information. The skill of the forecasts was measured and compared with forecasts created with only SWE information. It was found that forecasts using SWE only performed as well as forecasts incorporating climate during years of normal streamflow. The addition of climate information proved crucial in years of extreme events,

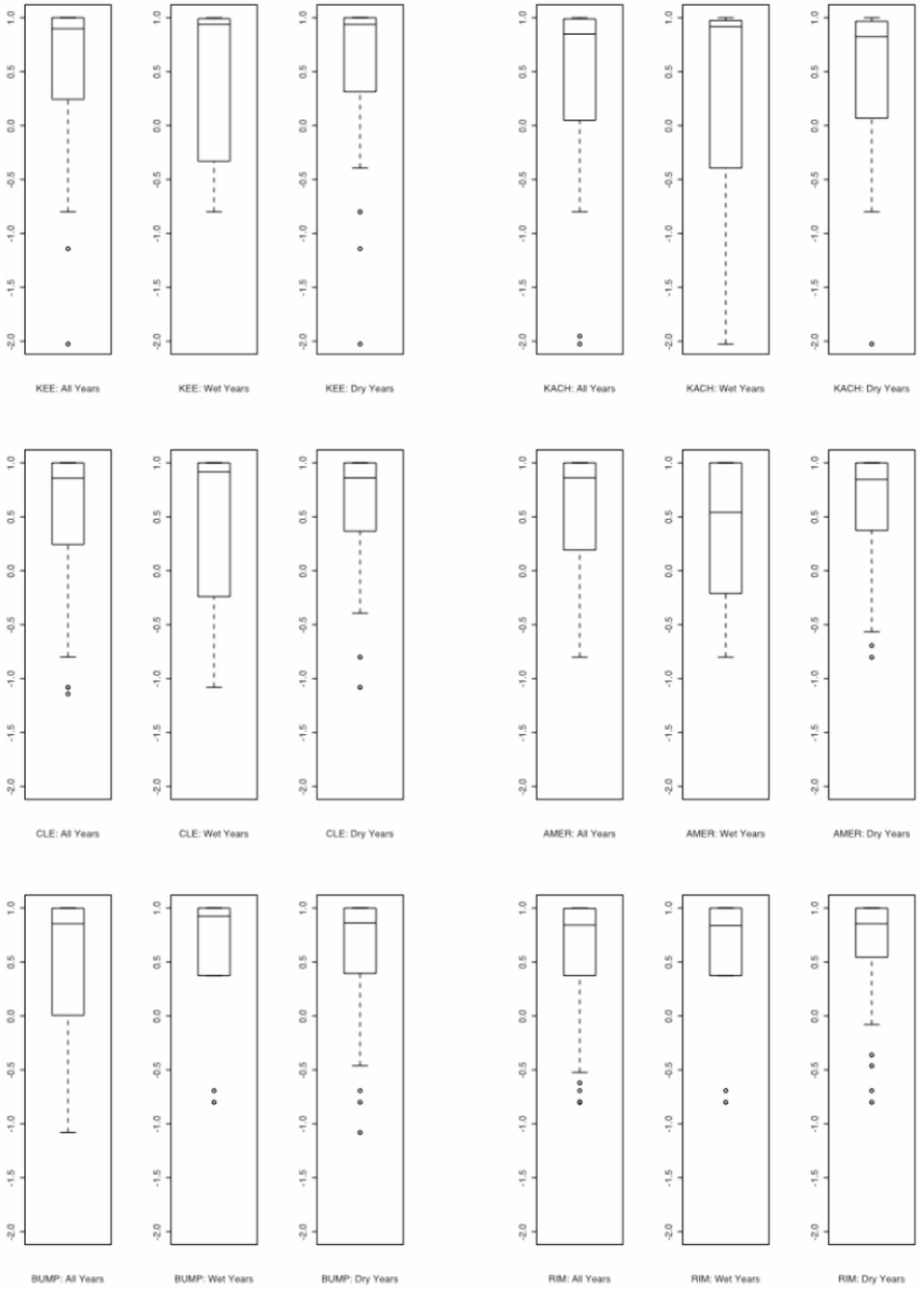
and pushed the forecasts closer to the observed values. Some drawbacks to the modified k-nn approach were noted, as well as, suggestions for improving the forecasts.

The fourth chapter examined the usefulness of the ensemble forecasts to water managers in the Yakima Basin. The possibility of incorporating the forecasts into a decision support system model of the basin, rendered in RiverWare, was explored. While the incorporation proved difficult, the implications of our forecasts for two different drought scenarios, 1977 and 2001, were examined. The range of streamflow values produced for the 2001 season tended to overpredict the spring runoff but still would have been useful to the forecasters. The 1977 drought was the worst drought on record for the Yakima Basin. The ability to capture the 1977 drought was key, because biological target flows did not exist at the time. Our forecasts accurately captured the 1977 drought as early as the beginning of February and indicated that proration levels would have been much more severe if target flows had been in place. The benefits to water users and managers resulting from the increased lead-time and accuracy of our forecasts were explored at the end of the fourth chapter.

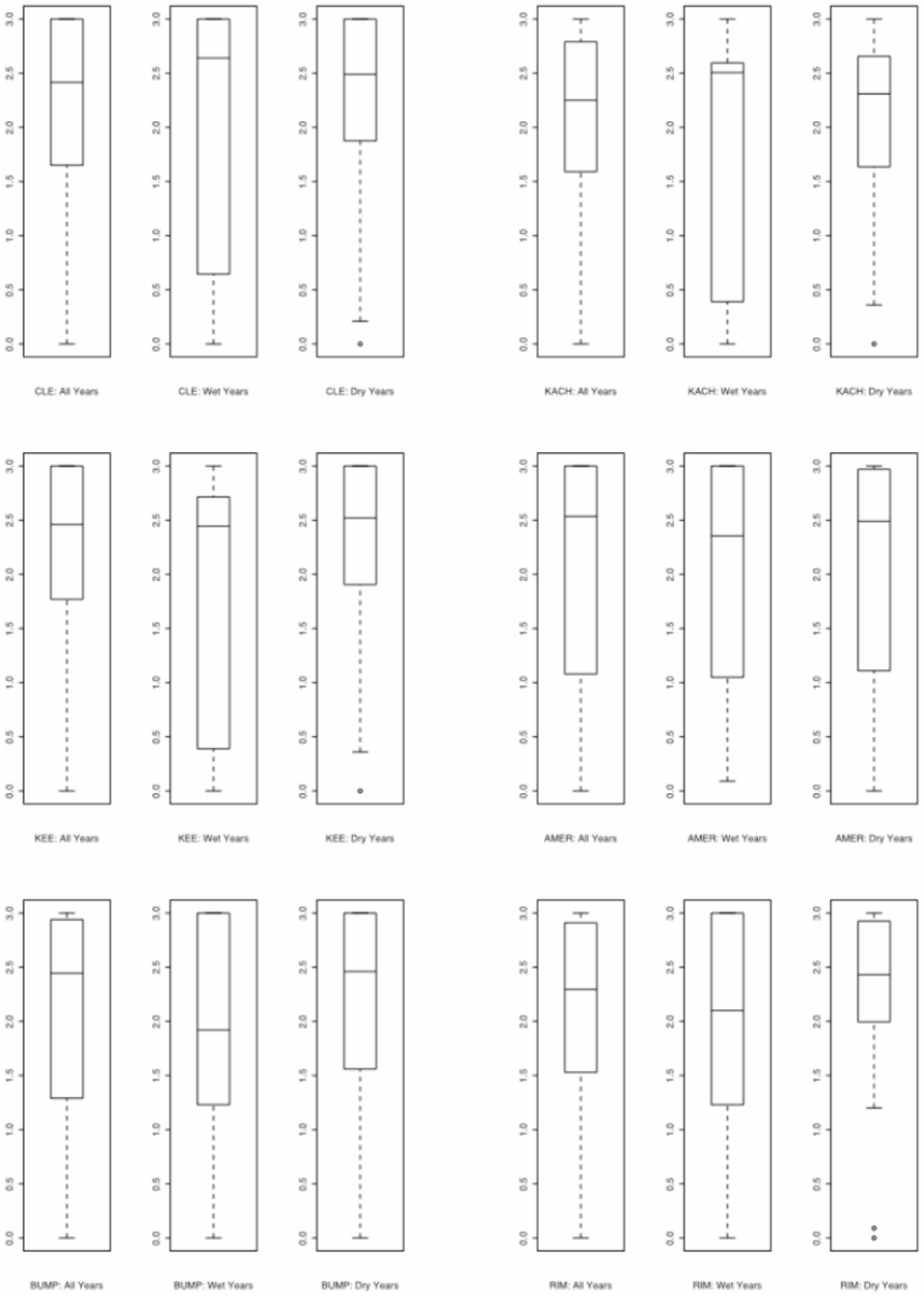
The fifth and final chapter addressed the affects of climate variability on salmon survival. Previous studies had only examined the role of large-scale climate features such as PDO and ENSO on salmon survival in the ocean. We demonstrated that the affects of climate variability extend inland to the Yakima River Basin. PDO and ENSO affect the timing and volume of the spring runoff, which dictates the water supply of the basin. Years of lower streamflow correspond strongly with the warm phases of PDO and ENSO. This knowledge has significant implications because lower streamflow is known to lead to higher water temperatures and increased concentrations of various contaminants that affect salmon mortality. Thus, we have shown that salmon survival instream is linked to climate variability. This thesis is concluded.

Appendix A

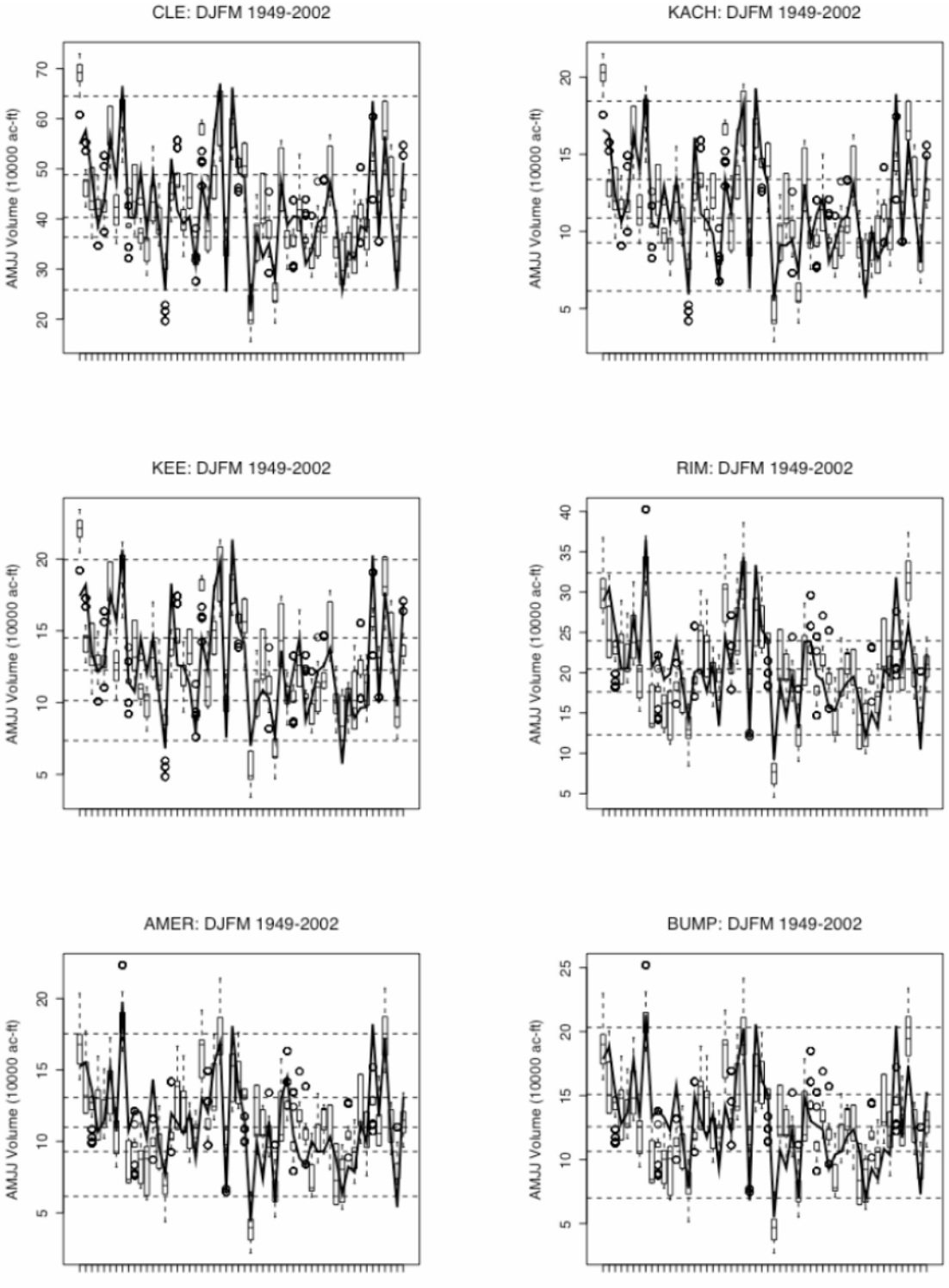
DJFM RPSS



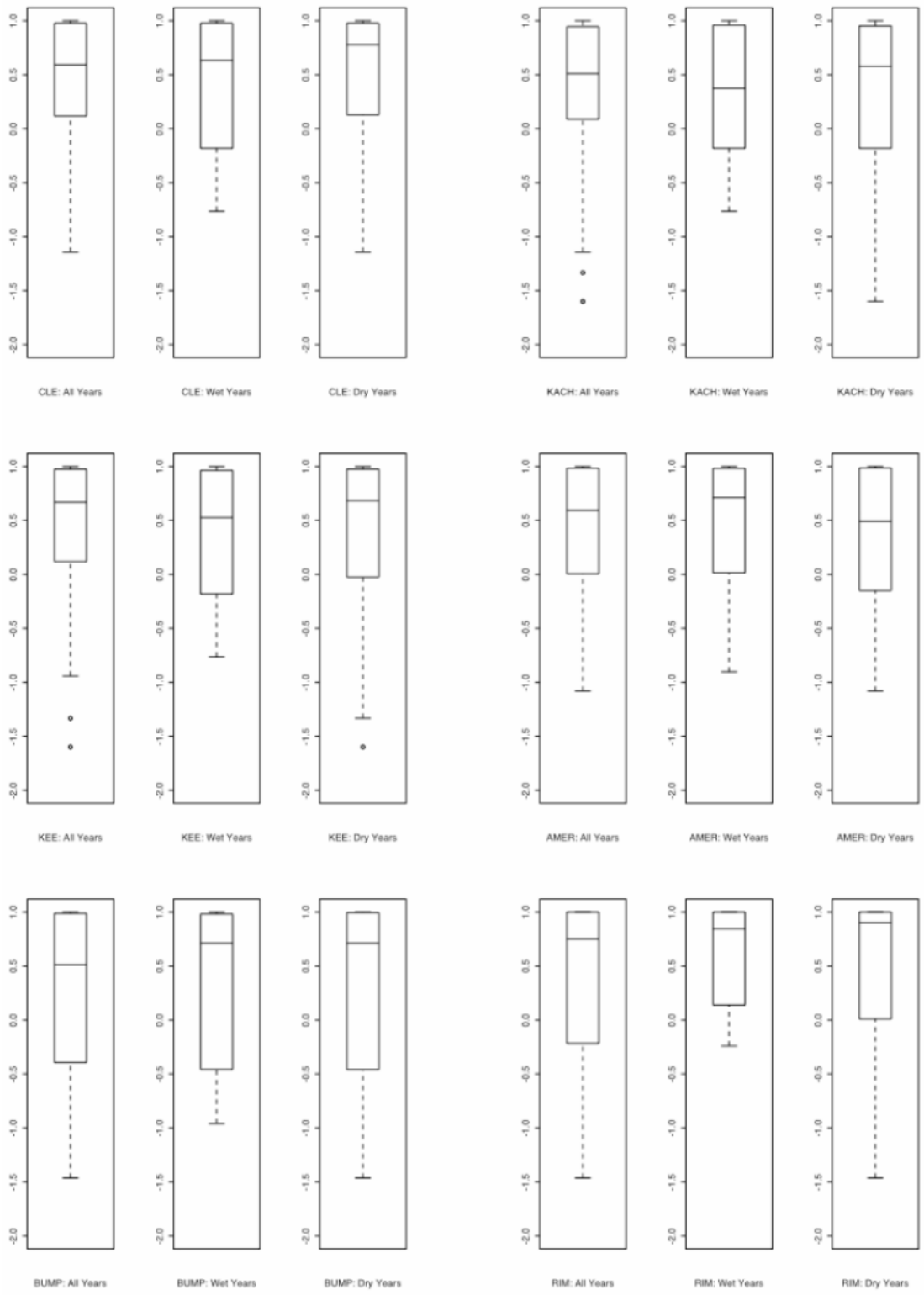
DJFM Likelihood Skill Scores



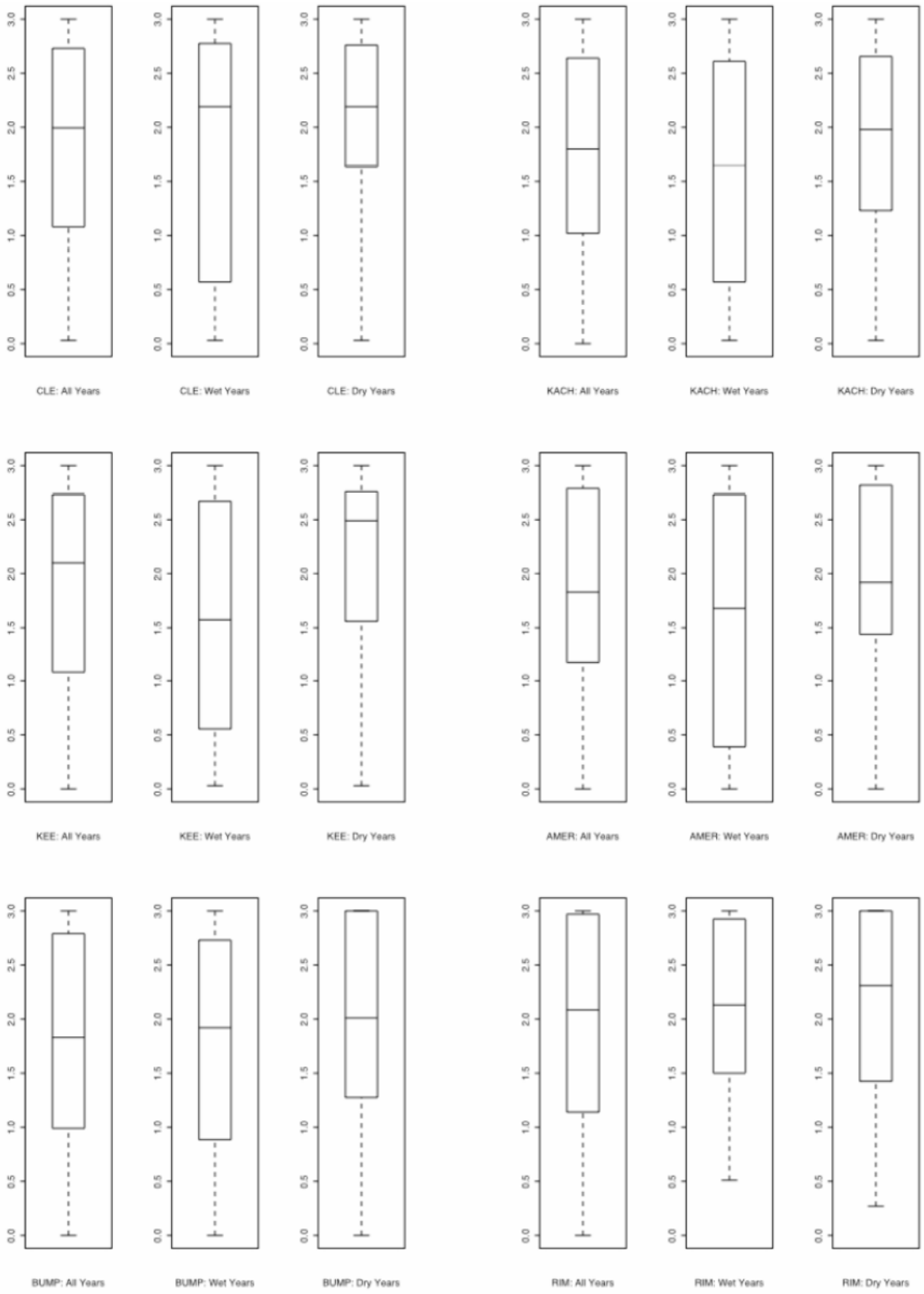
DJFM Ensemble Plots: 1949-2002



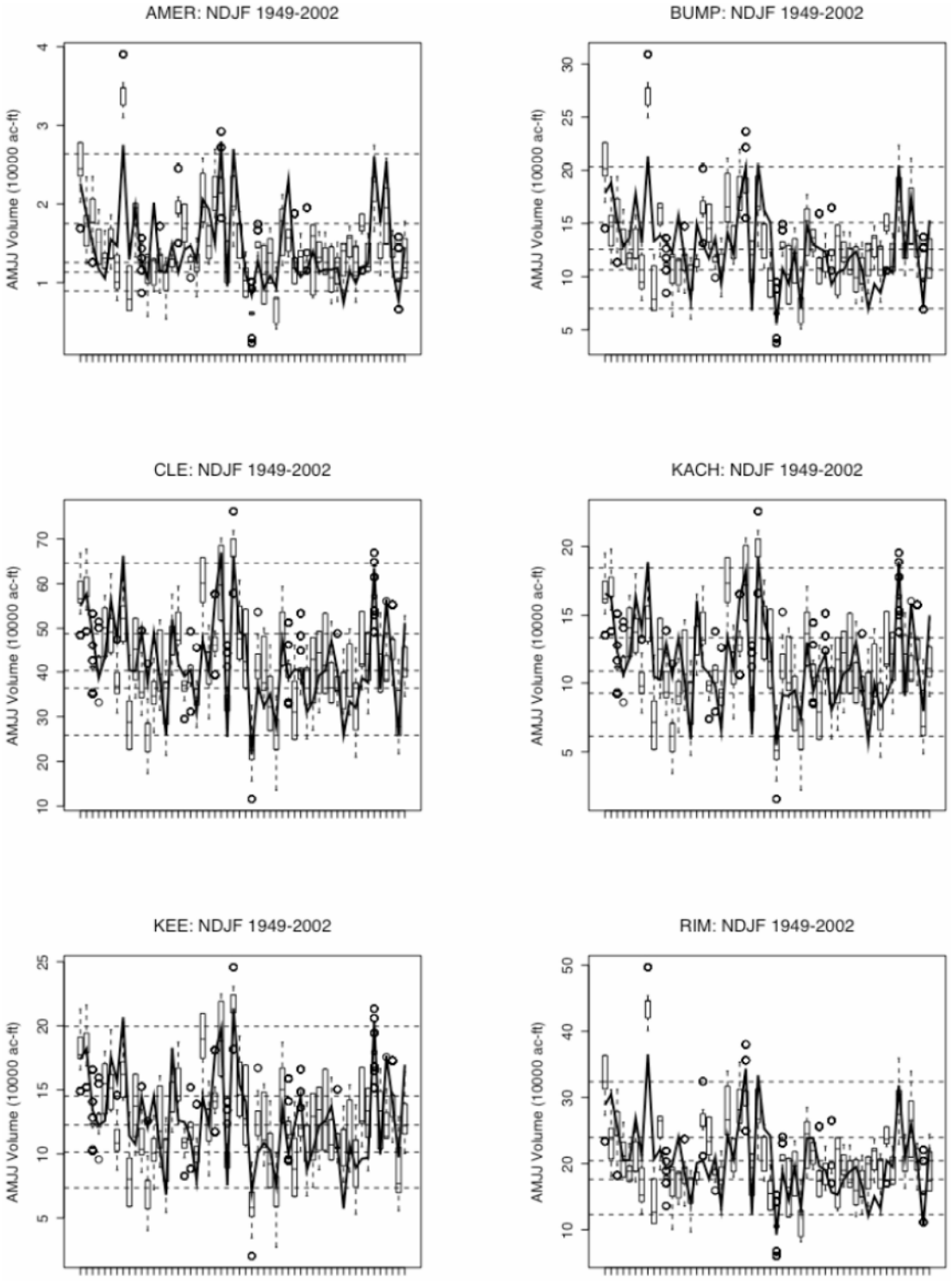
NDJF RPSS


CLE: All Years
CLE: Wet Years
CLE: Dry Years
KACH: All Years
KACH: Wet Years
KACH: Dry Years
KEE: All Years
KEE: Wet Years
KEE: Dry Years
AMER: All Years
AMER: Wet Years
AMER: Dry Years
BUMP: All Years
BUMP: Wet Years
BUMP: Dry Years
RIM: All Years
RIM: Wet Years
RIM: Dry Years

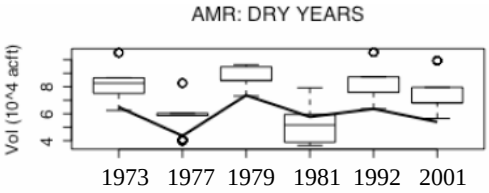
NDJF Likelihood Skill Score

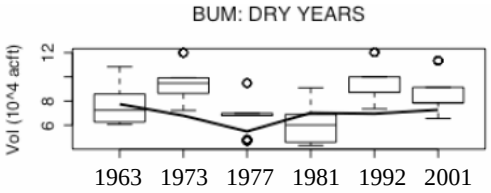


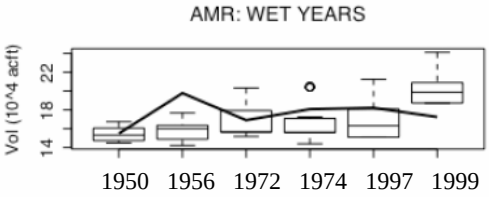
NDJF Ensembles: 1949-2002

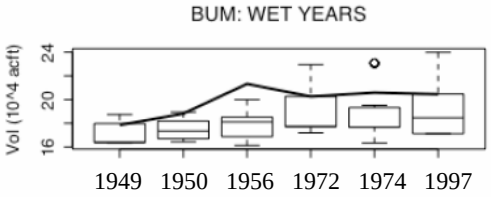


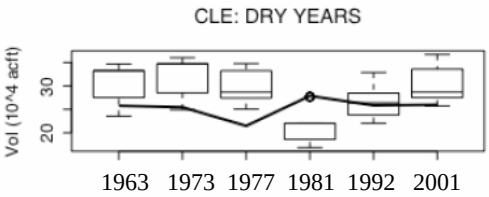
Ensemble plots for extreme years using only April 1st SWE

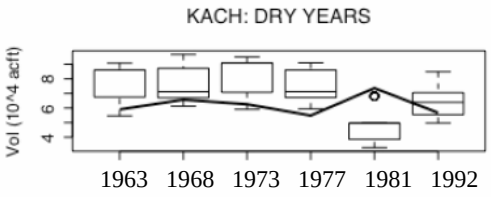


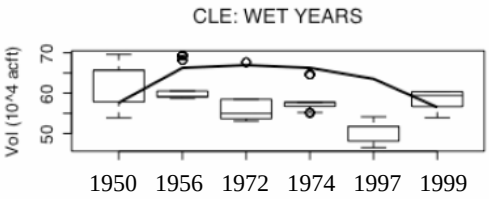


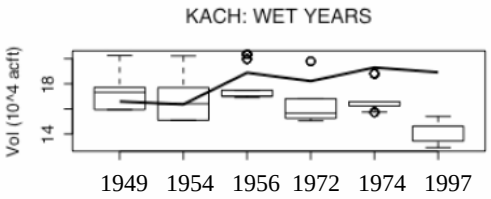


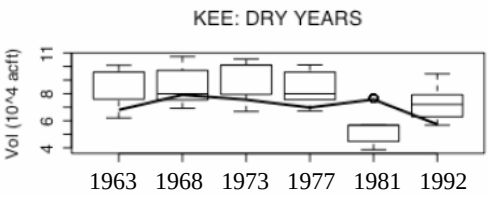


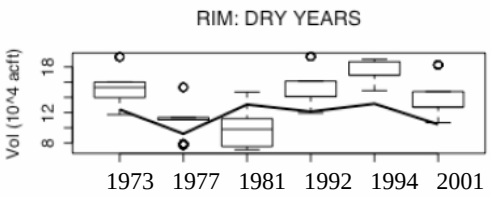


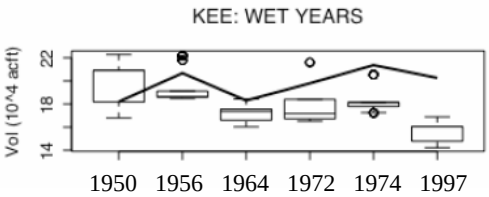


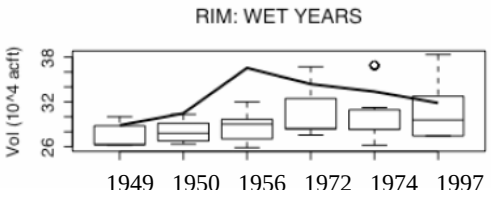




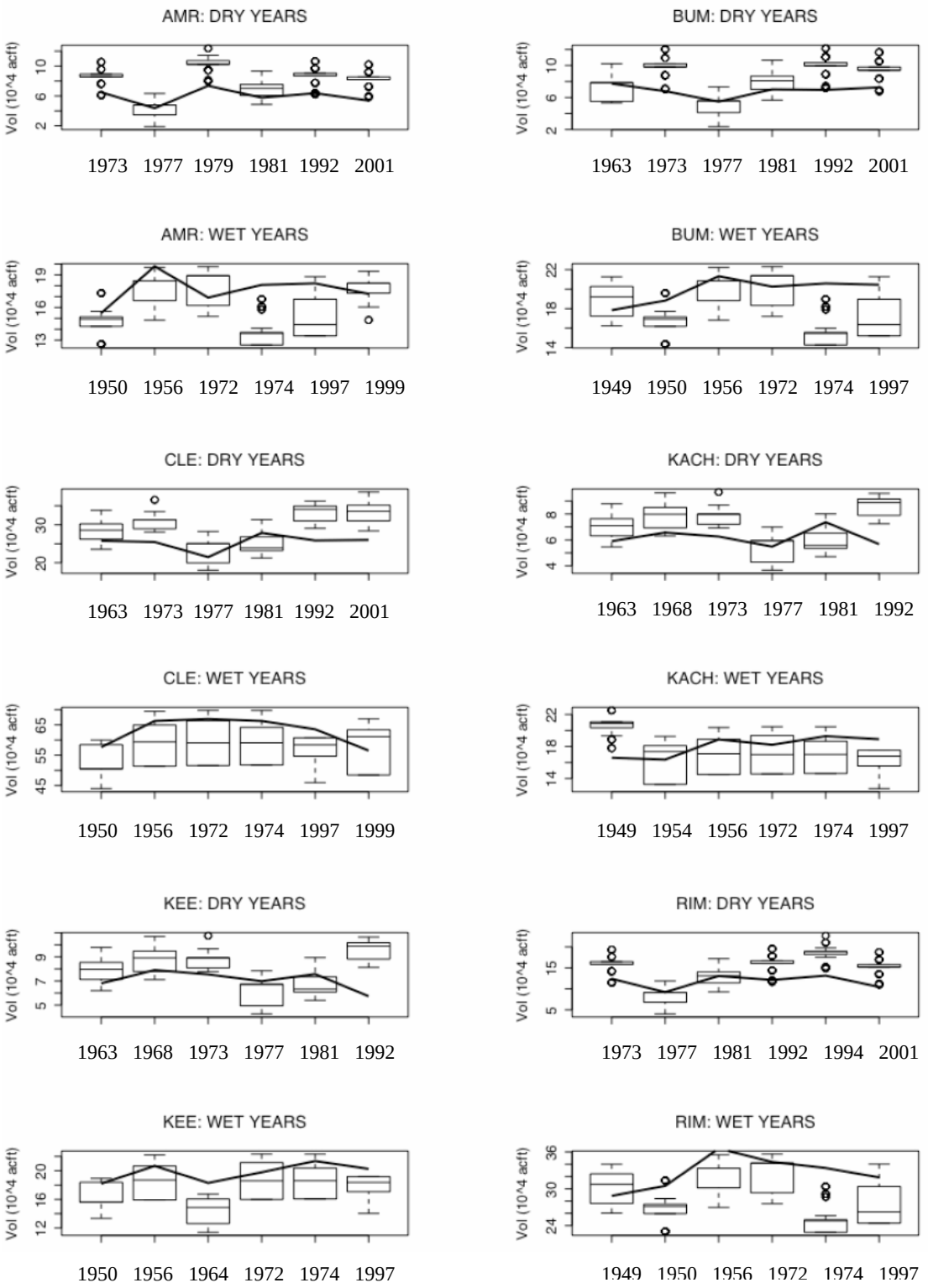








Ensembles of extreme years using only March 1st SWE



Appendix B

Glossary

Anadromous- refers to fish that migrate between fresh water and salt water. Anadromous specifically refers to fish that spawn in fresh water and live the majority of their lives in salt water.

Appurtenant- refers to a water body being adjacent to land. A requirement of riparian water rights

Fry- newly hatched salmon young that feed primarily on zooplankton until becoming large enough to consume benthic invertebrates and other insects.

Parr- older juvenile salmon that have developed stripe (parr) markings that help to camouflage the fish in their fresh water environment

Prior appropriation- a system of water rights common to the western United States. Water rights issued under the system of prior appropriation are separated from the adjacent lands. This allows water to be diverted and applied at locations away from the water body and must be applied to a beneficial use. The system follows the rule of first in time, first in right meaning that senior water rights have priority in being fulfilled before younger/junior water rights.

Redd- salmon nests

Riparian rights- a system of water rights more common to the eastern United States. The right to use water from a water body is tied to the land adjacent to the water body. Riparian right holders must share a supply equitably during years of scarcity.

Smolts- juvenile salmon that have begun undergoing physiological changes which allow them to live in a salt water environment. Smolts begin changing as they migrate toward the ocean.

Teleconnection- recurring, large-scale pressure and atmospheric circulation pattern anomalies that can persist for several months to years.

Appendix C

Modified K-NN algorithm adopted from Grantz (2003)

UPPER SUBBASIN

```
{
library(locfit)

# This file generates the ensemble forecasts for the upper subbasin using the K Nearest
# Neighbor algorithm developed by Prairie (2002) and adapted by Grantz (2003).

# These files contain the predictor sets and PC1 for the upper subbasin.
# The general format of the data file is according to column:
# 1) Year
# 2) PC1
# 3) Canadian ridge
# 4) Aleutian low
# 5) pressure difference between the two.
# 6) SST
# 7) and 8) SWE
OND = matrix(scan("USBR/UP_OND_predict.txt", skip=2), ncol=4, byrow=T)
#NDJ = matrix(scan("USBR/UP_NDJ_predict.txt", skip=2), ncol=6, byrow=T)
#DJF = matrix(scan("USBR/UP_DJF_predict.txt", skip =2), ncol = 8, byrow=T)
#JFM = matrix(scan("USBR/UP_JFM_predict.txt", skip=2), ncol=7, byrow=T)

# The predictors from which the forecasts are generated.
X = cbind(OND[,3], OND[,4])
#X = cbind(NDJ[,5], NDJ[,6])
#X = cbind(DJF[,5], DJF[,6])                    # change predictors as needed.
#X = cbind(JFM[,5], JFM[,6]), JFM[,7])          #(Pres Diff, SST, SWE)

XX = scale(X)                                  # Normalize the predictors
Y = OND[,2]                                    # Observed PC1s.

# Set the years over which to loop.  Start year changes with season of prediction, so make
#generic.

dummy = 1948                                   # year before data starts.
start = 1949 - dummy                           # start of loop.
end = 2002 - dummy                             # end of loop.
nyr = end - start + 1                          # number of years to predict
                                               #  at each year.

ensemble = matrix(nrow=100, ncol=nyr)           # empty array for ensemble
                                               # forecasts.

#_____________________________________________________________________
# Now perform the Ensemble forecasting. Need to cross-validate
# the predictions with the observed data by dropping a year
# from the model and trying to predict it with the remaining
# years.
```

```
  for( i in 1:nyr)
   {
        # Loop through the years, dropping one at a time.
        Xp = X[i,]       # the predicted predictors

        if( i == start)
          {
           xmodel = X[(start+1):end, ]
           ymodel = Y[(start+1):end]
          }

        if( i == end)
          {
           xmodel = X[start:(end-1), ]
           ymodel = Y[start:(end-1)]
          }

        if( i != end && i != start)
          {
           xmodel = rbind(X[start:(i-1), ],X[(i+1):end, ])
           ymodel = c(Y[start:(i-1)], Y[(i+1):end])
          }

        ym = length(xmodel[,1])                    # length minus the one year



  # EUCLIDEAN DISTANCE: the distance from one year's predictors to
  # another year's predictors.

     YR = ymodel
        XR = scale(xmodel)
        distance = 1:ym

        for(j in 1:ym)
       {
#distance[j]=sqrt(((XX[i,1]-XR[j,1])^2) + ((XX[i,2]-XR[j,2])^2) + ((XX[i,3]-XR[j,3])^2))
distance[j] = sqrt(((XX[i,1]-XR[j,1])^2) + ((XX[i,2]-XR[j,2])^2))
        }

           drank = rank(distance)                  # rank the distances, 1 being
                                                   # the nearest neighbor.
```

```
# Search for the best alpha using first degree polynomial fit as
# the multiple linear regression model worked so well for the
# mean forecast.

n1 = 10/ym
bestalpha = seq(n1, 1.0, by = 0.05)
zalph = gcvplot(YR~xmodel, alpha=bestalpha, deg=1, kern="bisq", ev="data",
          scale = TRUE)
zzf = zalph$values
zxf = order(zzf)
bestalpha = bestalpha[zxf[1]]

# fit model using LOCFIT to get expected value.
fit = locfit(YR~xmodel, alpha=bestalpha, deg=1, kern="bisq", scale=TRUE, maxk=250)
resids = residuals(fit)                                    # get the residuals.

# Make the mean forecast prediction.
Xp = rbind(Xp, xmodel)
Ypred = predict.locfit(fit, Xp, se.fit=T, band="global")

# Determine K, the number of nearest neighbors to keep.
# Weight the nearest neighbors and pick the ones with the greatest
# weights at random 100 times to generate ensemble forecast.  Try different combinations
#  of neighbors around the neighborhood k = sqrt(m).
                m = length(distance)
                kk = sqrt(m)
                kk = round(kk)
                kk = kk+2
                W = 1:kk
                W = 1/W
                W = W/sum(W)
                W = cumsum(W)

                # The ensemble generation...
                        for(k in 1:100)
                          {
                          random = runif(1,0,1)          # random number generator
                                                         # between 0 & 1
                          xy = c(random, W)
                          rankW = rank(xy)
                          position = order(drank)[rankW[1]]
                          residpos = resids[position]

                        # add residuals to Yp to get ensemble.
                          ensemble[k,i] = Ypred$fit[1] + residpos

                           }                 # end ensemble loop
        }     # end loop over all years
```

```
#_____________________________________________________________________________
# What the ensemble predicted were the PC1s of the streamflow.  Now,
# the actual flows need to be reconstituted by multiplying by the
# respective Eof's and unscaled.  Then write the ensemble flows to
# file.

Eofs = matrix(scan("Up_AMJJ_Eofs.txt"), ncol=3, byrow=T)

# Regain the spatial information for each of the six sites:
# Kachess, Keechelus, Cle Elum, American, Rimrock, Bumping
# The extra factors are the scaling variables.  The
# multiplicative factor is the standard dev.  The additive
# factor is the mean for each site.

KACHguess = ((ensemble*Eofs[1,1])*35153.05) + 116015.87
KEEguess = ((ensemble*Eofs[2,1])*37994.65) + 128007.08
CLEguess = ((ensemble*Eofs[3,1])*109210.05) + 424452.85


write(t(KACHguess), file =
"USBR/Ensembles/Non_parametric/Upper/OND9_KACH_ensemble.txt", ncol = 54)
write(t(KEEguess), file =
"USBR/Ensembles/Non_parametric/Upper/OND9_KEE_ensemble.txt", ncol = 54)
write(t(CLEguess), file =
"USBR/Ensembles/Non_parametric/Upper/OND9_CLE_ensemble.txt", ncol = 54)



}          # end of program
```

LOWER SUBBASIN

```
{
library(locfit)

# This file generates the ensemble forecasts for the lower subbasin using the K Nearest
# Neighbor algorithm developed by Prairie (2002) and adapted by Grantz (2003).

# These files contain the predictor sets and PC1 for the upper subbasin.
# The general format of the data file is according to column:
# 1) Year
# 2) PC1
# 3) Canadian ridge
# 4) Aleutian low
# 5) pressure difference between the two.
# 6) SST
# 7) and 8) SWE

# This file is for the forecast using the USBR's dataset of
# naturalized streamflows.
OND = matrix(scan("USBR/LO_OND_predict.txt", skip=2), ncol=4, byrow=T)
#NDJ = matrix(scan("USBR/LO_NDJ_predict.txt", skip=2), ncol=6, byrow=T)
#DJF = matrix(scan("USBR/LO_DJF_predict.txt", skip =2), ncol = 8, byrow=T)
#JFM = matrix(scan("USBR/LO_JFM_predict.txt", skip=2), ncol=7, byrow=T)

# The predictor sets:
X = cbind(OND[,3], OND[,4])
#X = cbind(NDJ[,5], NDJ[,6])
#X = cbind(DJF[,5], DJF[,6], DJF[,7])                  # change predictors as needed.
#X = cbind(JFM[,5], JFM[,6]), JFM[,7])

                                                      #(Pres Diff, SST, M1 SWE)
XX = scale(X)                                         # Normalize the predictors
Y = OND[,2]                                           # Observed PC1s.


# Set the years over which to loop.  Start year changes
# with season of prediction, so make generic.

dummy = 1948                                          # year before data starts.
start = 1949 - dummy                                  # start of loop.
end = 2002 - dummy                                    # end of loop.
nyr = end - start + 1                                 # number of years to predict
                                                      #  at each year.

ensemble = matrix(nrow=100, ncol=nyr)                 # empty array for ensemble
                                                      # forecasts.



#_____________________________________________________________________
```

```
# Now perform the Ensemble forecasting. Need to cross-validate
# the predictions with the observed data by dropping a year
# from the model and trying to predict it with the remaining
# years.

  for( i in 1:nyr)
   {
        # Loop through the years, dropping one at a time.
        Xp = X[i,]                                          # the predicted predictors

        if( i == start)
          {
           xmodel = X[(start+1):end, ]
           ymodel = Y[(start+1):end]
          }

        if( i == end)
          {
           xmodel = X[start:(end-1), ]
           ymodel = Y[start:(end-1)]
          }

        if( i != end && i != start)
          {
           xmodel = rbind(X[start:(i-1), ],X[(i+1):end, ])
           ymodel = c(Y[start:(i-1)], Y[(i+1):end])
          }

        ym = length(xmodel[,1])                             # length minus the one year


   # EUCLIDEAN DISTANCE: the distance from one year's predictors to
   # another year's predictors.

      YR = ymodel
         XR = scale(xmodel)
         distance = 1:ym

         for(j in 1:ym)
        {
#distance[j]=sqrt(((XX[i,1]-XR[j,1])^2) + ((XX[i,2]-XR[j,2])^2) + ((XX[i,3]-XR[j,3])^2))
distance[j] = sqrt(((XX[i,1]-XR[j,1])^2) + ((XX[i,2]-XR[j,2])^2))
        }

           drank = rank(distance)                           # rank the distances, 1 being
                                                            # the nearest neighbor.



    # Search for the best alpha using first degree polynomial fit as
```

```
# the multiple linear regression model worked so well for the
# mean forecast.

n1 = 10/ym
bestalpha = seq(n1, 1.0, by = 0.05)
zalph = gcvplot(YR~xmodel, alpha=bestalpha, deg=1, kern="bisq", ev="data",
          scale=TRUE)
zzf = zalph$values
zxf = order(zzf)
bestalpha = bestalpha[zxf[1]]

# fit model using LOCFIT to get expected value.
fit = locfit(YR~xmodel, alpha=bestalpha, deg=1, kern="bisq", scale=TRUE, maxk=250)
resids = residuals(fit)                                    # get the residuals.

# Make the mean forecast prediction.
Xp = rbind(Xp, xmodel)
Ypred = predict.locfit(fit, Xp, se.fit=T, band="global")

# Determine K, the number of nearest neighbors to keep.
# Weight the nearest neighbors and pick the ones with the greatest
# weights at random 100 times to generate ensemble forecast.
                m = length(distance)
                kk = sqrt(m)
                kk = round(kk)
                kk = kk+2
                W = 1:kk
                W = 1/W
                W = W/sum(W)
                W = cumsum(W)

                # The ensemble generation...
                        for(k in 1:100)
                          {
                          random = runif(1,0,1)                # random number generator
                                                               # between 0 & 1
                          xy = c(random, W)
                          rankW = rank(xy)
                          position = order(drank)[rankW[1]]
                          residpos = resids[position]

                         # add residuals to Yp to get ensemble.
                          ensemble[k,i] = Ypred$fit[1] + residpos

                           }                 # end ensemble loop
        }    # end loop over all years
```

#_____

```
# What the ensemble predicted were the PC1s of the streamflow.  Now,
# the actual flows need to be reconsituted by multiplying by the
# respective Eof's and unscaled.  Then write the ensemble flows to
# file.

Eofs = matrix(scan("lo_AMJ_Eofs.txt"), ncol=3, byrow=T)

# Regain the spatial information for each of the six sites:
# Kachess, Keechelus, Cle Elum, American, Rimrock, Bumping
# The extra factors are the scaling variables.  The
# multiplicative factor is the standard dev.  The additive
# factor is the mean for each site.

AMguess = ((ensemble*Eofs[1,1])*33664) + 113291.0
BUMguess = ((ensemble*Eofs[2,1])*37463) + 129017
RIMguess = ((ensemble*Eofs[3,1])*59446) + 207426

write(t(AMguess), file =
"USBR/Ensembles/Non_parametric/Lower/OND9_AMER_ensemble.txt", ncol = 54)
write(t(BUMguess), file =
"USBR/Ensembles/Non_parametric/Lower/OND9_BUM_ensemble.txt", ncol = 54)
write(t(RIMguess), file =
"USBR/Ensembles/Non_parametric/Lower/OND9_RIM_ensemble.txt", ncol = 54)

}           # end of program

#_____________________________________________________________________________
# Plot the ensemble forecasts.
{
# This file generates the boxplots of the ensemble forecasts for each
# of the six basin sites.

# The reconstructed streamflow ensembles:
#KACH =
matrix(scan("USBR/Ensembles/Non_parametric/Kach/DJFA7_KACH_ensemble.txt"), ncol
= 54, byrow = T)

#KEE =
matrix(scan("USBR/Ensembles/Non_parametric/Upper/11_neighbors/DJFM/DJFM_KEE_en
semble.txt"), ncol = 54, byrow = T)

#CLE =
matrix(scan("USBR/Ensembles/Non_parametric/Upper/11_neighbors/DJFM/DJFM_CLE_en
semble.txt"), ncol = 54, byrow = T)

#AMR =
matrix(scan("USBR/Ensembles/Non_parametric/Lower/10_neighbors/DJFM/DJFM_AMER_
ensemble.txt"), ncol = 54, byrow = T)
#BUM =
matrix(scan("USBR/Ensembles/Non_parametric/Lower/10_neighbors/DJFM/DJFM_BUM_e
nsemble.txt"), ncol = 54, byrow = T)
```

```
RIM =
matrix(scan("USBR/Ensembles/Non_parametric/Lower/10_neighbors/DJFM/DJFM_RIM_en
semble.txt"), ncol = 54, byrow = T)

#KACH = KACH/10000
#KEE = KEE/10000
#CLE = CLE/10000
#AMR = AMR/10000
#BUM = BUM/10000
RIM = RIM/10000

# Get the actual flows for comparison.
real = matrix(scan("USBR/Total_Volume_AMJJ.txt", skip=2), ncol=6, byrow=T)
KAob = real[,1]/10000
KEob = real[,2]/10000
CEob = real[,3]/10000
AMob = real[,4]/10000
BUob = real[,5]/10000
RMob = real[,6]/10000

# PLOTS:
dummy = 1948
plotbeg = 1949 - dummy
plotend = 2002 - dummy

nevals = plotend - plotbeg + 1
xevals = RIM[ ,plotbeg:plotend]
xs = 1:nevals

quantf = quantile(RMob, c(0.05, 0.25, 0.5, 0.75, 0.95))
zz=boxplot(split(t(xevals),xs),plot=F,cex=1.0,print=F,cex=1.0,ylim=range(0,500))
zz$names= rep("",length(zz$names))
zz2 = bxp(zz, xlab="", ylab="", cex=1.25)

title(main="RIM: DJFM 1949-2002")
title(ylab= "AMJJ Volume (10000 ac-ft)")
lines(zz2, RMob[plotbeg:plotend], lty=1, lwd=2)

          for(i in 1:5)
            {
             abline(h = quantf[i], lty = 2)
            }

}
```

RPSS Code adapted from Grantz (2003)

```
{
# Rank Probability Skill Score RPSS

# The reconstructed streamflows:
#KACH =
matrix(scan("USBR/Ensembles/Non_parametric/Upper/OND/OND9_KACH_ensemble.txt"),
ncol = 54, byrow = T)

#KEE =
matrix(scan("USBR/Ensembles/Non_parametric/Upper/OND/OND9_KEE_ensemble.txt"),
ncol = 54, byrow = T)

#CLE =
matrix(scan("USBR/Ensembles/Non_parametric/Upper/OND/OND9_CLE_ensemble.txt"),
ncol = 54, byrow = T)

#AMR =
matrix(scan("USBR/Ensembles/Non_parametric/Lower/OND/OND9_AMER_ensemble.txt")
, ncol = 54, byrow = T)

#BUM =
matrix(scan("USBR/Ensembles/Non_parametric/Lower/OND/OND9_BUM_ensemble.txt"),
ncol = 54, byrow = T)

RIM =
matrix(scan("USBR/Ensembles/Non_parametric/Lower/OND/OND9_RIM_ensemble.txt"),
ncol = 54, byrow = T)


# Get the actual flows for comparison.
real = matrix(scan("USBR/Total_Volume_AMJJ.txt", skip=2), ncol=6, byrow=T)
KAob = real[,1]
KEob = real[,2]
CEob = real[,3]
AMob = real[,4]
BUob = real[,5]
RMob = real[,6]

Volume = RMob                                   # Change variable as needed


# hist is the vector of historical values
# obsval is the observed value in any given year
# il  is the upper limit for the lower category
# iu  is the lower limit for the upper category

# isim is the vector of simulated values
# ncat is the number of categories (we can set it to 3 for now)
```

```
dummy = 1948                                                    # year before data starts.
start = 1949 - dummy                                            # start of loop.
end = 2002 - dummy                                              # end of loop.
nyr = end - start + 1                                           # number of years in dataset

rpss = length(nyr)

## USBR skill-score

   for(j in 1:nyr)
     {
          isim = RIM[,j]
      obsval = RMob[j]

#set the divisions at the 20th and 80th percentile of historical data
# (minus the year we're in), to capture the more extreme years.
           if(j == start)
          xxf = Volume[(start+1):end]
           if(j == end)
          xxf = Volume[start:(end-1)]
           if(j>start && j<end)
          xxf = c(Volume[start:(i-1)], Volume[(i+1):end])

il = quantile(xxf,0.2)
iu = quantile(xxf,0.8)


ncat = 3
probs = 1:ncat
pclim = 1:ncat
obscat = 1:ncat

ncat1 = ncat-1

# assigns which category each of the simulations falls in
simcat = 1:length(isim)
nsim = length(isim)
simcat[isim >= iu] = 3
simcat[isim <= il] = 1
simcat[isim > il & isim <iu] = 2

# Probability of simulations falling into each category
           for(i in 1:ncat)
             probs[i] = length(simcat[simcat == i])/nsim
             probs = cumsum(probs)

# raw probability of falling into each category
# 33% chance to fall in each one of 3 categories
           pclim = rep(1,ncat)
           pclim = pclim/ncat
```

```
        pclim = cumsum(pclim)

# where the observation actually falls
        obscat = rep(0,ncat)
          if(obsval >= iu)
            obscat[3] = 1
          if(obsval > il & obsval < iu)
            obscat[2] = 1
          if(obsval <= il)
            obscat[1] = 1
            obscat = cumsum(obscat)

# The RPSS:
    rpss[j] = 1 - ( sum((obscat-probs)^2) / sum((obscat-pclim)^2) )
        }          # end of for loop


# Plot RPSS for all years, wet years (80th%),
# dry years (20th%).  These percentiles are chosen arbitrarily.

par(mfrow = c(1,3))
#   ALL YEARS:
boxplot(rpss, xlab="RIM: All Years", ylim=c(-2,1))


 #  WET YEARS:
qup = quantile(xxf, 0.75)
wet = rpss[xxf > qup]
boxplot(wet, xlab="RIM: Wet Years", ylim=c(-2,1))


#   DRY YEARS:
qlo = quantile(xxf, 0.25)
dry = rpss[xxf > qlo]
boxplot(dry, xlab="RIM: Dry Years", ylim=c(-2,1))


}          # end of program
```

Likelihood Skill Score code, adapted from Grantz (2003)

```
{
# Likelihood Skill Score

# The reconstructed streamflow ensembles:
#KACH =
matrix(scan("USBR/Ensembles/Non_parametric/Kach/DJFA7_KACH_ensemble.txt"), ncol
= 54, byrow = T)

#KEE =
matrix(scan("USBR/Ensembles/Non_parametric/Upper/7_neighbors/DJFM/DJFM_KEE_ens
emble.txt"), ncol = 54, byrow = T)

#CLE =
matrix(scan("USBR/Ensembles/Non_parametric/Upper/7_neighbors/DJFM/DJFM_CLE_ens
emble.txt"), ncol = 54, byrow = T)

#AMR =
matrix(scan("USBR/Ensembles/Non_parametric/Lower/7_neighbors/NDJF/NDJF_AMER_e
nsemble.txt"), ncol = 54, byrow = T)

#BUM =
matrix(scan("USBR/Ensembles/Non_parametric/Lower/7_neighbors/NDJF/NDJF_BUM_ens
emble.txt"), ncol = 54, byrow = T)

RIM =
matrix(scan("USBR/Ensembles/Non_parametric/Lower/7_neighbors/NDJF/NDJF_RIM_ense
mble.txt"), ncol = 54, byrow = T)

# Get the actual flows for comparison.
real = matrix(scan("USBR/Total_Volume_AMJJ.txt", skip=2), ncol=6, byrow=T)
KAob = real[,1]
KEob = real[,2]
CEob = real[,3]
AMob = real[,4]
BUob = real[,5]
RMob = real[,6]

Volume = RMob                                   # Change variable as needed


# hist is the vector of historical values
# obsval is the observed value in any given year
# il  is the upper limit for the lower category
# iu  is the lower limit for the upper category
```

```
# isim is the vector of simulated values
# ncat is the number of categories (we can set it to 3 for now)

dummy = 1948                                    # year before data starts.
start = 1949 - dummy                            # start of loop.
end = 2002 - dummy                              # end of loop.
nyr = end - start + 1                           # number of years in dataset

likeskill = length(nyr)

## USBR skill-score

   for(j in 1:nyr)
     {
          isim = RIM[,j]
       obsval = RMob[j]

#set the divisions at the 20th and 80th percentile of historical data
# (minus the year we're in)
           if(j == start)
         xxf = Volume[(start+1):end]
           if(j == end)
         xxf = Volume[start:(end-1)]
           if(j>start && j<end)
         xxf = c(Volume[start:(i-1)], Volume[(i+1):end])

il = quantile(xxf,0.20)
iu = quantile(xxf,0.80)


ncat = 3
probs = 1:ncat
pclim = 1:ncat
obscat = 1

# assigns which category each of the simulations falls in
simcat = 1:length(isim)
nsim = length(isim)
simcat[isim >= iu] = 3
simcat[isim <= il] = 1
simcat[isim > il & isim <iu] = 2

# Probability of simulations falling into each category
           for(i in 1:ncat)
             probs[i] = length(simcat[simcat == i])/nsim

# raw probability of falling into each category
# 33% chance to fall in each one of 3 categories
           pclim = rep(1,ncat)
           pclim = pclim/ncat
```

```
# where the observation actually falls
        if(obsval >= iu)
              obscat = 3
            if(obsval > il & obsval < iu)
              obscat = 2
            if(obsval <= il)
              obscat = 1

#  The Loglikihood skill
          logskill[j] = probs[obscat]/pclim[obscat]

          }  # End of for loop.

# Plot Logskill for all years, wet years (80th%),
# dry years (20th%).  These percentiles are chosen arbitrarily.

par(mfrow = c(1,3))
#   ALL YEARS:
          boxplot(logskill, xlab="RIM: All Years", ylim=c(0,3))

#   WET YEARS:
          qup = quantile(xxf, 0.80)
          wet = logskill[xxf > qup]
          boxplot(wet, xlab="RIM: Wet Years", ylim=c(0,3))

#   DRY YEARS:
          qlo = quantile(xxf, 0.20)
          dry = logskill[xxf > qlo]
          boxplot(dry, xlab="RIM: Dry Years", ylim=c(0,3))


}         # end of program
```

References

References arranged by chapter. References used in multiple chapters are only noted once under the chapter in which they are first used.

Chapter 1

Getches, D.H., 1997: Water Law in a Nutshell, 3rd Ed., West Pub. Co., MN.

Matsin, M.C. and Vaccaro, J.J., 2002: "Watershed Models for Decision Support in the Yakima River Basin, Washington," USGS Openfile Report 02-404, http://water.usgs.gov/pubs/of/2002/ofr02404/.

The Systems Operations Advisory Committee (SOAC), 1999: "Report on Biologically Based Flows for the Yakima River Basin,".

Tri-County Water Resource Agency, 2003: "Watershed Assessment Yakima River Basin Plan," www.co.yakima.wa/us/tricnty/watershedplan.htm.

U.S. Bureau of Reclamation, 2004: "Yakima Project Washington", http://www.usbr.gov/dataweb/html/yakima.html

U.S. Bureau of Reclamation, "Interim Operating Plan," Nov. 2002, CD-ROM.

U.S. Environmental Protection Agency, 2004: "Endangered Species Act Status of West Coast Salmon and Steelhead," www.nwr.noaa.gov/1salmon/salmesa/

U.S. Environmental Protection Agency, 1998: "Columbia River Basin Fish Contaminant Survey," yosemite.epa.gov/r10/oea.nsf/0/C3A9164ED269353788256C09005D36B7?OpenDocument

Walton, S., 2000: The Complete Guide to Wine, pp. 211-212, Hermes House, New York

Washington State Department of Ecology, 2004: Yakima Basin Water Quality Homepage, www.ecy.wa.gov/programs/wq/tmdl/watershed/Yakima_wq/intro.html

"WA Dept. of Ecology v. Yakima Reservation Irrigation District," April 1993, www.mrsc.org/mc/courts/supreme/121wn2d/121wn2d0257.htm

WA Dept. of Fish and Wildlife, 1998: "1998 Salmonid Stock Inventory Bull Trout/ Dolly Varden Volume: Columbia River," wdfw.wa.gov/fish/sassi/bulldolly.htm

Yakima Subbasin Fish and Wildlife Planning Board, 2004: "Yakima River Subbasin Plan," 200-218, www.co.yakima.wa.us/Yaksubbain/default.htm

Chapter 2

Aydin, K.Y., 2004: "ECOPATH Modeling and the Alaskan Gyre," School of Fisheries, Univ. of WA, www.fisheries,ubc.ca/Projects/PWSound/AlaskaEco/GYRE/AYDIN1.htm

Cayan, D.R., 1996: "Internannual Climate Variability and Snowpack in the Western United States," Jn. Clim., Amer. Met. Soc.

Clark, M.P., Serezze, M.C., and McCabe, G.J., 2001: "Historical Effects of El Niño and La Niña events on the seasonal evolution of the montane snowpack in the Columbia and Colorado River Basins," Water Res. Res., 37 (3)

Climate Prediction Center, National Weather Service, 2004: "Pacific/North American (PNA)," www.cpc.noaa.gov/data/teledoc/pna.html

_____, "The ENSO Cycle," www.cpc.noaa.gov/products/analysis_monitering/ensocycle/enso_cycle.html, 2004.

Gershunov, A., and Barnett, T.P., 1996: "Interdecadal Modulation of ENSO Teleconnections," Bull. Amer. Met. Soc.

Hamlet, A.F., and Lettenmaier, D.P., 1999: "Columbia River Streamflow Forecasting Based on ENSO and PDO Climate Signals," ASCE. Jn. Of Water Res. Assoc., 35 (6)

Helsel, D.R., and Hirsch, R.M., 1995: Statistical Methods in Water Resources. Amsterdam: Elesevier Science B.V.

Hoerling, M.P., Kumar, A., and Zhong, M., 1997: "El Niño, La Niña, and the Nonlinearity of Their Teleconnections," Jn. Clim., Amer. Met. Soc.

Kalnay, E. and Coauthors, 1996: "The NCEP/NCAR Reanalysis 40-year Project," Bull. Amer. Met. Soc., 77

Latif, M., and Barnett, T.P., 1996: "Decadal Climate Variability over the North Pacific and North America: Dynamics and Predictability," Jn. Clim., Amer. Met. Soc.

Lowry, R., 2002: Concept and Applications of Inferential Statstics, Chpts. 2 and 3, vassun.vassar.edu/~lowry/webtext.html

Mantua, N., 2004: "PDO Index", www.jisao.washington.edu/pdo/PDO.latest

Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M, and Francis, R.C., 1997: "A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production," Bull. Amer. Met. Soc., 78

McCabe, G.J., and Dettinger, M.D., 1999: "Decadal Variations in the Strength of ENSO Teleconnections with Precipitation in the Western United States," Int. Jn. Climat., 19

McCabe, G.J, and Dettinger, M.D., 2002: "Primary Modes and Predictability of Year-to-Year Snowpack Variations in the Western United States from Teleconnections with Pacific Ocean Climate," Jn. Hydrom.

Natural Resources Conservation Service, U.S. Dept. of Agriculture, "Snow Course and Monthly SNOTEL Data Tables," www.wcc.nrcs.usda.gov/cgibin/state-site.pl?state=WA&report=snowcourse, 2004.

Polovina, J.J., Mitchum, G.T., and Evans, G.T., 1995: "Decadal and basin-scale variation in mixed layer depth and the impact on biological production in the Central and North Pacific, 1960-1988," Deep-Sea Res.

Slack, J.R., and Landwehr, J.M., 2002: "Hydro-Climatic Data Network: A U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1874-1988," USGS Openfile Report 92-129, www.pubs.usgs.gov/of/hcdn_report/content.html

U.S. Bureau of Reclamation, Hydromet Database, www.usbr.gov/pn/hydromet/yakima/yakwebarcread.html, 2004.

Wallace, J.M, and Gutzler, D.S., 1980: "Teleconnections in the Geopotential Height Field during the Northern Hemispheric Winter," Mon. Weather Rev., Amer. Met. Soc.

Zhang, Y., Wallace, J.M., and Battisti, D.S., 1997: "ENSO-like Interdecadal Variability," Jn. Clim., Amer. Met. Soc.

Chapter 3

Baoli, L., Shiwen, Y., and Qin, L., 2003: "An Improved k-Nearest-Neighbor Algorithm for Text Categorization," Proc. 20th Int. Conf. On Comp. Proc. Orient. Lang.

Basic methods in Theoretical Biology, Ed. Kooijman, S.A.L.M, Ferreira J., Kooi, B.W., Zonneveld, 1999: C., Dept. Theoretical Biology, Vrije Universiteit, Amsterdam, http://www.bio.vu.nl/thb/course/tb/

ENSO Index according to JMA SST (1868-present), Center for Ocean-Atmospheric Prediction Studies (COAPS), http://elviejo.coaps.fsu.edu/products/jma_index.php, 2004.

Garen, D.C., 1992: "Improved Techniques in Regression-Based Streamflow Volume Forecasting," Jn. Water Res. Plan. And Mang., 118 (6)

Grantz, K.A., 2003: Using Large-Scale Climate Information to Forecast Seasonal Streamflow in the Truckee and Carson Rivers, Diss. U of Colorado

Grantz, K.A., Rajagopalan, B., Clark, M., and Zagona, E., 2004: "A Technique for Incorporating Large-Scale Climate Information in Basin-Scale Ensemble Streamflow Forecasts", in review.

Kumar, A., Barnston, A.G., and Hoerling, M.P., 2001: "Seasonal Predictions, Probabilistic Verifications, and Ensemble Size," Jn. Clim., 14

Lea A. and Saunders, M., 2004: "Verification of Authors' Seasonal Forecast for Winter 2003/04 NAO," forecast.mssl.ucl.ac.uk/docs/NAO2003=4Verification.pdf

Loader, C.R., 1999: “Locfit: An Introduction,” cm.bell-labs.com/stat/project/locfit

Lynch, C., U.S. Bureau of Reclamation, email communication with the author on 09-09-03 and 09-24-03.

Murphy, A.H., 1972: “Scalar and Vector Partitions of the Ranked Probability Score,” Mont. Weather Rev., 100 (10)

Perry, K.M. and Stanley, J.R., 1995: “Generalized Cross-Validation as a Stopping Rule for the Richardson-Lucy Algorithm,” Int. Jn. Imag. Sys. and Tech., 6, or available online: www.stsci.edu/stsci/meetings/irw/proceedings/perryk.dir/perryk.html.

Prairie, J.R., 2002: Long-Term Salinity Prediction with Uncertainty Analysis: Application for Colorado River above Glenwood Springs, CO, Diss. U of Colorado

Rajagopalan, B. and Lall, U., 1999: “A k-nearest-neighbor simulator for daily precipitation and other weather variables,” Water Res. Res., 35 (10)

Sharma, A., Tarboton, D.G., and Lall, U., 1997: “Streamflow Simulation: A nonparametric approach,” Water Res. Res., 33 (2)

Sharp, W., U.S. Bureau of Reclamation, phone conference with Edith Zagona and author, 01-14-04.

Singhrattna, N., Rajagopalan, B., Clark, M., and Kumar, K.K., 2004: “Forecasting Thailand Summer Monsoon Rainfall,” Int. Jn. Clim.

Weisstein, E.W., “Likelihood,” From MathWorld—A Wolfram Web Resource, http://mathworld.wolfram.com/Likelihood.html.

Chapter 4

Economic and Engineering Services Inc., 2002: “Technical Memorandum Yakima River Basin Watershed Plan: Voluntary Water Transfers as a Strategy for Meeting Planning Objectives”

Lynch, C., personal communication with author via email and fax, 06-15-04.

Scott, M.J., Vail, L.W., Jaksch, J., Stöckle, C.O., and Kemanian, A., 2004: “Water Exchanges: Tools to Beat El Niño Climate Variability in Irrigated Agriculture,” Jn. Amer. Water Res. Assc.

U.S. Bureau of Reclamation, “News Release April 10, 2001: Yakima Basin April Water Supply Forecast,” www.usbr.gov/pn/01new/yakapril.html, 2001.

Zagona, E.A., Fulp, T.J., Shane, R., Magee, T., Goranflo, M.H., 2001: “Riverware: A Generalized Tool for Complex Reservoir System Modeling,” Jn. Amer. Water Res. Assc., 37 (4)

Chapter 5

Anderson, J.J., 2003: "An analysis of smolt survival with implications to flow management", *to be submitted for publication*

Anderson, J. J., 1996: "Decadal climate cycles and declining Columbia River salmon," Proceedings of the Sustainable Fisheries Conference Victoria B.C. Canada, Eric Knudsen Editor. Special publication of the American Fisheries Society

Cayan, D.R., 1996: "Interannual Climate Variability and Snowpack in the Western United States," Jn. Clim., (9)

Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, "Climate Change 2001, The Scientific Basis: Technical Summary," Cambridge Univ. Press, 2001.

Francis, R.C., S.R. Hare, A.B. Hollowed, and Wooster, W.S., "Effects of interdecadal climate variability on the oceanic ecosystems of the Northeast Pacific," Fish. Oceanogr. 7: 1-21, 1998.

Finney, B.P., Gregory-Eaves, I., Sweetman, J., Douglas, M.S.V., and Smol, J.P., 2000: "Impacts of Climate Change and Fishing on Pacific Salmon Abundance Over the Past 300 Years," Science, 290 (5492)

Hare, S.R., Mantua, N.J., and Francis, R.C., 1999: "Inverse Production Regimes: Alaska and West Coast Pacific Salmon," Fisheries, 24 (1)

Klaus, W., "Multivariate ENSO Index," www.cdc.noaa.gov/ENSO/enso.mei_index.html, 2004.

Lluch-Cota, D.B., Wooster, W.S., Hare, S.R., Lluch-Belda, D., and Parés-Sierra, A., 2003: "Principal Modes and Related Frequencies of Sea Surface Temperature Variability in the Pacific Coast of North America," Jn. Ocean., 59

Miles, E.L., Snover, A.K., Hamlet, A.F., Callahan, B., and Fluharty, D., 2000: "Pacific Northwest Regional Assessment: The impacts of climate change variability and climate change on the water resources of the Columbia River Basin," Jn. Amer. Water Res. Assc., 36 (2)

Mote, P.W., 2003: "Trends in snow water equivalent in the Pacific Northwest and their climatic causes," Geo. Res. Ltrs., 30 (12)

Natural Resources Defense Council, "America's Animal Factories, How States Fail to Prevent Pollution from Livestock Waste", Ch. 28, http://www.nrdc.org/water/pollution/factor/aafinx.asp, 1998.

Smith, J.J., "Winter Steelhead and Chinook and Coho Salmon Life Cycles and Habitat Requirements," Dept. Biological Sciences, San Jose State Univ.

United Nations Framework Convention on Climate Change, “Article 1 Definitions,” http://unfccc.int/resource/conv/conv_003.html.

U.S. Environmental Protection Agency, Columbia Basin Tribes, and the States of Idaho, Oregon, and Washington, “Columbia/Snake Rivers Temperature TMDL” DRAFT, 11/2002.

U.S. Global Change Research Program, “US National Assessment of the Potential Consequences of Climate Variability and Change: Chpt. 9 Pacific Northwest,” Ed. Parson, E.A., 2003.

Washington State Department of Ecology: Yakima Basin Water Quality Homepage, www.ecy.wa.gov/programs/wq/tmdl/watershed/Yakima_wq/intro.html, 2004.

Washington State Department of Ecology, "A Suspended Sediment and DDT Maximum Daily Load Evaluation Report for the Yakima River", p. 4, 1997. Accessible through above link.