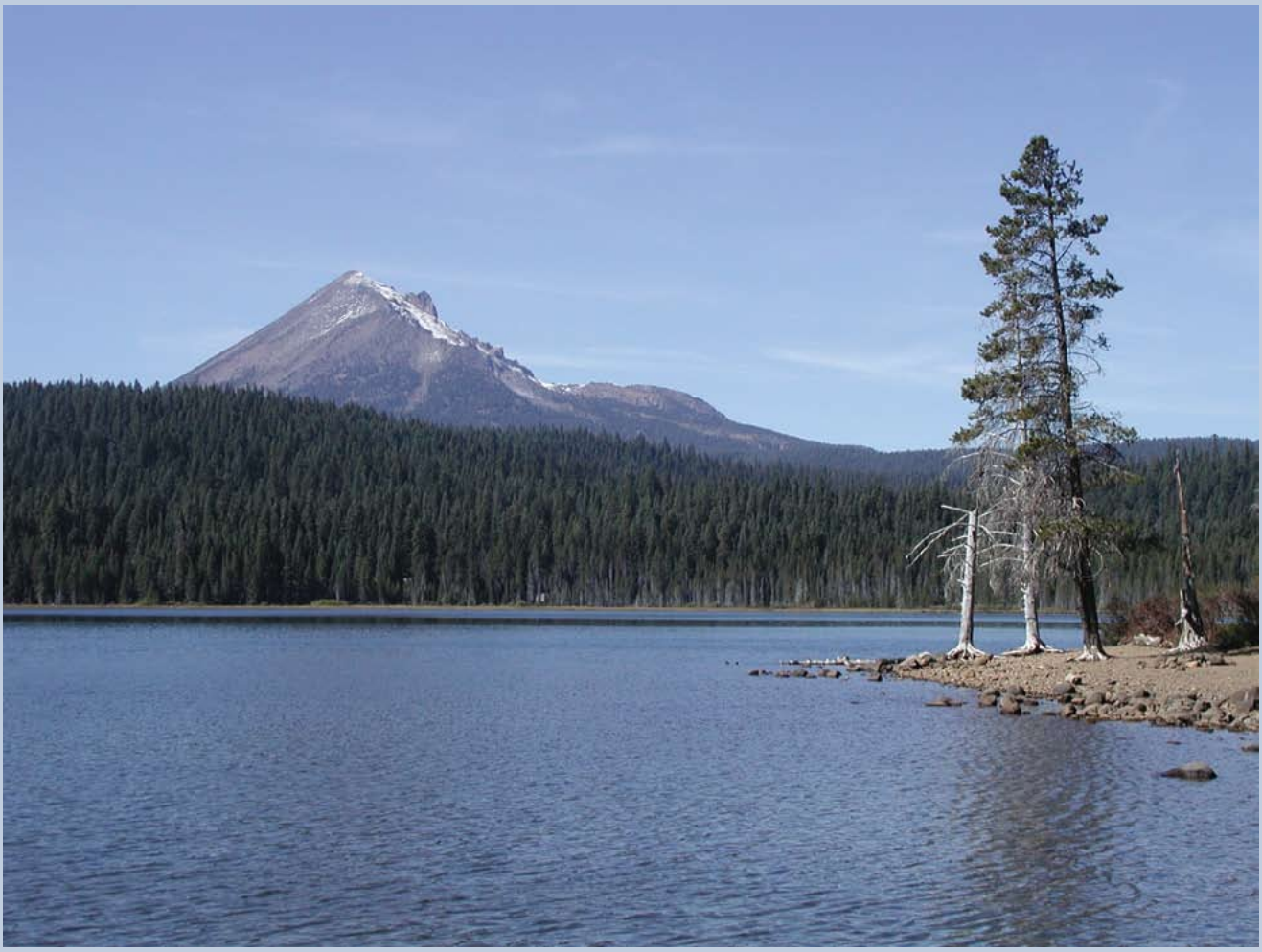
USGS
science for a changing world

In cooperation with the
Bureau of Reclamation and Natural Resources Conservation Service

An Analysis of Statistical Methods for Seasonal Flow Forecasting in the Upper Klamath River Basin of Oregon and California



Scientific Investigations Report 2005–5177

U.S. Department of the Interior
U.S. Geological Survey

Cover photograph:
Lake of the Woods with Mt. McLaughlin in the background (Photograph by Marshall W. Gannett, U.S. Geological Survey).

An Analysis of Statistical Methods for Seasonal Flow Forecasting in the Upper Klamath River Basin of Oregon and California

By John C. Risley and Marshall W. Gannett, U.S. Geological Survey,
Jolyne K. Lea, Natural Resources Conservation Service, and
Edwin A. Roehl Jr., Advanced Data Mining Services, LLC

In cooperation with the
Bureau of Reclamation and Natural Resources Conservation Service

Scientific Investigations Report 2005–5177

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
Gale A. Norton, Secretary

U.S. Geological Survey
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia: 2005

For product and ordering information:
World Wide Web: *http://www.usgs.gov/pubprod*
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:
World Wide Web: *http://www.usgs.gov*
Telephone: 1-888-ASK-USGS

Suggested citation:
Risley, J.C., Gannett, M.W., Lea, J.K., and Roehl, E.A., Jr., 2001, An analysis of statistical methods for seasonal flow forecasting in the Upper Klamath River Basin of Oregon and California: U.S. Geological Survey Scientific Investigations Report 2005–5177, 44 p.

Contents

Figures

Tables

Conversion Factors and Vertical Datum

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter cm)
foot (ft)	0.3048	meter (m)
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
acre-foot (acre-ft)	1,233	cubic meter (m^3)

Vertical coordinate information is referenced to the National Geodetic Vertical Datum of 1929 (NGVD 29).

An Analysis of Statistical Methods for Seasonal Flow Forecasting in the Upper Klamath River Basin of Oregon and California

By John C. Risley, Marshall W. Gannett, Jolyne K. Lea, and Edwin A. Roehl

Abstract

Water managers in the upper Klamath Basin, located in south-central Oregon and northeastern California, use forecasts of spring and summer streamflow to optimally allocate increasingly limited water supplies for various demands that include irrigation for agriculture, habitat for endangered fishes, and hydropower production. Flow forecasts are made by the Natural Resources Conservation Service using statistical models that use current snow and precipitation data collected at nearby monitoring sites as input. The forecasts for five upper Klamath Basin sites (Williamson River, Sprague River, Upper Klamath Lake, Gerber Reservoir, and Clear Lake Reservoir) are made at the beginning of each month from January through June.

In 2003, the U.S. Geological Survey, the Natural Resources Conservation Service, and the Bureau of Reclamation began a collaborative study to reduce uncertainty and error in seasonal flow forecasting in the upper Klamath Basin. The main objectives included (1) evaluating nonregression statistical modeling approaches, such as artificial neural networks, for their efficacy in reducing model error, (2) finding and evaluating potential model variables that better described long-term climate-trend conditions, and (3) analyzing the efficacy of upper Klamath Basin snow-water equivalent and precipitation data in forecast models.

The modeling approaches evaluated included principal components regression, nonautoregressive artificial neural networks, and autoregressive artificial neural networks. For the Upper Klamath Lake forecast site, the nonautoregressive artificial neural network models had lower error than the other models for the January, February, and March forecasts. However, the principal components regression model performed better for the April forecast. Both models performed roughly the same for the May and June forecasts. For the Sprague River forecast site, the nonautoregressive artificial neural network models performed far better than the other models for the January, February, March, and June forecasts. However, the principal components regression models performed better for the April and May forecasts. For the Williamson River and Gerber Reservoir forecast sites, the principal components regression models generally, but not always, performed better than the other models. For the Clear Lake Reservoir forecast site, the nonautoregressive artificial neural network models performed far better than the other models for the months of January, February, and March. However, the Clear Lake Reservoir autoregressive artificial neural network model performed better than the other models for the month of April.

For the Williamson River, Upper Klamath Lake, and Gerber Reservoir forecast models, the inclusion of new long-term climate-trend variables reduced model error in many, but not all, instances.

The relationships between the upper Klamath Basin snow-water equivalent, precipitation, and flow data were analyzed to determine the ability and the extent to which current snow-water equivalent and precipitation conditions can be used to forecast future flow conditions. The analyses were made by decomposing the flow time series into annual periodic, long-term climatic, and chaotic components, and then lag correlating the snow-water equivalent and precipitation time series with the chaotic component time series. After 120 days (approximately 4 months), all of the snow-water equivalent and precipitation correlation coefficients were less than 0.4.

Introduction

Most irrigated land in the upper Klamath Basin, located in a semiarid region of southern Oregon and northern California, is supplied water by the Bureau of Reclamation (BOR). Established in 1905, the BOR Klamath Project currently operates three lake regulation dams (Link River, Clear Lake, and Gerber) and includes approximately 240,000 acres of irrigable land, in addition to National Wildlife Refuge land. Although the actual number of irrigated acres varies each year, the project has generally provided water annually to 200,000 acres per year. Water delivered to most of the irrigated land is stored in Upper Klamath Lake. Although the lake is natural, its stage has been regulated since 1921 by the Link River Dam. During the spring and summer, BOR management decisions

depend on balancing the amount of water (1) passed over the Link River Dam, the outlet of Upper Klamath Lake, for threatened Chinook and coho salmon in the lower Klamath River, (2) delivered from the lake through the A-Canal to irrigate about 185,000 acres, (3) retained in Upper Klamath Lake to protect water quality and habitat for two endangered sucker species, (4) used for hydroelectric power production, and (5) needed for wildlife refuges. Accurate forecasts of spring and summer streamflow are essential for optimally managing these limited water supplies. From January to June each year, water supply forecasts for five upper Klamath Basin sites are provided by the United States Department of Agriculture, Natural Resources Conservation Service (NRCS) on the first of each month, with biweekly updates as required. These sites include Clear Lake Reservoir, Upper Klamath Lake, Gerber Reservoir, Williamson River near Chiloquin, and Sprague River near Chiloquin (*fig. 1*). In addition to these upper Klamath Basin sites, the NRCS also provides similar forecasts for hundreds of other locations throughout the Western States. Typically, a total volume of flow (in thousand acre-feet) is forecasted at a site for two different 6-month periods after the date of the forecast. The 6-month periods that are often forecasted include February through July and April through September. The forecasted flow volumes are determined from a combination of statistical model results, data analyses, and coordination with National Weather Service (NWS) River Forecast Offices. Prior to this study, input variables to the NRCS upper Klamath Basin statistical models have generally included just current and recent monthly snowpack depths (measured as snow-water equivalent [SWE]) and precipitation measured at locations near the forecast sites.

Background

In water year 2001, the upper Klamath Basin experienced one of the worst droughts on record. In January and February 2001, basinwide snowpack and cumulative precipitation were only 50 percent of average. The January median forecasts (50 percent exceedance probability) for the five sites ranged from 54 to 70 percent of average. By April, the median forecasts dropped to 6 to 39 percent of average because of the continued lack of precipitation during the early spring. Typically, by early April and prior to planting crops, farmers want to know if and how much irrigation water will be delivered to them during the coming summer. Because of the combination of well-below-average volume forecasts and legal obligations under the Endangered Species Act to protect suckers and Chinook and coho salmon, the BOR ceased irrigation water deliveries to 75 percent of the Klamath Project (approximately 150,000 acres) in early April 2001.

The median forecasted April–September inflow to Upper Klamath Lake on April 1, 2001, was 160 thousand acre-feet (with a 40 percent confidence interval of 108 to 212 thousand acre-feet and a 90 percent confidence interval of 5 to 328 thousand acre-feet). The actual inflow for that period was 231 thousand acre-feet. Although well within the realm of possibility based on prediction confidence limits, it exceeded the median forecast by 71 thousand acre-feet and exceeded the 70 percent exceedance probability forecast (used by BOR for planning) by 123 thousand acre-feet. Towards the end of July, after realizing there was an additional volume of water available in the lake, the U.S. Department of the Interior and BOR released approximately 70 thousand acre-feet of water for irrigation within the BOR Project. However, at that point many farmers had already incurred financial losses.

In 2003, the USGS, NRCS, and BOR began a collaborative study to determine how some of the uncertainty and errors in seasonal flow forecasting in the upper Klamath Basin could be reduced. Uncertainty will exist in any seasonal flow forecasting method because of limitations in predicting the weather several months into the future, data network deficiencies, and processes not well represented in forecast models. In this study, we sought to reduce forecast uncertainty through model improvement. These efforts included improving the representation of long-term climate and interannual hydrologic processes in the models and evaluating the efficacy of alternative statistical techniques.

Flow forecasts in the upper Klamath Basin have larger uncertainty than other basins in the West. For example, the standard error of the April 1st forecast for the inflow to Upper Klamath Lake is about 20 percent of the long-term mean inflow. The standard error of other NRCS flow forecast sites in the West, such as in the northern Rocky Mountains, is about 10 percent. Part of the large uncertainty in the upper Klamath Basin may be due to the effects of regional ground-water flow on the hydrologic system.

Much of the upper Klamath Basin is underlain by permeable late Tertiary to Quaternary volcanic deposits. The youngest and most permeable deposits are in the Cascade Volcanic Arc, which is also the region of largest precipitation in the basin. As a result, a large proportion of precipitation in the Cascade Range infiltrates to the regional ground-water system, and many of the major streams in the basin have a large component of regional ground water. Because of its scale, the regional ground-water system tends to integrate climate conditions over multiple years, with the effects of anomalously dry or wet periods persisting in ground-water discharge for 2 or 3 years after conditions change.

These regional ground-water effects could partly explain the underforecasted inflows to Upper Klamath Lake during 2001. SWE and precipitation in water years 2000 and 2001 were below long-term averages. Water year 1999, in contrast, was an extremely wet year, and SWE was measured at record or near-record levels at many sites in the basin (*fig. 2*). In both 2000 and 2001, flow forecasts underpredicted the actual flows. The underprediction may have stemmed from the fact that the forecasting models had no input variables containing information about climatic conditions in previous years and could not account for the larger than average baseflow still resulting from the anomalously wet 1999 water year.

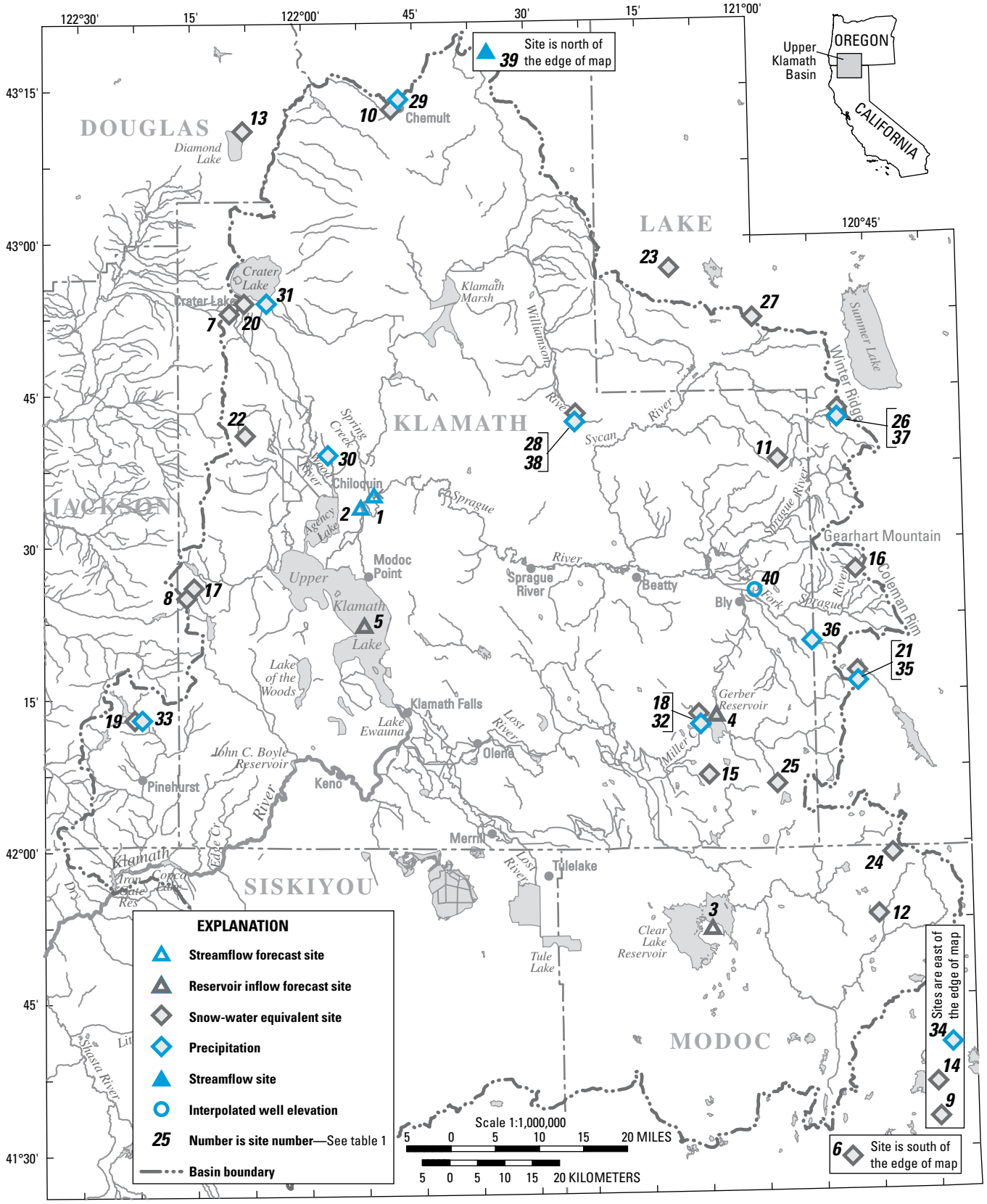


Figure 1. Upper Klamath Basin study area.

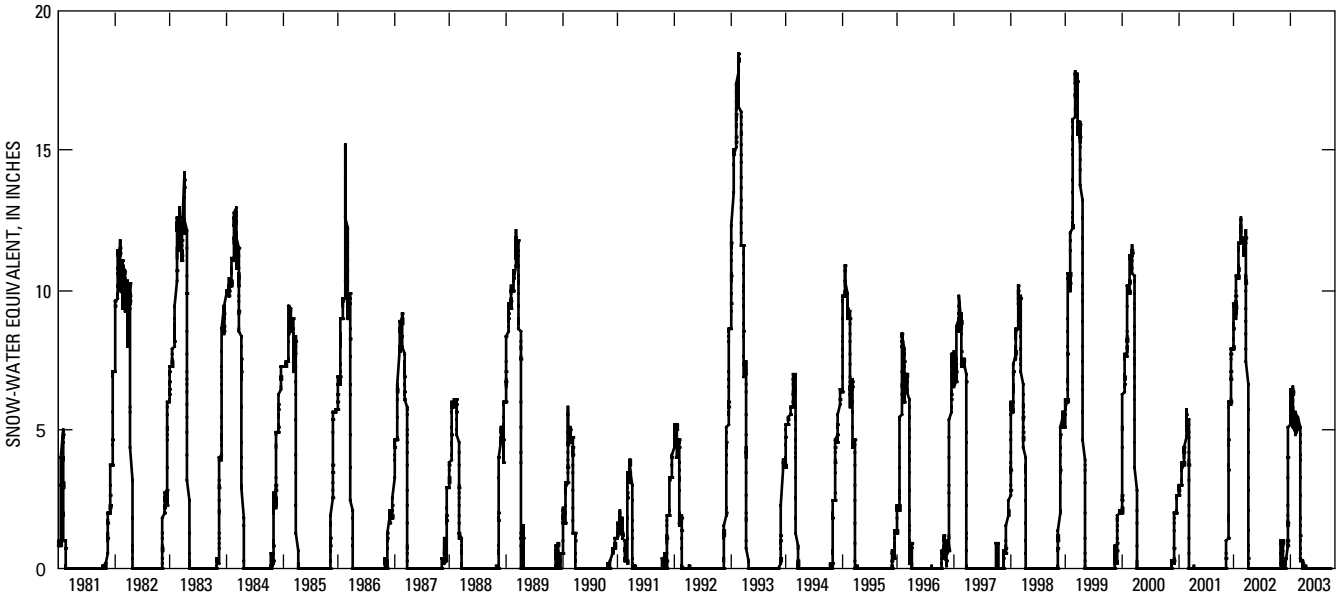


Figure 2. Daily snow-water equivalent at Chemult, Oregon, 1981–2003.

Flow forecasts in the upper Klamath Basin could be improved if the models contain new input variables that represent antecedent conditions from the preceding several years. Antecedent conditions could be represented by variables that indicate the state of the regional ground-water system, such as water-level time series from long-term observation wells or flow data from ground-water dominated streams. Another useful variable that might reflect interannual flow conditions would be precipitation, SWE, or streamflow volumes from the previous 1 or 2 years. Lastly, interannual climate conditions might be represented by certain ocean-climate indicators such as the Pacific Decadal Oscillation (PDO) or the El Niño Southern Oscillation.

Another possible avenue for improving forecast models is the application of statistical techniques that are an alternative to multiple-regression models. These methods, available and gaining wider acceptance in many hydrologic applications, include artificial neural networks (ANN), genetic algorithms, multivariate adaptive regression splines, and partial least squares. Like multiple regression, these other methods also use time series data for input and output variables. Unlike multiple-regression models, which assume a linear relationship between variables, these other methods are capable of efficiently modeling nonlinear processes that typically occur in natural systems.

Purpose and Scope

This report presents:

1. A comparison of three different modeling approaches to forecast spring and summer flows at five upper Klamath Basin forecast sites.
2. An evaluation of long-term climate-trend variables, used in some of the models, to determine their efficacy in reducing forecast model error.
3. An analysis of the limitations of upper Klamath Basin SWE and precipitation data in forecast models.

Study Area

The upper Klamath Basin (*fig. 1*) is located in south-central Oregon and northeastern California. The Klamath River originates at the southern end of the Upper Klamath Lake and is a tributary to the Pacific Ocean in northern California. The study area, which includes the drainage basin upstream of Iron Gate Dam, California and the closed Crater Lake, Lost River, and Butte Creek Basins, encompasses approximately 8,055 square miles. Elevations range from 2,162 feet at Iron Gate Dam, California to over 9,000 feet in the Oregon Cascade Range. The Williamson River Basin (including the Sprague River Basin) is approximately 3,000 square miles and encompasses 79 percent of the total drainage area that contributes to Upper Klamath Lake. Together, the Williamson and Sprague Rivers supply about one-half of the inflow to Upper Klamath Lake. The Lost River Basin, located south of the upper Klamath Lake Basin, includes Gerber Reservoir, Clear Lake Reservoir, Tule Lake, and Lower Klamath Lake drainages and has an area of approximately 3,000 square miles.

Most of the upper Klamath Basin is located on the western fringe of the Basin and Range geologic province (Dicken and Dicken, 1985), a region characterized by strong relief.

Extensive, broad, flat, poorly drained uplands, valleys, and marshlands are located throughout the province. The northern, eastern, and southern boundaries of the upper Klamath Basin are formed by inactive volcanoes, rims, scarps, buttes, and fault-block mountains; the western boundary is formed by the Cascade Range volcanic arc. The basin containing Upper Klamath Lake and Agency Lake is a fault-bounded trough, or graben (Gonthier, 1984). Much of the upper Klamath Basin has a poorly developed drainage system that includes many small streams that discharge into marshes and intermittent streams that disappear into pumice or porous lava.

Major land use activities in the upper Klamath Basin include irrigated agriculture, ranching, and timber production, and are controlled by elevation. Most of the irrigated agriculture areas are at lower elevations adjacent to the major rivers. Agriculture in the Williamson and Sprague River Basins is primarily irrigated pasture. In the Lost River Basin, irrigation is mostly used for crops such as oats, barley, wheat, potatoes, and sugar beets. Rangelands are mainly on the tablelands, benches, and terraces, and forest is predominant on the slopes of the buttes and mountains. Livestock grazing can occur on irrigated pastureland, rangeland, and forestland throughout the basin. Although forestland accounts for 56 percent of the upper Klamath Basin, it is not homogeneous; second- and third-growth stands are in varying stages of regeneration. Cropland accounts for 6 percent of the upper Klamath Basin. Range, wetlands, water bodies, and urban areas compose the remaining 38 percent of upper Klamath Basin land use (U.S. Geological Survey, 1999). Generating over $300 million annually, agriculture is the mainstay of the upper Klamath Basin economy (Bureau of Reclamation, 2005).

Mean annual precipitation in the upper Klamath Basin study area ranges from approximately 12 inches in the Tule Lake region to 70 inches in the Cascade Range (Daly and others, 1994, 1997). The Cascade Range creates a rain shadow that affects the areal distribution of precipitation throughout much of the upper Klamath Basin. Mean annual precipitation in the upper Williamson River Basin (not including the Sprague River), Sprague River Basin, and the Lost River Basin is approximately 28, 23, and 17 inches, respectively. For the entire upper Klamath Basin it is 23 inches. Snowfall represents 30 percent of the annual precipitation in the valleys and more than 50 percent of the total at higher elevations (Taylor and Hannan, 1999). For most upper Klamath Basin streams, seasonal high flows occur from February to May and low flows occur from July through September.

Acknowledgements

The authors thank Phil Pasteris and David Garen, of the U.S. Department of Agriculture Natural Resources Conservation Service, Portland, Oregon, and John Rasmussen of the Bureau of Reclamation, Klamath Falls, Oregon, for their assistance in this study.

Data Networks

Snow-water equivalent, precipitation, air temperature, streamflow, and net reservoir inflow data were used to create both the dependent and independent variables of the models presented in this report (*table 1*). The USGS, NRCS, BOR, and NWS maintain long-term data monitoring networks for these data in the study area. The data period of record used for many of the models in the study ranged from 1960 to the present.

Snow-Water Equivalent

The NRCS National Water and Climate Center operates and maintains an extensive network of snow measurement sites in the upper Klamath Basin and throughout the West. The network comprises both manually measured snow course sites and automatic Snow Telemetry (SNOTEL) sites. NRCS began snow-course measurements in the 1930s in the upper Klamath Basin. At these sites snow-water equivalent data are collected once a month throughout the winter and spring seasons. SNOTEL sites were established in the upper Klamath Basin in the late 1970s and early 1980s (*table 2*). Having permanent snow-pillow and telemetry instrumentation, it is possible to collect SWE data continuously and make it available in real time. Daily precipitation and daily minimum and maximum air temperatures are also collected at SNOTEL sites.

Precipitation

Precipitation data used in the models were collected at NWS Cooperative program and NRCS SNOTEL sites. Although precipitation data generally are collected on a daily basis, total monthly volumes were used as input to most of the models. None of the models used precipitation data collected prior to 1960. However, the precipitation site with the longest record is at Crater Lake and was established in 1931.

Streamflow

Two of the forecast sites, Williamson and Sprague Rivers, are co-located with USGS streamflow gages. Flow data from these two gages were used as dependent variables in the models. Daily flow records for the gages began in the 1920s. However, their entire records were not used in the model development because most of the SWE and precipitation time series, used as input variables to the models, did not extend back that far. The Sprague River gage (11501000) is located above the confluence with the Williamson River near Chiloquin, Oregon, and has an upstream drainage area of 1,580 square miles. The Williamson River gage (11502500) is located just below the confluence and has an upstream drainage area of approximately 3,000 square miles.

Table 1. Upper Klamath Basin flow and climate data used as monthly model variables.

[fig., figure; ID, identification; deg., degrees; min., minutes; sec., seconds; feet (NGVD29), feet above the 1929 National Geodetic Vertical Datum; FLOW FS, streamflow forecast site; INFLOW FS, calculated net reservoir inflow forecast site; SWE, snow-water equivalent; PREC, precipitation; WELL, interpolated well elevation; NA, not available; USGS, U.S. Geological Survey; BOR, Bureau of Reclamation; NRCS, Natural Resources Conservation Service; NOAA, National Oceanic and Atmospheric Administration; OWRD, Oregon Water Resources Department]

Site no. (fig. 1)	Data type	Site description	Agency site ID	Agency	Latitude			Longitude			Elevation (feet, NGVD29)
					(deg.)	(min.)	(sec.)	(deg.)	(min.)	(sec.)	
1	FLOW FS	Sprague River, Oregon	11501000	USGS	42	35	5	121	50	55	4,202
2	FLOW FS	Williamson River, Oregon	11502500	USGS	42	33	54	121	52	42	4,148
3	INFLOW FS	Clear Lake, California	NA	BOR	41	52	0	121	7	0	4,350
4	INFLOW FS	Gerber Reservoir, Oregon	NA	BOR	42	13	0	121	6	0	4,840
5	INFLOW FS	Upper Klamath Lake, Oregon	NA	BOR	42	15	0	121	48	0	4,098
6	SWE	Adin Mountain, California	20H08	NRCS	41	14	0	120	47	0	6,350
7	SWE	Annie Springs, Oregon	22G06	NRCS	42	53	0	122	10	0	6,020
8	SWE	Billie Creek Divide, Oregon	22G13	NRCS	42	25	0	122	16	0	5,300
9	SWE	Cedar Pass, California	20H06	NRCS	41	35	0	120	18	0	7,100
10	SWE	Chemult, Oregon	21F22	NRCS	43	13	0	121	48	0	4,760
11	SWE	Crazyman Flat, Oregon	20G12	NRCS	42	38	0	120	57	0	6,100
12	SWE	Crowder Flat, California	20H02	NRCS	41	53	0	120	45	0	5,200
13	SWE	Diamond Lake, Oregon	22F18	NRCS	43	11	0	122	8	0	5,320
14	SWE	Dismal Swamp, California	20H12	NRCS	41	58	0	120	10	0	7,050
15	SWE	Dog Hollow, Oregon	21G06	NRCS	42	7	0	121	7	0	4,900
16	SWE	Finley Corrals, Oregon	20G14	NRCS	42	27	0	120	47	0	6,000
17	SWE	Fourmile Lake, Oregon	22G12	NRCS	42	26	0	122	15	0	6,000
18	SWE	Gerber Reservoir, Oregon	21G04	NRCS	42	12	0	121	8	0	4,850
19	SWE	Howard Prairie, Oregon	22G26	NRCS	42	13	0	122	23	0	4,500
20	SWE	Crater Lake, Oregon	22G05	NRCS	42	54	0	122	8	0	6,550
21	SWE	Quartz Mountain, Oregon	20G06	NRCS	42	16	0	120	47	0	5,320
22	SWE	Sevenmile Marsh, Oregon	22G33	NRCS	42	41	0	122	8	0	5,720
23	SWE	Silver Creek, Oregon	21F12	NRCS	42	57	0	121	11	0	4,900
24	SWE	State Line, California	20H01	NRCS	41	59	0	120	43	0	5,750
25	SWE	Strawberry, Oregon	20G09	NRCS	42	6	0	120	58	0	5,760
26	SWE	Summer Rim, Oregon	20G02	NRCS	42	42	0	120	49	0	7,200
27	SWE	Sycan Flat, Oregon	21G09	NRCS	42	52	0	121	0	0	5,500
28	SWE	Taylor Butte, Oregon	21G03	NRCS	42	42	0	121	24	0	5,100

Table 1. Upper Klamath Basin flow and climate data used as monthly model variables.–Continued

[fig., figure; ID, identification; deg., degrees; min., minutes; sec., seconds; feet (NGVD29), feet above the 1929 National Geodetic Vertical Datum; FLOW FS, streamflow forecast site; INFLOW FS, calculated net reservoir inflow forecast site; SWE, snow-water equivalent; PREC, precipitation; WELL, interpolated well elevation; NA, not available; USGS, U.S. Geological Survey; BOR, Bureau of Reclamation; NRCS, Natural Resources Conservation Service; NOAA, National Oceanic and Atmospheric Administration; OWRD, Oregon Water Resources Department]

Site no. (fig. 1)	Data type	Site description	Agency site ID	Agency	Latitude			Longitude			Elevation (feet, NGVD29)
					(deg.)	(min.)	(sec.)	(deg.)	(min.)	(sec.)	
29	PREC	Chemult, Oregon	1546	NOAA	43	14	0	121	47	0	4,760
30	PREC	Chiloquin, Oregon	1571	NOAA	42	39	0	121	57	0	4,160
31	PREC	Crater Lake, Oregon	1946	NOAA	42	54	0	122	8	0	6,470
32	PREC	Gerber Reservoir, Oregon	3232	NOAA	42	12	0	121	8	0	4,850
33	PREC	Howard Prairie, Oregon	4060	NOAA	42	13	0	122	22	0	4,570
34	PREC	Lakeview, Oregon	4670	NOAA	42	13	0	120	22	0	4,780
35	PREC	Quartz Mountain, Oregon	S367	NRCS	42	16	0	120	47	0	5,700
36	PREC	Round Grove, Oregon	7354	NOAA	42	20	0	120	53	0	4,890
37	PREC	Summer Rim, Oregon	S068	NRCS	42	42	0	120	49	0	7,100
38	PREC	Taylor Butte, Oregon	S058	NRCS	42	42	0	121	24	0	5,100
39	FLOW	Fall River, Oregon	14057500	OWRD	43	47	48	121	34	18	4,220
40	WELL	Upper Sprague River, Oregon	2145	OWRD	42	25	7	121	0	28	4,360

Table 2. Natural Resources Conservation Service, Upper Klamath Basin Snow Telemetry (SNOTEL) sites.

[ID, identification; deg., degrees; min., minute; sec., second; feet (NGVD29), feet above the National Geodetic Vertical Datum of 1929]

Site description	Site ID	Latitude			Longitude			Elevation (feet, NGVD29)	Period of record water year
		(deg.)	(min.)	(sec.)	(deg.)	(min.)	(sec.)		
Adin Mountain, California	20H13S	41	15	0	120	46	0	6,200	1985–04
Big Red Mountain, Oregon	22G21S	42	3	0	122	51	0	6,250	1981–04
Billie Creek Divide, Oregon	22G13S	42	25	0	122	17	0	5,300	1985–04
Cedar Pass, California	20H13S	41	35	0	120	18	0	7,100	1979–04
Chemult, Oregon	21F22S	43	13	0	121	48	0	4,760	1981–04
Cold Springs Camp, Oregon	22G24S	42	32	0	122	11	0	6,100	1982–04
Diamond Lake, Oregon	22F18S	43	11	0	122	8	0	5,315	1983–04
Fish Lake, Oregon	22G14S	42	23	0	122	25	0	4,665	1982–04
Fourmile Lake, Oregon	22G12S	42	24	0	122	13	0	6,000	1981–04
New Crescent Lake, Oregon	21F10S	43	29	0	121	58	0	4,800	1981–04
Quartz Mountain, Oregon	20G06S	42	16	0	120	47	0	5,700	1981–04
Sevenmile Marsh, Oregon	22G33S	42	41	0	122	8	0	6,200	1981–04
Silver Creek, Oregon	21F12S	42	57	0	121	11	0	5,720	1981–04
Strawberry, Oregon	20G09S	42	6	0	121	51	0	5,760	1981–04
Summer Rim, Oregon	20G02S	42	42	0	120	49	0	7,100	1979–04
Summit Lake, Oregon	22F14S	43	27	0	122	8	0	5,600	1982–04
Taylor Butte, Oregon	21G03S	42	42	0	121	24	0	5,100	1979–04

Flow data collected at the USGS gage on Fall River near La Pine, Oregon (14507500), was used as an independent variable for some of the models. Fall River flow consists entirely of spring discharge. Although the river is north of the upper Klamath Basin in the Deschutes River Basin, its flow record was included in the modeling because it closely reflects regional climate patterns. Fall River has a geologic and geographic setting very similar to many upper Klamath Basin streams. It is also located on the east side of the Cascades and emerges from young volcanic material. Fall River flow integrates climate conditions over multiple years, so the effects of anomalous dry or wet years can persist for 1 or 2 years, and change in climatic conditions may take 2 or 3 years to be manifest fully (*fig. 3*).

Reservoir Net Inflow

Monthly net inflow records for three of the forecast sites, Upper Klamath Lake, Gerber Reservoir, and Clear Lake Reservoir, were computed by BOR and were used as dependent variables in the models. For each reservoir, monthly net inflow was computed as the sum of the change in storage volume and the volume of all measured flow diversions or releases. The change in storage volume is computed using a stage-storage curve, unique for each reservoir, and the end of month (last hour of the last day) stage measurements of a given month and its proceeding month.

Climate-Trend Variables

In addition to the ground-water dominated Fall River flow records, other variables representing interannual flow conditions in the upper Klamath Basin were assembled and evaluated for their potential use as model input variables. As discussed earlier, these included well-water elevation data, precipitation records, and ocean climate indicators.

Well hydrograph data selected for the study were collected by the Oregon Water Resources Department (OWRD), which maintains a network of approximately 20 long-term observation wells in the upper Klamath Basin (*http://or.water.usgs.gov/projs_dir/or180/waterlevels/nwis/index.html*). Many of the wells have records that began in the early 1960s. However, many of the records were not useful indicators of climate trends because of the effects of nearby irrigation pumping. However, one long-term record from a well (#2145) located in the eastern headwaters of the Sprague River, away from significant pumping impacts, was suitable for use. Like most of the wells in the network, measurements at this

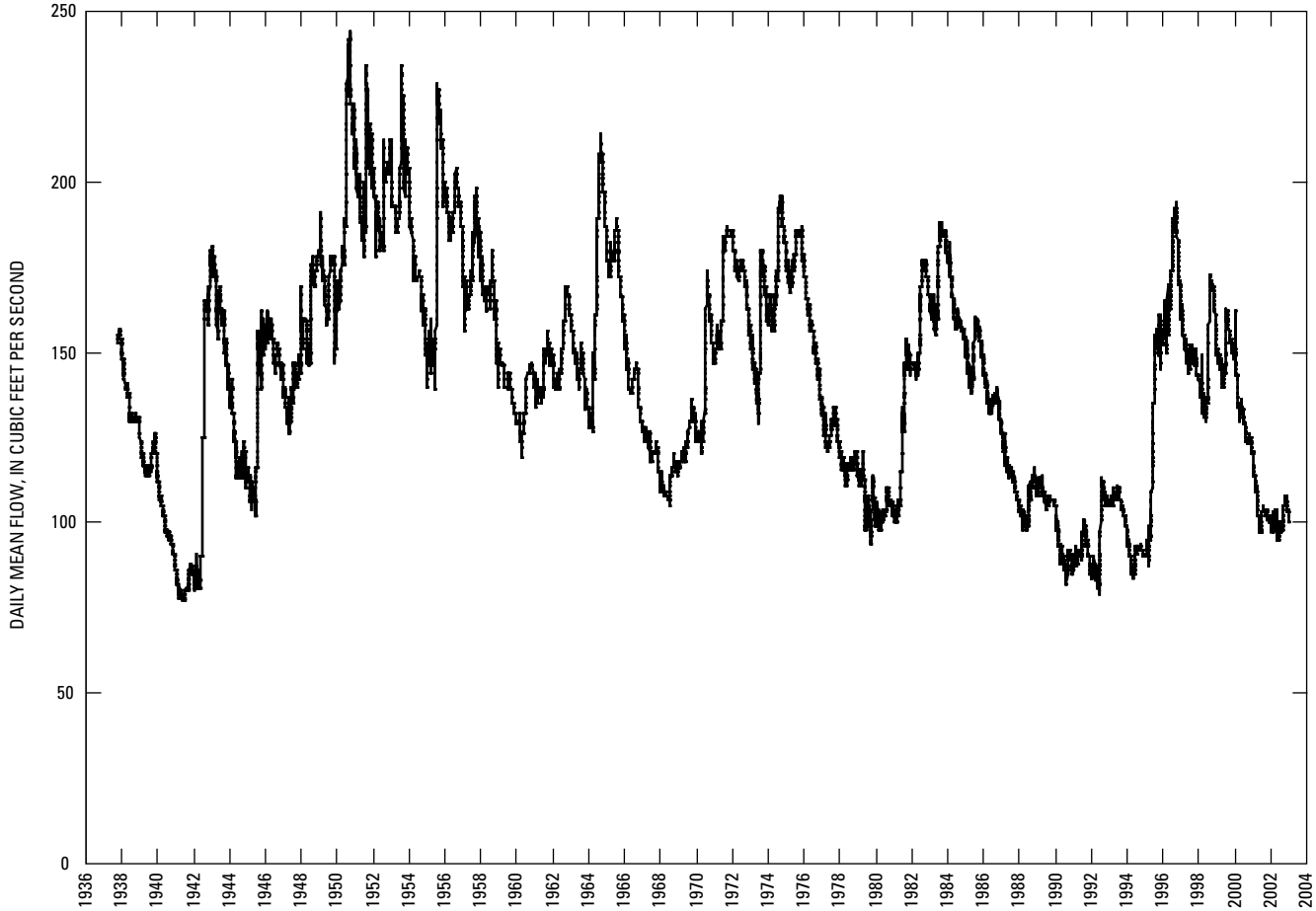


Figure 3. Daily mean flow at Fall River near La Pine, Oregon (14057500), 1938–2003.

well have been made manually every few months, at irregular intervals, throughout its 43-year period. The first measurement in the record is April 3, 1962. Some adjustments to the original record were made to make it more usable as a model variable. The record was first smoothed with a 5-point moving average filter. To create an annual time series, a January 1 well depth for every year was interpolated from the smoothed record (*fig. 4*). It was assumed that if the record were affected by any possible nearby summer time pumping, the aquifer would have recovered by the following January.

Another indicator variable of climate trends was created by computing the cumulative departure from normal of the Crater Lake, Oregon, and Prospect, Oregon, annual precipitation records. Although Prospect is located west of the upper Klamath Basin in the nearby Rogue River Basin, it and the Crater Lake precipitation records were selected because they are the longest and most reliable in the region. A cumulative departure from normal was computed by first subtracting the average annual water year precipitation for the period of 1932–2004 from annual precipitation for each water year. These annual differences were then successively cumulated starting with the first year and ending with the last year in the period. The cumulative departures from normal computed from both Crater Lake and Prospect records were then averaged to create a regional variable. By computing a cumulative departure from normal time series from a precipitation time series, the dry and wet trends usually become more apparent than they would be if only using the precipitation time series.

The Pacific Decadal Oscillation (PDO) index is another indicator of long-term climate trends that was evaluated as a model variable. The PDO index is the difference between the mid-North Pacific Ocean and North American coastline sea surface temperatures (SST) (Mantua and others, 1997).

$$\text{PDO index} = (\text{Coast line SST} - \text{mid-ocean SST}) \qquad (1)$$

For many locations in the Pacific Northwest, negative PDO values coincide with cool and wet periods (1890–1924, 1947–1976). Positive PDO values coincide with warm and dry periods (1925–1946, 1977–present).

Models

Three different sets of models used to forecast spring and summer flows at five upper Klamath Basin forecast sites are presented. The first model set was developed by NRCS and uses principal components regression. Current or recent monthly SWE and precipitation from climate sites near the forecast sites were used as input data. The second set of models, developed by the USGS, also used current or recent monthly SWE and precipitation as input data, except these models were created from artificial neural networks. The third set of models, developed by Advanced Data Mining LLC, Greenville, South Carolina, was also created from artificial neural networks. However, these models are autoregressive and use model input variables that are derived from components of the flow time series of the forecast site.

Principal Components Regression

Principal components regression is an alternative to multiple regression that is appropriate for situations where the independent variables are correlated with each other (Garen, 1992). The NRCS uses principal components regression because the most commonly used input variables to predict seasonal flows, SWE and precipitation, are usually highly cross-correlated. Highly correlated independent variables in a multiple regression can often produce inaccurately estimated variable coefficients that do not make physical sense. For example, variables that are expected to be positively correlated with flow, such as SWE and precipitation, could be given negative signs.

Principal components regression uses principal components analysis (PCA) techniques that can transform a set of cross-correlated variables into non-cross-correlated variables that can be used for descriptive or predictive (or regression) purposes. In the first step of the analysis, the eigenvalues and eigenvectors of the correlation matrix of the original variables are computed. The eigenvalues describe the fractional importance of each of the principal components of the matrix. The eigenvectors are used as weights on each of the original variables to compute the new set of variables.

$$W_1 = ev_{1,1}Z_1 + ev_{1,2}Z_2 + ev_{1,3}Z_3 + \ldots + ev_{1,n}Z_n \quad (2)$$

$$W_2 = ev_{2,1}Z_1 + ev_{2,2}Z_2 + ev_{2,3}Z_3 + \ldots + ev_{2,n}Z_n$$

$$W_3 = ev_{3,1}Z_1 + ev_{3,2}Z_2 + ev_{3,3}Z_3 + \ldots + ev_{3,n}Z_n$$

$$W_n = ev_{n,1}Z_1 + ev_{n,2}Z_2 + ev_{n,3}Z_3 + \ldots + ev_{n,n}Z_n$$

Where W is a principal component variable, ev is an eigenvector of the n by n correlation matrix of the original snow and precipitation variables, and Z is an original snow and

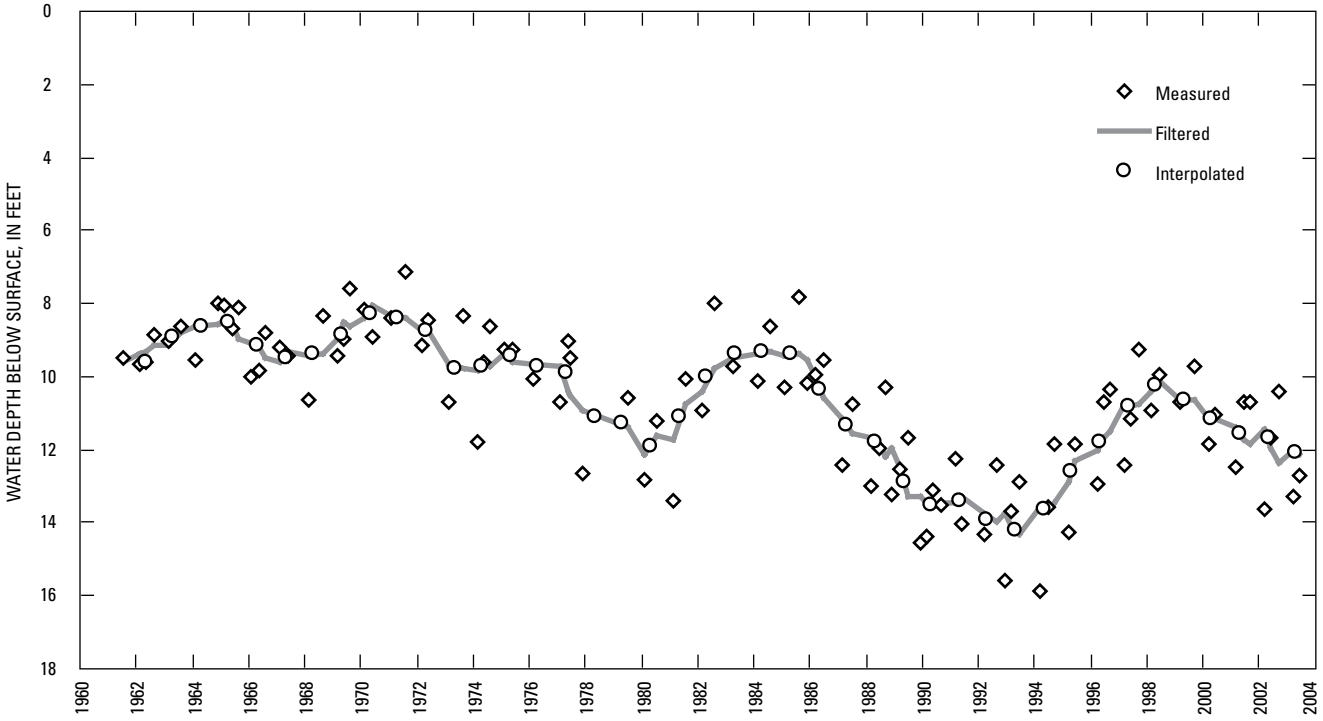


Figure 4. Upper Sprague River Basin well hydrograph time series, 1962–2003.

precipitation variable. Using the equations above, it is possible to transform the original data set of Z_n variables into a data set of W_n variables having an equal number of cases (or years). The W_n variables are then used as input variables in a multiple-regression model to predict the flow volume output variable. Typically, most of the variability will be explained by just the first two variables (W_1 and W_2). Additional W variables usually are insignificant and are not used. Many NRCS principal components regression equations will use original variables of SWE and precipitation, sometimes from a dozen different snow-course sites that may surround a forecast site. However, the actual regression may be based on only the first and second principal component variables. This approach has the similar effect of averaging climate data from multiple sites into a climate index site.

To create the forecast equations, the NRCS uses a semi-automated approach to construct the most optimal equations from a pool of available input variables. To create a single equation, over several thousand combinations of variables are typically tested using an optimization program. However, the most optimal equation (with the highest correlation coefficient and the lowest error) is not necessarily selected as the finalized equation by the NRCS. Because separate equations are used for each forecast date from January to June, the finalized equations are manually chosen for the purpose of maintaining input variable consistency from month to month. Doing this reduces the chance that flow forecasts will abruptly change from month to month. However, the results from the optimization program are still used in the process as guidance.

The NRCS optimization program also uses a cross-validation technique or a "jackknife test" to lesson the tendency of equations to overfit themselves to the data sets used to create them. The NRCS cross-validation technique is done systematically by repeatedly pulling 1 year from the record, recomputing the regression equation using the remaining years, and testing the equation on the year pulled from the record.

As more years of climate and flow data are collected in the upper Klamath Basin, the NRCS updates their equations for the five forecast sites every few years. The equations, updated for this study, are shown in *tables 3–7*. Some of the equation time periods vary because the availability of data from different climate sites also vary. However, none of the equation time periods extend back earlier than 1960. The equations predict a total volume of water in thousand acre-feet that is expected at a site for the 6-month periods of February through July and April through September. The NRCS flow forecasts are officially presented as a range of probable flow volumes (90-, 70-, 50-, 30-, 10-percent chance of exceedance) rather than as a single number.

It is important to note that the NRCS does not necessarily use the model output as their posted official flow forecast. Sometimes the model output flow forecasts are adjusted on the basis of other available information. For some sites, where there is overlap, the NRCS flow forecasts are coordinated with flow forecasts from the National Weather Service to avoid discrepancies between the two agencies. However, in this report, only NRCS model output flow forecasts are presented in the analyses.

Table 3. Principal components regression Williamson River flow forecast models.

[JR^2, jackknife coefficient of determination; vol., volume; P, precipitation; S, snow-water equivalent; F, flow; T, air temperature; W, Upper Sprague River Basin well hydrograph; Will. R., Williamson River; Bt., Butte; Lk., Lake; Spr., Spring; Ck., Creek; Res., Reservoir]

April-September models								
Forecast date	**JR^2**	**Dependent variable**	**Independent variables (month-site-parameter)**					
Jan. 1	0.54	Apr.–Sep. flow vol.	Oct.-Chiloquin-P	Dec.-Chiloquin-P	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Crater Lk.-S	Jan.-Silver Ck.-S
			Oct.-Will. R.-F	Jan.-Sprague R.-W				
Feb. 1	0.75	Apr.–Sep. flow vol.	Oct.-Chiloquin-P	Jan.-Chiloquin-P	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Feb.-Crater Lk.-S
			Feb.-Silver Ck.-S	Feb.-Taylor Bt.-S	Oct.-Will. R.-F	Jan.-Sprague R.-W		
March 1	0.85	Apr.–Sep. flow vol.	Feb.-Chiloquin-P	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Mar.-Crater Lk.-S	Mar.-Silver Ck.-S
			Mar.-Taylor Bt.-S	Jan.-Sprague R.-W				
April 1	0.94	Apr.–Sep. flow vol.	Mar.-Chiloquin-P	Dec.-Gerber Res.-P	Mar.-Gerber Res.-P	Apr.-Annie Spr.-S	Apr.-Crater Lk.-S	Apr.-Taylor Bt.-S
			Mar.-Fall R.-F	Mar.-Will. R.-F	Jan.-Sprague R.-W			
May 1	0.82	May–Sep. flow vol.	Apr.-Gerber Res.-P	May-Annie Spr.-S	May-Crater Lk.-S	Apr.-Taylor Bt.-S	Mar.-Fall R.-F	Mar.-Will. R.-F
			Apr.-Will. R.-F	Mar.-Crater Lk.-T	Jan.-Sprague R.-W			
June 1	0.82	Jun.–Sep. flow vol.	Jun.-Annie Spr.-S	Mar.-Fall R.-F	Apr.-Will. R.-F	May-Will. R.-F	May-Crater Lk.-T	Jan.-Sprague R.-W
February-July models								
Forecast date	**JR^2**	**Dependent variable**	**Independent variables (month-site-parameter)**					
Jan. 1	0.54	Feb.–Jul. flow vol.	Oct.-Chiloquin-P	Dec.-Chiloquin-P	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Crater Lk.-S	Jan.-Silver Ck.-S
			Oct.-Will. R.-F	Jan.-Sprague R.-W				
Feb. 1	0.70	Feb.–Jul. flow vol.	Oct.-Chiloquin-P	Jan.-Chiloquin-P	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Feb.-Crater Lk.-S
			Feb.-Silver Ck.-S	Feb.-Taylor Bt.-S	Oct.-Will. R.-F	Jan.-Sprague R.-W		
March 1	0.85	Mar.–Jul. flow vol.	Feb.-Chiloquin-P	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Mar.-Crater Lk.-S	Mar.-Silver Ck.-S
			Mar.-Taylor Bt.-S	Jan.-Sprague R.-W				
April 1	0.94	Apr.–Jul. flow vol.	Mar.-Chiloquin-P	Dec.-Gerber Res.-P	Mar.-Gerber Res.-P	Apr.-Annie Spr.-S	Apr.-Crater Lk.-S	Apr.-Taylor Bt.-S
			Mar.-Fall R.-F	Mar.-Will. R.-F	Jan.-Sprague R.-W			
May 1	0.78	May–Jul. flow vol.	Apr.-Gerber Res.-P	May-Annie Spr.-S	May-Crater Lk.-S	Apr.-Taylor Bt.-S	Mar.-Fall R.-F	Mar.-Will. R.-F
			Apr.-Will. R.-F	Mar.-Crater Lk.-T	Jan.-Sprague R.-W			
June 1	0.78	Jun.–Jul. flow vol.	Jun.-Annie Spr.-S	Mar.-Fall R.-F	Apr.-Will. R.-F	May-Will. R.-F	May-Crater Lk.-T	Jan.-Sprague R.-W

Table 4. Principal components regression Sprague River flow forecast models.

[JR^2, jackknife coefficient of determination; vol., volume; P, precipitation; S, snow-water equivalent; Bt., Butte; Sum., Summer; Mtn., Mountain; Fl., Flat; Res., Reservoir; NA, not available]

April-September models								
Forecast date	**JR^2**	**Dependent variable**	**Independent variables (month-site-parameter)**					
Jan. 1	0.51	April–Sept. flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Dec.-Round Grove-P	Jan.-Gerber Res.-S		
Feb. 1	0.71	April–Sept. flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Nov.-Round Grove-P	Nov.-Sum. Rim-S	Feb.-Crazyman Fl.-S
			Jan.-Taylor Bt.-S					
March 1	0.79	April–Sept. flow vol.	Nov.-Gerber Res.-P	Jan.-Gerber Res.-P	Nov.-Quartz Mtn.-S	Dec.-Quartz Mtn.-S	Mar.-Crazyman Fl.-S	Mar.-Sum. Rim-S
			Mar.-Taylor Bt.-S					
April 1	0.89	April–Sept. flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Apr.-Crazyman Fl.-S	Apr.-Sum. Rim-S	Apr.-Taylor Bt.-S
May 1	0.82	May–Sept. flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Apr.-Crazyman Fl.-S	May-Sum. Rim-S	
June 1	NA							
February-July models								
Forecast date	**JR^2**	**Dependent variable**	**Independent variables (month-site-parameter)**					
Jan. 1	0.52	Feb.–July flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Dec.-Round Grove-P	Jan.-Gerber-S		
Feb. 1	0.76	Feb.–July flow vol.	Nov.-Gerber Res.-P	Jan.-Gerber Res.-P	Dec.-Sum. Rim-S	Feb.-Sycan Flat-S		
March 1	0.80	Mar.–July flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Nov.-Quartz Mtn.-S	Dec.-Quartz Mtn.-S	Feb.-Crazyman Fl.-S
			Mar.-Crazyman Fl.-S	Mar.-Sum. Rim-S	Feb.-Taylor Bt.-S			
April 1	0.87	Apr.–July flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Apr.-Sum. Rim-S	Apr.-Taylor Bt.-S	
May 1	0.80	May–July flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Apr.-Gerber Res.-P	Apr.-Crazyman Fl.-S	May-Sum. Rim-S
			Apr.-Taylor Bt.-S					
June 1	NA							

Table 5. Principal components regression Upper Klamath Lake net inflow forecast models.

[JR^2, jackknife coefficient of determination; vol., volume; P, precipitation; S, snow-water equivalent; T, air temperature; F, flow; N, net inflow; W, Upper Sprague River Basin well hydrograph; Lk., Lake; Res., Reservoir; Spr., Spring; Bt., Butte]

April–September models								
Forecast date	**JR^2**	**Dependent variable**	**Independent variable (month-site parameter)**					
Jan. 1	0.52	Apr.–Sep. flow vol.	Nov.-Crater Lk.-P	Dec.-Crater Lk.-P	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Nov.-Round Grove-P	Jan.-Annie Spr.-S
			Jan.-Chemult-S	Jan.-Crater Lk.-S	Oct.-Fall R.-F	Nov.-Fall R.-F	Nov.-U. Klamath Lk.-N	Jan.-Sprague R.-W
Feb. 1	0.70	Apr.–Sep. flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Feb.-Annie Spr.-S	Feb.-Chemult-S	Feb.-Cold Spr.-S
			Feb.-Crater Lk.-S	Feb.-Sevenmile-S	Oct.-Fall R.-F	Nov.-Fall R.-F	Nov.-U. Klamath Lk.-N	Jan.-Sprague R.-W
March 1	0.78	Apr.–Sep. flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Feb.-Gerber Res.-P	Mar.-Annie Spr.-S	Mar.-Chemult-S
			Mar.-Cold Spr.-S	Mar.-Crater Lk.-S	Mar.-Sevenmile-S	Mar.-Summer Rim-S	Mar.-Taylor Bt.-S	Oct.-Fall R.-F
			Nov.-Fall R.-F	Nov.-U. Klamath Lk.-N	Jan.-Sprague R.-W			
April 1	0.95	Apr.–Sep. flow vol.	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Mar.-Gerber Res.-P	Mar.-Round Grove-P	Apr.-Annie Spr.-S	Apr.-Crater Lk.-S
			Mar.-Sevenmile-S	Mar.-Summer Rim-S	Apr.-Taylor Bt.-S	Oct.-Fall R.-F	Nov.-Fall R.-F	Mar.-U. Klamath Lk.-N
			Jan.-Sprague R.-W	Mar.-Crater Lk.-T				
May 1	0.91	May–Sep. flow vol.	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Mar.-Gerber Res.-P	Apr.-Gerber Res.-P	Mar.-Round Grove-P	Apr.-Annie Spr.-S
			Apr.-Crater Lk.-S	Mar.-Sevenmile-S	May-Summer Rim-S	Apr.-Taylor Bt.-S	Oct.-Fall R.-F	Nov.-Fall R.-F
			Apr.-U. Klamath Lk.-N	Jan.-Sprague R.-W	Mar.-Crater Lk.-T			
June 1	0.92	Jun.–Sep. flow vol.	Dec.-Gerber Res.-P	Apr.-Gerber Res.-P	Mar.-Round Grove-P	June-Annie Spr.-S	May-Crater Lk.-S	Apr.-Taylor Bt.-S
			Oct.-Fall R. F	Nov.-Fall R. F	May-U. Klamath Lk. N	Jan.-Sprague R. W	Mar.-Crater Lk.T	May-Crater Lk. T
February–July models								
Forecast date	**JR^2**	**Dependent variable**	**Independent variable (month-site parameter)**					
Jan. 1	0.48	Feb.–Jul. flow vol.	Nov.-Crater Lk. P	Dec.-Crater Lk. P	Nov.-Gerber Res. P	Dec.-Gerber Res. P	Nov.-Round Grove P	Jan.-Annie Spr. S
			Jan.-Chemult S	Jan.-Crater Lk. S	Oct.-Fall R. F	Nov.-Fall R. F	Nov.-U. Klamath Lk. N	Jan.-Sprague R. W
Feb. 1	0.71	Feb.–Jul. flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Feb.-Annie Spr.-S	Feb.-Chemult-S	Feb.-Cold Spr.-S
			Feb.-Crater Lk.-S	Feb.-Sevenmile-S	Oct.-Fall R.-F	Nov.-Fall R.-F	Nov.-U. Klamath Lk.-N	Jan.-Sprague R.-W
March 1	0.72	Mar.–Jul. flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Feb.-Gerber Res.-P	Mar.-Annie Spr.-S	Mar.-Chemult-S
			Mar.-Cold Spr.-S	Mar.-Crater Lk.-S	Mar.-Sevenmile-S	Oct.-Fall R.-F	Nov.-Fall R.-F	Nov.-U. Klamath Lk.-N
			Jan.-Sprague R.-W					
April 1	0.94	Apr.–Jul. flow vol.	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Mar.-Gerber Res.-P	Mar.-Round Grove-P	Apr.-Annie Spr.-S	Apr.-Crater Lk.-S
			Mar.-Sevenmile-S	Mar.-Summer Rim-S	Apr.-Taylor Bt.-S	Oct.-Fall R.-F	Nov.-Fall R.-F	Mar.-U. Klamath Lk.-N
			Jan.-Sprague R.-W	Mar.-Crater Lk.-T				
May 1	0.87	May–Jul. flow vol.	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Mar.-Gerber Res.-P	Apr.-Gerber Res.-P	Mar.-Round Grove-P	Apr.-Annie Spr.-S
			Apr.-Crater Lk.-S	Mar.-Sevenmile-S	May-Summer Rim-S	Apr.-Taylor Bt.-S	Oct.-Fall R.-F	Nov.-Fall R.-F
			Apr.-U. Klamath Lk.-N	Jan.-Sprague R.-W	Mar.-Crater Lk.-T			
June 1	0.79	Jun.–Jul. flow vol.	Jun.-Annie Spr.-S	Nov.-Fall R.-F	May-U. Klamath Lk.-N	Jan.-Sprague R.-W	May-Crater Lk.-T	

Table 6. Principal components regression Clear Lake Reservoir net inflow forecast models.

[JR^2, jackknife coefficient of determination; vol., volume; P, precipitation; S, snow-water equivalent; Res., Reservoir; NA, not available]

April–September models								
Forecast date	**JR^2**	**Dependent variable**	**Independent variables (month-site-parameter)**					
Jan. 1	0.40	Apr.–Sep. flow vol.	Jan.-Cedar Pass-S	Nov.-Gerber Res.-P	Nov.-Lakeview-P	Dec.-Lakeview-P		
Feb. 1	0.67	Apr.–Sep. flow vol.	Jan.-Cedar Pass-S	Nov.-Gerber Res.-P	Jan.-Gerber Res.-P	Nov.-Lakeview-P	Dec.-Lakeview-P	Feb.-Strawberry-S
March 1	0.65	Apr.–Sep. flow vol.	Mar.-Dismal Swamp-S	Nov.-Gerber Res.-P	Jan.-Gerber Res.-P	Dec.-Lakeview-P	Feb.-Strawberry-S	
April 1	0.81	Apr.–Sep. flow vol.	Mar.-Dismal Swamp-S	Apr.-State Line-S	Nov.-Gerber Res.-P	Jan.-Gerber Res.-P	Mar.-Gerber Res.-P	Dec.-Lakeview-P
			Mar.-Lakeview-P	Feb.-Strawberry-S				
May 1	0.50	May–Sep. flow vol.	Feb.-Cedar Pass-S	Apr.-Cedar Pass-S	May-Cedar Pass-S	Mar.-Dismal Swamp-S	Feb.-State Line-S	Apr.-State Line-S
			Jan.-Gerber Res.-P	Mar.-Gerber Res.-P	Jan.-Lakeview-P			
June 1	NA							
February–July models								
Forecast date	**JR^2**	**Dependent variable**	**Independent variables (month-site-parameter)**					
Jan. 1	0.34	Feb.–Jul. flow vol.	Dec.-Gerber Res-P	Nov-Lakeview-P				
Feb. 1	0.70	Feb.–Jul. flow vol.	Feb.-Cedar Pass-S	Nov.-Gerber Res.-P	Jan.-Gerber Res.-P	Nov.-Lakeview-P	Feb.-Dog Hollow-S	Feb.-Strawberry-S
March 1	0.71	Mar.–Jul. flow vol.	Feb.-Cedar Pass-S	Nov.-Gerber Res.-P	Jan.-Gerber Res.-P	Nov.-Lakeview-P	Feb.-Strawberry-S	Mar.-Strawberry-S
April 1	0.76	Apr.–Jul. flow vol.	Apr.-Dismal Swamp-S	Apr.-State Line-S	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Jan.-Gerber Res.-P	Mar.-Gerber Res.-P
			Nov.-Lakeview-P	Dec.-Lakeview-P	Feb.-Strawberry-S			
May 1	0.59	May–Jul. flow vol.	Apr.-Cedar Pass-S	May-Cedar Pass-S	Apr.-State Line-S	Jan.-Gerber Res.-P	Mar.-Gerber Res.-P	Apr.-Gerber Res.-P
			Apr.-Lakeview-P					
June 1	NA							

Table 7. Principal components regression Gerber Reservoir net inflow forecast models.

[JR^2, jackknife coefficient of determination; vol., volume; P, precipitation; S, snow-water equivalent; N, net inflow; Res., Reservoir; Mtn., Mountain; Fl., Flat]

April–September models								
Forecast date	**JR^2**	**Dependent variable**	**Independent variables (month-site-parameter)**					
Jan. 1	0.25	Apr.–Sep. flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Nov.-Lakeview-P	Oct.-Round Grove-P	Jan.-Quartz Mtn.-S	
Feb. 1	0.45	Apr.–Sep. flow vol.	Nov.-Gerber Res.-P	Oct.-Round Grove-P	Feb.-Strawberry-S			
March 1	0.46	Apr.–Sep. flow vol.	Nov.-Gerber Res.-P	Oct.-Round Grove-P	Feb.-Gerber Res.-S	Feb.-Strawberry-S	Mar.-Strawberry-S	
April 1	0.79	Apr.–Sep. flow vol.	Apr.-Crowder Fl.-S	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Oct.-Lakeview-P	Dec.-Lakeview-P	Oct.-Round Grove-P
			Apr.-Gerber Res.-S	Jan.-Quartz Mtn.-S	Apr.-Strawberry-S	Mar.-Gerber Res.-N	Apr.-State Line-S	
May 1	NA							
June 1	NA							
February–July models								
Forecast date	**JR^2**	**Dependent variable**	**Independent variables (month-site-parameter)**					
Jan. 1	0.41	Feb.–Jul. flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Nov.-Lakeview-P	Oct.-Round Grove-P	Jan.-Gerber Res.-S	Jan.-Quartz Mtn.-S
Feb. 1	0.52	Feb.–Jul. flow vol.	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Nov.-Lakeview-P	Feb.-Quartz Mtn.-S	Feb.-Strawberry-S	
March 1	0.67	Mar.–Jul. flow vol.	Nov.-Gerber Res.-P	Mar.-Dog Hollow-S	Jan.-Quartz Mtn.-S	Feb.-Strawberry-S	Mar.-Strawberry-S	
April 1	0.70	Apr.–Jul. flow vol.	Apr.-Crowder Fl.-S	Oct.-Gerber Res.-P	Nov.-Gerber Res.-P	Dec.-Gerber Res.-P	Nov.-Lakeview-P	Dec.-Lakeview-P
			Oct.-Round Grove-P	Feb.-Strawberry-S	Apr.-Strawberry-S	Mar.-Gerber Res.-N	Apr.-State Line-S	
May 1	0.37	May–Jul. flow vol.	Dec.-Gerber Res.-P	Apr.-Round Grove-P	Apr.-Gerber Res.-S	Jan.-Quartz Mtn.-S	Mar.-Gerber Res.-N	Apr.-Gerber Res.-N
			Apr.-State Line-S					
June 1	0.35	Jun.–Jul. flow vol.	Dec.-Gerber Res.-P	Nov.-Round Grove-P	Apr.-Dog Hollow-S	Jan.-Quartz Mtn.-S	Mar.-Gerber Res.-N	May-Gerber Res.-N
			Apr.-State Line-S					

Nonautoregressive Artificial Neural Networks

Developed by the USGS for the upper Klamath Basin, the nonautoregressive ANN models used precipitation and SWE monthly data sets similar to those used in the NRCS principal components regression models. An objective in the study was determining how well techniques alternative to regression, such as ANN models, would perform in flow forecasting. Unlike linear regression models, which must assume linear relationships between input and output variables, an ANN model is a flexible mathematical structure capable of describing complex nonlinear relationships. Although used in industrial applications for years, ANN modeling is increasingly being used in environmental sciences, particularly for problems where the characteristics of the processes are difficult to simulate using a physically based or regression modeling approach.

Within hydrologic studies, ANN modeling has been used for a variety of purposes in recent years. Kuligowski and Barros (1998) used ANN modeling to estimate missing rainfall data. Karunainithi and others (1994) and Hsu and others (1998) used ANN modeling for streamflow forecasting. River stage also has been forecasted using ANN modeling (Thirumalaiah and Deo, 1998). Hsu and others (1995) and Shamseldin (1997) describe applications of ANN modeling to rainfall-runoff processes. Conrads and Roehl (1999, 2000), and Conrads, Roehl, and Cook (2002), and Conrads, Roehl, and Martello (2002) used ANN models to simulate salinity, temperature, dissolved oxygen, and flow movement in various tidal estuary applications. Morshed and Kaluarachchi (1998) present an ANN model used in complex ground-water flow and contaminant transport simulations. Cannon and Whitfield (2001) modeled transient pH depressions using an ANN model. Risley and others (2003) used ANN models to predict summer water temperatures in small upland streams in western Oregon.

The architecture of ANN models is loosely based on the biological nervous system (Hinton, 1992). ANNs contain interconnected units that are analogous to neurons. The function of the synapse is modeled by a modifiable weight which is associated with each connection. A commonly used ANN model is the feed-forward neural network shown in *figure 5*.

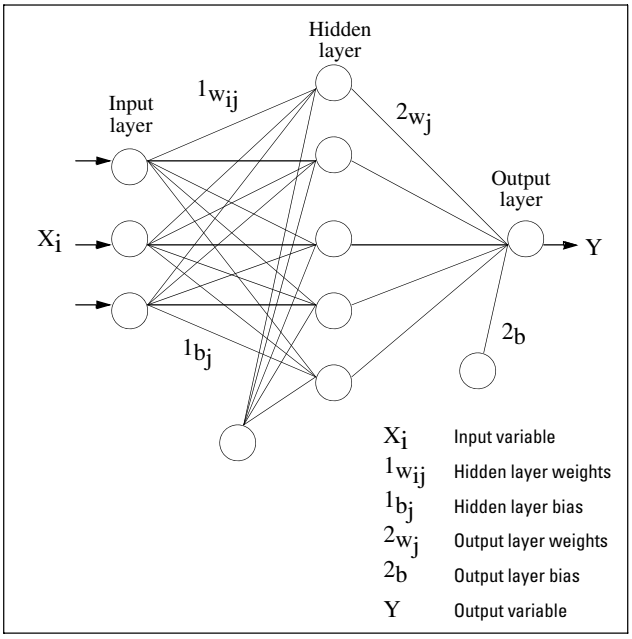


Figure 5. Feed-forward neural network architecture with three inputs, five hidden-layer nodes, and a single output.

This example contains three nodes in the input layer, five nodes in the hidden layer, and a single node in the output layer. The model output is generated by feeding input data through the model from left to right. The output from each hidden-layer node h_j is computed in the following equation:

$$h_j = \tanh\left[\sum_i X_i {}^1w_{ij} + {}^1 b_j\right] \quad (3)$$

where

h_j is the computed output from each hidden-layer node,

j is the hidden-layer node index,

tanh is the hyperbolic tangent,

i is the input layer node index,

X_i is the input variable,

${}^1w_{ij}$ is the hidden-layer weight, and

${}^1b_j$ is hidden-layer bias.

Output from the ANN model is computed in the following:

$$Y = \sum_j h_j {}^2w_j + {}^2 b \quad (4)$$

where

Y is the output variable

${}^2w_j$ is the output layer weights, and

${}^2b$ is the output layer bias.

Nonlinear relationships in the model can be represented by the hyperbolic tangent function, a sigmoid-shaped function, in the hidden-layer nodes. However, the output variable, Y, is a linear function of the weighted hidden-layer outputs.

The root mean square error (RMSE) of the ANN model is defined as:

$$E = \sqrt{\frac{1}{N}\sum_{cases}\left(Y - Y_{obs}\right)^2} \quad (5)$$

where

E is the root mean square error,

N is the number of input and output cases,

Y is the predicted output, and

Y_{obs} is the observed output.

Training an ANN model typically involves minimizing the RMSE by adjusting the model weights and bias terms. Usually, training is accomplished using a nonlinear multivariate optimization algorithm. The backpropagation algorithm (or gradient descent) is commonly used in many training applications.

Similar to the NRCS equations, separate ANN models were created for each forecast site and for each forecast date from January to June (*tables 8–12*). The length in years of the time periods used to create the models varied because of data availability. However, none of the time periods were begun earlier than 1960.

Table 8. Nonautoregressive artificial neural network Williamson River flow forecast models.

[R^2, coefficient of determination; vol., volume; P, precipitation; S, snow-water equivalent; F, flow; Will. R., Williamson River; Index, averaged climate sites; (1), one-year lag; Res., Reservoir; Lk., Lake; Ck., Creek; Spr., Spring]

April-September models								
Forecast date	**Training R^2**	**Testing R^2**	**Dependent variable**	**Independent variables (month-site-parameter)**				
Jan. 1	0.58	0.59	Apr.–Sep. flow vol.	Dec.-Will. R.-F	Jan.-Index-S	Oct.-Gerber Res.-P	Nov.-Chemult-P	Jan.(1)-Will. R.-F
				Jan.-Diamond Lk.-S				
Feb. 1	0.62	0.62	Apr.–Sep. flow vol.	Jan.-Will. R.-F	Feb.-Index-S	Feb.(1)-Will. R.-F	Jan.-Crater Lk.-S	Dec.-Chilloquin-P
March 1	0.80	0.80	Apr.–Sep. flow vol.	Feb.-Will. R.-F	Mar.-Index-S	Jan.-Crater Lk.-S	Mar.(1)-Will. R.-F	Feb.-Fall R.-F
Apr. 1	0.84	0.84	Apr.–Sep. flow vol.	Mar.-Will. R.-F	Apr.-Index-S	Apr.(1)-Will. R.-F	Feb.-Crater Lk.-S	Mar.-Fall R.-F
				Jan.-Silver Ck.-S				
May 1	0.76	0.75	May–Sep. flow vol.	Apr.-Will. R.-F	May(1)-Will. R.-F	Feb.-Crater Lk.-S	Apr.-Crater Lk.-P	Apr.-Fall R.-F
				May-Annie Spr.-S				
June 1	0.87	0.85	Jun.–Sep. flow vol.	May-Will. R.-F	May(1)-Will. R.-F			
February–July models								
Forecast date	**Training R^2**	**Testing R^2**	**Dependent variable**	**Independent variables (month-site-parameter)**				
Jan. 1	0.62	0.59	Feb.–Jul. flow vol.	Dec.-Will. R.-F	Jan.-Index-S	Nov.-Crater Lk.-P	Dec.-Chemult-P	Jan.(1)-Will. R.-F
Feb. 1	0.73	0.72	Feb.–Jul. flow vol.	Jan.-Will. R.-F	Feb.-Index-S	Jan.-Crater Lk.-S	Feb.(1)-Will. R.-F	Dec.-Chiloquin-P
				Jan.-Fall R.-F				
March 1	0.87	0.86	Mar.–Jul. flow vol.	Feb.-Will. R.-F	Dec.-Chemult-P	Mar.(1)-Will. R.-F	Mar.-Index-S	Feb.-Fall R.-F
				Feb.-Crater Lk.-S				
Apr. 1	0.93	0.92	Apr.–Jul. flow vol.	Mar.-Will. R.-F	Dec.-Gerber Res.-P	Apr.(1)-Will. R.-F	Mar.-Crater Lk.-S	Apr.-Index-S
				Feb.-Diamond Lk.-S				
May 1	0.89	0.85	May–Jul. flow vol.	Apr.-Will. R.-F	Apr.-Crater Lk.-P	Apr.-Chiloquin-P	Apr.-Diamond Lk.-S	May(1)-Will. R.-F
				May-Crater Lk.-S				
June 1	0.91	0.92	Jun.–Jul. flow vol.	May-Will. R.-F	Apr.-Diamond Lk.-S	May(1)-Will. R.-F	Apr.-Index-P	Mar.-Crater Lk.-S

Table 9. Nonautoregressive artificial neural network Sprague River flow forecast models.

[R^2, coefficient of determination; vol., volume; P, precipitation; S, snow-water equivalent; F, flow; Index, averaged climate sites; OND, sum of October, November, and December data; (1), one-year lag; Co., Corrals; Sum., Summer; Fl., Flat; Res., Reservoir; Mtn., Mountain; NA, not available]

April-September models								
Forecast date	**Training R^2**	**Testing R^2**	**Dependent variable**	**Independent variables (month-site parameter)**				
Jan. 1	0.78	0.78	Apr.–Sep. flow vol.	Jan(1)-Sprague R.-F	Jan.-Quartz Mtn.-S	Dec.-Sprague R.-F	Dec.-Chiloquin-P	Nov.-Round Grove-P
Feb. 1	0.76	0.77	Apr.–Sep. flow vol.	Jan.-Sprague R.-F	Feb(1)-Sprague R.-F	OND-Index-P	Feb.-Index-S	Nov.-Chiloquin-P
March 1	0.76	0.75	Apr.–Sep. flow vol.	Feb.-Sprague R.-F	Mar.-Index-S	Dec.-Chiloquin-P	Feb.-Finley Co.-S	Mar.-Sycan Fl.-S
				Mar.-Sum. Rim-S				
April 1	0.82	0.81	Apr.–Sep. flow vol.	Mar.-Sprague R.-F	Apr(1)-Sprague R.-F	Apr.-Index-S	OND-Index-P	Mar.-Finley Co.-S
				Feb.-Finley Co.-S				
May 1	0.82	0.83	May–Sep. flow vol.	Apr.-Sprague R.-F	May(1)-Sprague R.-F	Apr.-Index-P	Apr.-Sycan Fl.-S	Mar.-Sum. Rim-S
June 1	NA	NA	Jun.–Sep. flow vol.					
February–July models								
Forecast date	**Training R^2**	**Testing R^2**	**Dependent variable**	**Independent variables (month-site parameter)**				
Jan. 1	0.71	0.72	Feb.–Jul. flow vol.	Dec.-Sprague R.-F	Jan.-Quartz Mtn.-S	Dec.-Chiloquin-P	Dec.-Round Grove-P	Nov.-Chiloquin-P
				Feb(1)-Index-S				
Feb. 1	0.76	0.76	Feb.–Jul. flow vol.	Jan.-Sprague R.-F	Nov.-Round Grove-P	OND-Index-P	Feb.-Index-S	
March 1	0.68	0.68	Mar.–Jul. flow vol.	Feb.-Sprague R.-F	Mar.-Crazyman Fl.-S	Dec.-Round Grove-P	Mar(1)-Crazyman Fl.-S	
April 1	0.82	0.81	Apr.–Jul. flow vol.	Mar.-Sprague R.-F	Apr.-Index-S	Dec.-Chiloquin-P	Feb.-Finley Co.-S	
May 1	0.82	0.91	May–Jul. flow vol.	Apr.-Sprague R.-F	Apr.-Gerber Res.-P	Mar(1)-Crazyman Fl.-S	May-Finley Co.-S	
June 1	NA	NA	Jun.–Jul. flow vol.					

Table 10. Nonautoregressive artificial neural network Upper Klamath Lake net inflow forecast models.

[R^2, coefficient of determination; vol., volume; P, precipitation; S, snow-water equivalent; F, flow; N, net inflow; Index, averaged climate sites; OND, sum of October, November, and December data; Lk., Lake; Ck., Creek; Di., Divide; Pr., Prairie; (1), one-year lag; Res., Reservoir; NA, not available]

April-September models								
Forecast date	**Training R^2**	**Testing R^2**	**Dependent variable**	**Independent variables (month-site-parameter)**				
Jan. 1	0.65	0.64	Apr.–Sep. flow vol.	Dec.-U. Klamath Lk.-N	Dec.-Fall R.-F	Jan.-Index-S	OND-Chiloquin-P	Dec.-Gerber Res.-P
				Jan.-Crater Lk.-S				
Feb. 1	0.81	0.85	Apr.–Sep. flow vol.	Jan.-U. Klamath Lk. N	Feb.-Index S	OND-Chiloquin P	Jan.-Fall R. F	Feb.(1)-Index S
				Feb.-Fourmile Lk. S				
March 1	0.84	0.83	Apr.–Sep. flow vol.	Jan.(1)-Chemult-S	Feb.-U. Klamath Lk.-N	Feb.-Billie Ck. Di-S	Feb.-Fall R.-N	Mar.-Index-S
				Dec.-Round Grove-P				
April 1	0.89	0.89	Apr.–Sep. flow vol.	Mar.-U. Klamath Lk.-N	OND-Index-P	Mar.-Index-S	Feb.(1)-Index-S	Dec.-Gerber Res.-P
				Mar.-Fourmile Lk.-S				
May 1	0.93	0.83	May–Sep. flow vol.	Apr.-U. Klamath Lk.-N	Dec.-Howard Pr.-P	Mar.-Index-S	Mar.-Billie Ck. Di.-S	
June 1	NA	NA						
February-July models								
Forecast date	**Training R^2**	**Testing R^2**	**Dependent variable**	**Independent variables (month-site-parameter)**				
Jan. 1	0.67	0.49	Feb.–Jul. flow vol.	Jan.-Index-S	Dec.-U. Klamath Lk.-N	Dec.-Gerber Res.-P	Jan.(1)-U. Klamath Lk.-N	Nov.-Crater Lk.-P
				Jan.-Fourmile Lk.-S				
Feb. 1	0.62	0.62	Feb.–Jul. flow vol.	Jan.-U. Klamath Lk.-N	Feb.-Index-S	Jan.-Fourmile Lk.-S	Dec.-Gerber Res.-P	Dec.-Howard Pr.-P
				Jan.-Fall R.-F				
March 1	0.83	0.82	Mar.–Jul. flow vol.	Feb.-U. Klamath Lk.-N	Mar.-Index-S	Mar.(1)-Index-S	Dec.-Crater Lk.-P	Feb.-Fourmile Lk.-S
				Jan.-Crater Lk.-S				
April 1	0.77	0.77	Apr.–Jul. flow vol.	Mar.-U. Klamath Lk.-N	Mar.-Fall R.-F	OND-Index-P	Mar.-Index-S	Mar.(1)-Index-S
				Dec.-Howard Pr.-P				
May 1	0.70	0.71	May–Jul. flow vol.	Apr.-U. Klamath Lk.-N	Apr.-Fall R.-F	Mar.(1)-Index-S	Mar.-Sevenmile-S	Jan.-Fourmile Lk.-S
June 1	0.80	0.79	Jun.–Jul. flow vol.	May-Fall R.-F	Mar.(1)-Index-S	Feb.-Sevenmile-S	May-U. Klamath Lk.-N	Jan.-Fourmile Lk.-S

Table 11. Nonautoregressive artificial neural network Clear Lake Reservoir net inflow forecast models.

[R^2, coefficient of determination; vol., volume; P, precipitation; S, snow-water equivalent; N, net inflow; Index, averaged climate sites; OND, sum of October, November, and December data; Lk., Lake; Ck., Creek; Di., Divide; Pr., Prairie; (1), one-year lag; Fl., Flat; Res., Reservoir; NA, not available]

April-September models								
Forecast date	**Training R^2**	**Testing R^2**	**Dependent variable**	**Independent variables (month-site-parameter)**				
Jan. 1	0.61	0.60	Apr.–Sep. flow vol.	Jan.-Cedar Pass-S	Nov.-Lakeview-P	Dec.-Lakeview-P	Dec.-Clear Lk. Res.-N	
Feb. 1	0.68	0.68	Apr.–Sep. flow vol.	Jan.-Gerber Res.-P	Nov.-Gerber Res.-P	Feb.-Cedar Pass-S	Jan.-Clear Lk. Res.-N	Dec.-Lakeview-P
March 1	0.87	0.86	Apr.–Sep. flow vol.	Mar.-Cedar Pass-S	Jan.-Gerber Res.-P	Nov.-Gerber Res.-P	Dec.-Lakeview-P	Feb.-Strawberry-S
April 1	0.71	0.71	Apr.–Sep. flow vol.	Jan.-Gerber Res.-P	Nov.-Gerber Res.-P	Apr.-Cedar Pass-S	Mar.-Strawberry-S	Jan.-Lakeview-P
				Mar.-Clear Lk. Res.-N				
May 1	0.51	0.49	May–Sep. flow vol.	Apr.-Strawberry-S	Mar.-Gerber Res.-P	Apr.-Cedar Pass-S	Feb.-Crowder Fl.-S	May-Cedar Pass-S
				Mar.-Clear Lk. Res.-N				
June 1	NA	NA						
February-July models								
Forecast date	**Training R^2**	**Testing R^2**	**Dependent variable**	**Independent variables (month-site-parameter)**				
Jan. 1	0.43	0.41	Feb.–July flow vol.	Feb.(1)-Index-S	Dec.-Lakeview-P	OND-Index-P	Jan.(1)-Clear Lk. Res.-N	
Feb. 1	0.69	0.68	Feb.–July flow vol.	OND-Index-P	Feb.-Index-S	Mar.(1)-Index-S	Jan.-Lakeview-P	Feb.-Crowder Fl.-S
				Feb.(1)-Clear Lk. Res.-N				
March 1	0.61	0.60	Mar.–July flow vol.	Mar(1)-Index-S	Jan.-Index-P	Mar.-Index-S	Mar.-Crowder Fl.-S	Feb.-Lakeview-P
				Feb.-Clear Lk. Res.-N				
April 1	0.70	0.70	Apr.–July flow vol.	Mar.(1)-Index-S	Apr.-Index-S	Mar.-Index-P	Mar.-Clear Lk. Res.-N	Mar.-Strawberry-S
				Feb.-Dog Hollow-S				
May 1	0.52	0.51	May–July flow vol.	Apr.-Index-P	Apr.-Dog Hollow-S	Mar.(1)-Index-S	Apr.(1)-Index-P	Apr.-Clear Lk. Res.-N
June 1	NA	NA						

Table 12. Nonautoregressive artificial neural network Gerber Reservoir net inflow forecast models.

[R^2, coefficient of determination; vol., volume; P, precipitation; S, snow-water equivalent; F, flow; N, net inflow; Index, averaged climate sites; (1), one-year lag; Fl., Flat; Mtn., Mountain; Res., Reservoir; NA, not available]

April-September models								
Forecast date	**Training R^2**	**Testing R^2**	**Dependent variable**	**Independent variables (month-site-parameter)**				
Jan. 1	0.39	0.39	Apr.–Sep. flow vol.	Dec.-Fall R.-F	Dec.-Round Grove-P	Jan.-Index-S	Mar.(1)-Index-S	Dec.-Gerber Res.-N
Feb. 1	0.50	0.50	Apr.–Sep. flow vol.	Nov.-Round Grove-P	Feb.-Index-S	Dec.-Index-P	Feb.-Strawberry-S	Feb(1)-Index-S
				Jan.-Gerber Res.-N				
March 1	0.57	0.56	Apr.–Sep. flow vol.	Nov.-Gerber Res.-P	Oct.-Round Grove-P	Feb.-Index-S	Feb.-Strawberry-S	Mar.-Dog Hollow-S
April 1	0.62	0.61	Apr.–Sep. flow vol.	Dec.-Round Grove-P	Dec.-Index-P	Mar.-Gerber Res.-N	Apr.-Index-S	Apr.(1)-Gerber Res.-P
				Mar.-Dog Hollow-S				
May 1	0.39	0.38	May–Sep. flow vol.	Apr.-Gerber Res.-N	Mar.-Strawberry-S	Apr.-Index-S	Mar.-Gerber-S	Feb.-Crowder Fl.-S
June 1	NA	NA						
February-July models								
Forecast date	**Training R^2**	**Testing R^2**	**Dependent variable**	**Independent variables (month-site-parameter)**				
Jan. 1	0.53	0.52	Feb.–Jul. flow vol.	Dec.-Fall R.-F	Dec.-Round Grove-P	Jan.-Index-S	Mar.(1)-Index-S	Dec.-Gerber Res.-N
Feb. 1	0.93	0.65	Feb.–Jul. flow vol.	Nov.-Round Grove-P	Feb.-Index-S	Dec.-Index-P	Feb.-Strawberry-S	Feb.(1)-Index-S
				Jan.-Gerber Res.-N				
March 1	0.81	0.82	Mar.–Jul. flow vol.	Dec.-Round Grove-P	Mar.-Index-S	Dec.-Index-P	Mar.(1)-Index-S	Dec.-Gerber Res.-N
April 1	0.80	0.79	Apr.–Jul. flow vol.	Dec.-Round Grove-P	Dec.-Index-P	Mar.-Gerber Res.-N	Apr.-Index-S	Apr.(1)-Gerber Res.-P
				Mar.-Dog Hollow-S				
May 1	0.36	0.36	May–Jul. flow vol.	Apr.-Index-P	Apr.-Gerber Res.-N	Feb.-Crowder Fl.-S	Mar.-Quartz Mtn.-S	Feb.(1)-Index-S
				Apr.-Crowder Fl.-S				
June 1	0.25	0.24	Jun.–Jul. flow vol.	May-Gerber Res.-N	Mar.-Index-S			

Autoregressive Artificial Neural Networks

The autoregressive artificial neural network models forecast seasonal flow volumes using input variables that are derived from components of the flow time series of the forecast site, rather than recent snow or precipitation conditions. In the past, an autoregressive approach to forecasting flow was not practical because real-time flow data were not available to the forecaster. However, as more and more real-time flow data from USGS gages become available on the Internet, these types of models become more practical and feasible.

The autoregressive models presented here are based on the flow or net inflow for the period of 1979 to 2003 and operate using a weekly time step. Weekly 24-year time series (1979–2003) were assembled for each of the five Klamath forecast sites. For the Williamson and Sprague River sites, daily flow data were simply aggregated into weekly values. Weekly Sprague River flows for 1979–2003 are shown in *figure 6*. However, for the Upper Klamath Lake, Clear Lake Reservoir, and Gerber Reservoir sites, monthly net inflow volumes were disaggregated into weekly net inflow volumes.

Input and output variables for the models were created from further processing of the weekly flow and net inflow time series as shown in *figure 7*. Dependent and independent model variables for all five forecast sites are shown in *tables 13–17*. A set of six autoregressive ANN models were created for each forecast site for six forecast dates from early January to early June. The models shown in these tables for each forecast date are designed to forecast the 6-month, or 26-week, period from approximately April 1 to approximately September 30.

The dependent, or output, variable in the models, forecast_site##, is a weekly time series containing the sum flow volume of the previous number of weeks for any given week. (In *figure 7*, this variable is labeled "SPRA26".) This variable for the January, February, March, and April forecast models is the sum flow volume of the previous 26 weeks. However, the dependent variables for the May and June models use the sum flow of 22 and 17 weeks, respectively. Those models are forecasting only 22 (May 1–September 30) and 17 (June 1–September 30) weeks, respectively, into the future.

The independent variables, or input variables, for the models are derived from components of the flow time series of the forecast site. These include real-time and recent flow conditions, annual periodic conditions, and long-term climate trends.

forecast_site_S## is a weekly time series that represents the annual periodic component of the flow time series. Flow for every week of the year is the average of all flows for that week from every year in the record.

forecast_siteD##, is a weekly time series containing the difference between the first and last week of the 26-week accumulated flow time series.

*SINETHETA i*s a fitted synthetic sine function based on the weekly flow time series for each forecast site, and is a representation of interannual wet and dry trend in the time series.

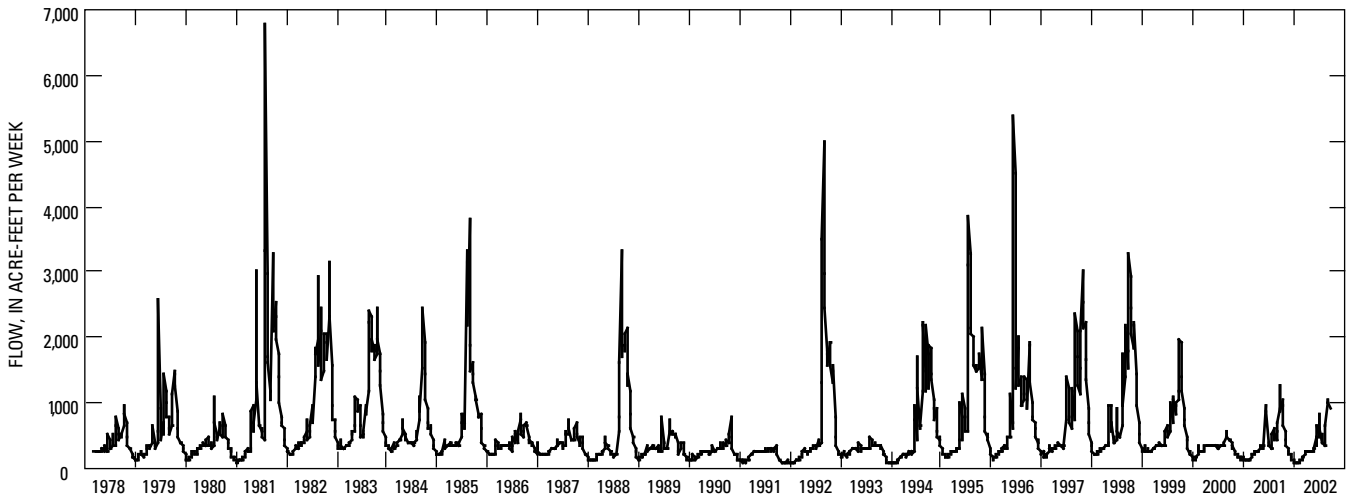


Figure 6. Weekly flow at Sprague River near Chiloquin, Oregon (11501000), 1979–2003.

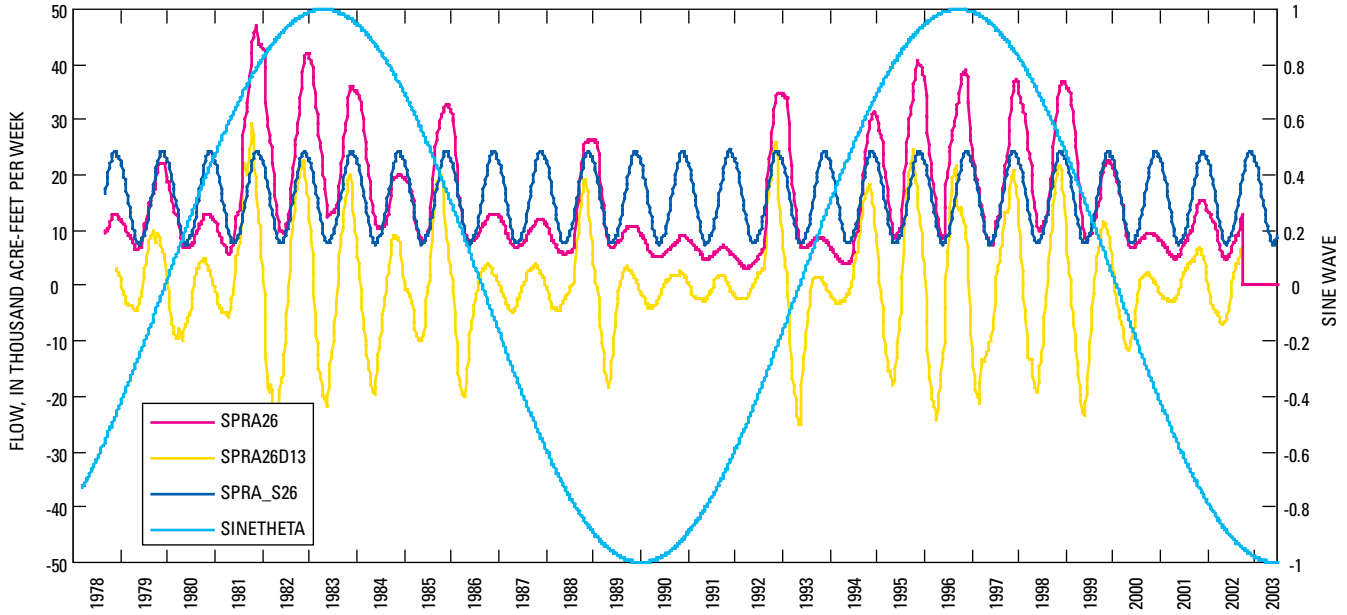


Figure 7. Sprague River autoregressive artificial neural network forecast model variables, 1979–2003.

Table 13. Autoregressive artificial neural network Williamson River flow forecast models.

[R^2, coefficient of determination; (AS), approximate period from April 1–September 30; (MS), approximate period from May 1–September 30; (JS), approximate period from June 1–September 30; (##), lag of ## weeks; WILL##, accumulated flow for the previous ## weeks; SINETHETA, synthetic sine wave created from weekly flow time series; WILL_S##, accumulated flow standard; WILL26D13(##), difference between first and last week of 26-week accumulated flow period]

Approximate forecast date	Training R^2	Testing R^2	Predicted flow volume period	Dependent variable	Independent variables			
Early Jan.	0.48	0.46	26 weeks (AS)	WILL26	SINETHETA	WILL_S26	WILL26(39)	WILL26D13(58)
Early Feb.	0.47	0.47	27 weeks (AS)	WILL26	SINETHETA	WILL_S26	WILL26(35)	WILL26D13(54)
Early March	0.45	0.43	28 weeks (AS)	WILL26	SINETHETA	WILL_S26	WILL26(30)	WILL26D13(40)
Early April	0.51	0.21	29 weeks (AS)	WILL26	SINETHETA	WILL_S26	WILL26(26)	WILL26D13(45)
Early May	0.74	0.75	22 weeks (MS)	WILL22	SINETHETA	WILL_S22	WILL26D13(42)	WILL26D13(55)
Early June	0.84	0.75	17 weeks (JS)	WILL17	WILL_S17	WILL17(17)	WILL26D13(37)	WILL26D13(50)

Table 14. Autoregressive artificial neural network Sprague River flow forecast models.

[R^2, coefficient of determination; (AS), approximate period from April 1–September 30; (MS), approximate period from May 1–September 30; (JS), approximate period from June 1–September 30; (##), lag of ## weeks; SPRA##, accumulated flow for the previous ## weeks; SINETHETA, synthetic sine wave created from weekly flow time series; SPRA_S##, Accumulated flow standard; and SPRA26D13(##), difference between first and last week of 26-week accumulated flow period]

Approximate forecast date	Training R^2	Testing R^2	Predicted flow volume period	Dependent variable	Independent variables			
Early Jan.	0.54	0.56	26 weeks (AS)	SPRA26	SINTHETA	SPRA_S26	SPRA26(39)	SPRA26D13(58)
Early Feb.	0.54	0.56	27 weeks (AS)	SPRA26	SINTHETA	SPRA_S26	SPRA26(35)	SPRA26D13(54)
Early March	0.66	0.59	28 weeks (AS)	SPRA26	SINTHETA	SPRA_S26	SPRA26(30)	SPRA26D13(49)
Early April	0.66	0.67	29 weeks (AS)	SPRA26	SPRA_S26	SPRA26(26)	SPRA26D13(45)	
Early May	0.64	0.61	22 weeks (MS)	SPRA22	SPRA_S22	SPRA22(22)	SPRA26D13(42)	SPRA26D13(55)
Early June	0.81	0.76	17 weeks (JS)	SPRA17	SPRA_S17	SPRA17(17)	SPRA26D13(37)	SPRA26D13(50)

Table 15. Autoregressive artificial neural network Upper Klamath Lake net inflow forecast models.

[R^2, coefficient of determination; (AS), approximate period from April 1–September 30; (MS), approximate period from May 1–September 30; (JS), approximate period from June 1–September 30; (##), lag of ## weeks; UKL##, accumulated flow for the previous ## weeks; SINETHETA, synthetic sine wave created from weekly net inflow time series; UKL_S##, accumulated flow standard; UKL26D13(##), difference between first and last week of 26-week accumulated flow period; YEARDAY, Julian calendar day]

Approximate forecast date	Training R^2	Testing R^2	Predicted flow volume period	Dependent variable	Independent variables			
Early Jan.	0.47	0.47	26 weeks (AS)	UKL26	SINTHETA	UKL_S26	UKL26(39)	YEARDAY
Early Feb.	0.46	0.48	27 weeks (AS)	UKL26	SINTHETA	UKL_S26	UKL26(35)	YEARDAY
Early March	0.42	0.49	28 weeks (AS)	UKL26	UKL_S26	UKL26(30)		
Early April	0.75	0.69	29 weeks (AS)	UKL26	UKL_S26	UKL26(26)	UKL26D13(46)	
Early May	0.74	0.71	22 weeks (MS)	UKL22	UKL_S22	UKL22(22)	UKL26D13(40)	
Early June	0.83	0.78	17 weeks (JS)	UKL17	UKL_S17	UKL17(17)	UKL26D13(37)	UKL26D13(50)

Table 16. Autoregressive artificial neural network Clear Lake Reservoir net inflow forecast models.

[R^2, coefficient of determination; (AS), approximate period from April 1–September 30; (MS), approximate period from May 1–September 30; (JS), approximate period from June 1–September 30; (##), lag of ## weeks; CLE##, accumulated flow for the previous ## weeks; SINETHETA, synthetic sine wave created from weekly net inflow time series; CLE_S##, accumulated flow standard; CLE26D13(##), difference between first and last week of 26-week accumulated flow period]

Approximate forecast date	Training R^2	Testing R^2	Predicted flow volume period	Dependent variable	Independent variables			
Early Jan.	0.44	0.27	26 weeks (AS)	CLE26	SINETHETA	CLE_S26	CLE26(39)	CLE26D13(58)
Early Feb.	0.5	0.24	27 weeks (AS)	CLE26	SINETHETA	CLE_S26	CLE26(35)	CLE26D13(54)
Early March	0.6	0.27	28 weeks (AS)	CLE26	SINETHETA	CLE_S26	CLE26(30)	CLE26D13(49)
Early April	0.62	0.68	29 weeks (AS)	CLE26	CLE_S26	CLE26(26)	CLE26D13(45)	
Early May	0.8	0.8	22 weeks (MS)	CLE22	CLE_S22	CLE22(22)	CLE26D13(40)	CLE26D13(55)
Early June	0.41	0.43	17 weeks (JS)	CLE17	CLE_S17	CLE17(17)	CLE26D13(37)	CLE26D13(50)

Table 17. Autoregressive artificial neural network Gerber Reservoir net inflow forecast models.

[R^2, coefficient of determination; ; (AS), approximate period from April 1–September 30; (MS), approximate period from May 1–September 30; (JS), approximate period from June 1–September 30; (##), lag of ## weeks; GER##, accumulated flow for the previous ## weeks; SINETHETA, synthetic sine wave created from weekly net inflow time series; GER_S##, accumulated flow standard; GER26D13(##), difference between first and last week of 26-week accumulated flow period]

Forecast date	Training R^2	Testing R^2	Predicted flow volume period	Dependent variable	Independent variables			
Early Jan.	0.45	0.4	26 weeks (AS)	GER26	SINETHETA	GER_S26	GER26(39)	GER26D13(58)
Early Feb.	0.45	0.38	27 weeks (AS)	GER26	SINETHETA	GER_S26	GER26(35)	GER26D13(54)
Early March	0.49	0.4	28 weeks (AS)	GER26	SINETHETA	GER_S26	GER26(30)	GER26D13(49)
Early April	0.51	0.5	29 weeks (AS)	GER26	GER_S26	GER26(26)	GER26D13(45)	
Early May	0.68	0.73	22 weeks (MS)	GER22	GER_S22	GER22(22)	GER26D13(40)	GER26D13(55)
Early June	0.81	0.55	17 weeks (JS)	GER17	SINETHETA	GER_S17	GER17(17)	GER26D13(37)

Results and Discussion

Major objectives of the study included an evaluation and comparison of different statistical modeling approaches, in addition to an evaluation of the effects of using long-term climate-trend variables to improve forecast accuracy.

Model Comparisons

The performance of the three models is shown in this section. Because these models are very different from each other and use different data sets, it is difficult to fairly compare the models strictly on their correlation coefficient values and standard errors. For example, a model may have a very high correlation coefficient, but if it is constructed from only 15 or 20 cases of data it could be just a very good fit for a narrow time period. It may perform poorly for climate periods that are very different from its own data set. Whereas other models, which are created from larger data sets but have lower correlation coefficients, may be better at capturing the full dynamics of a system. Nonetheless, some understanding of how well the models forecast flows at different sites and at different dates during the forecast season can be derived by analyzing model residual errors.

Robustness

Comparisons of the three models at each of the five sites are shown in *figures 8 through 12*. Residual error is computed as observed minus predicted flow. In each figure, model errors are based on the April 1 forecast of the 6-month flow volume from April to September. This forecast is shown because it is considered the most critical forecast for water managers and farmers. By that date, farmers need to know if projected water deliveries will be sufficiently reliable to warrant investments in seed and fertilizer.

The scatter of points in figures 8 through 12 can be used to assess model robustness. Ideally, the model residual error should be evenly distributed along the Y-axis zero line. However, figures 8 through 12 show a diagonal pattern. Many of the residuals from all three models are spread over the lower left and upper right sides of the plot. Because the residual is computed as observed minus predicted flow, this would indicate that the models have a tendency to overpredict dry years and underpredict wet years. Overall, these plots show that for the forecast month of April, the principal components regression models seem to provide more accuracy and less bias than the other two models at all five forecast sites. However, the principal components regression models did not necessarily perform better for the other forecast months.

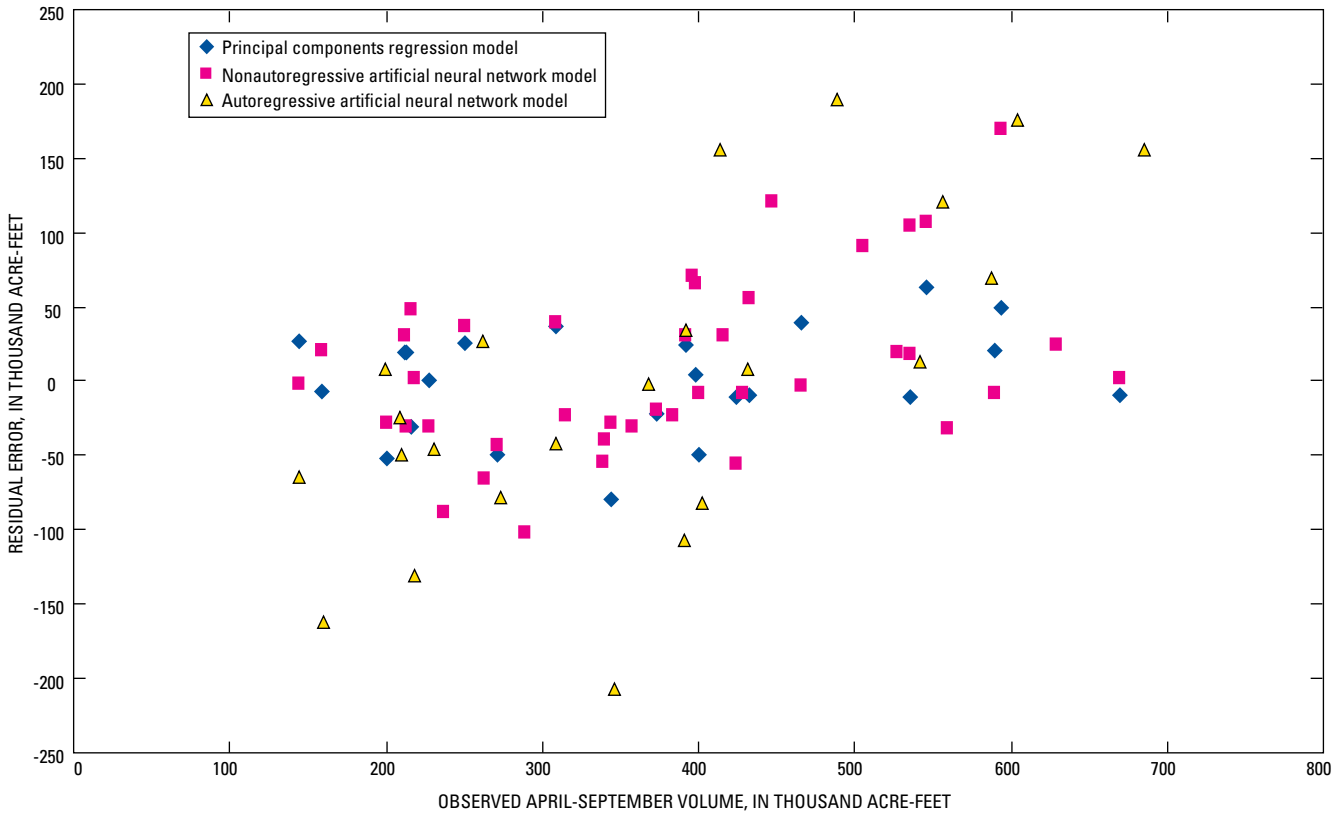


Figure 8. Comparison of residuals from three models predicting April through September Williamson River flow volume on April 1.

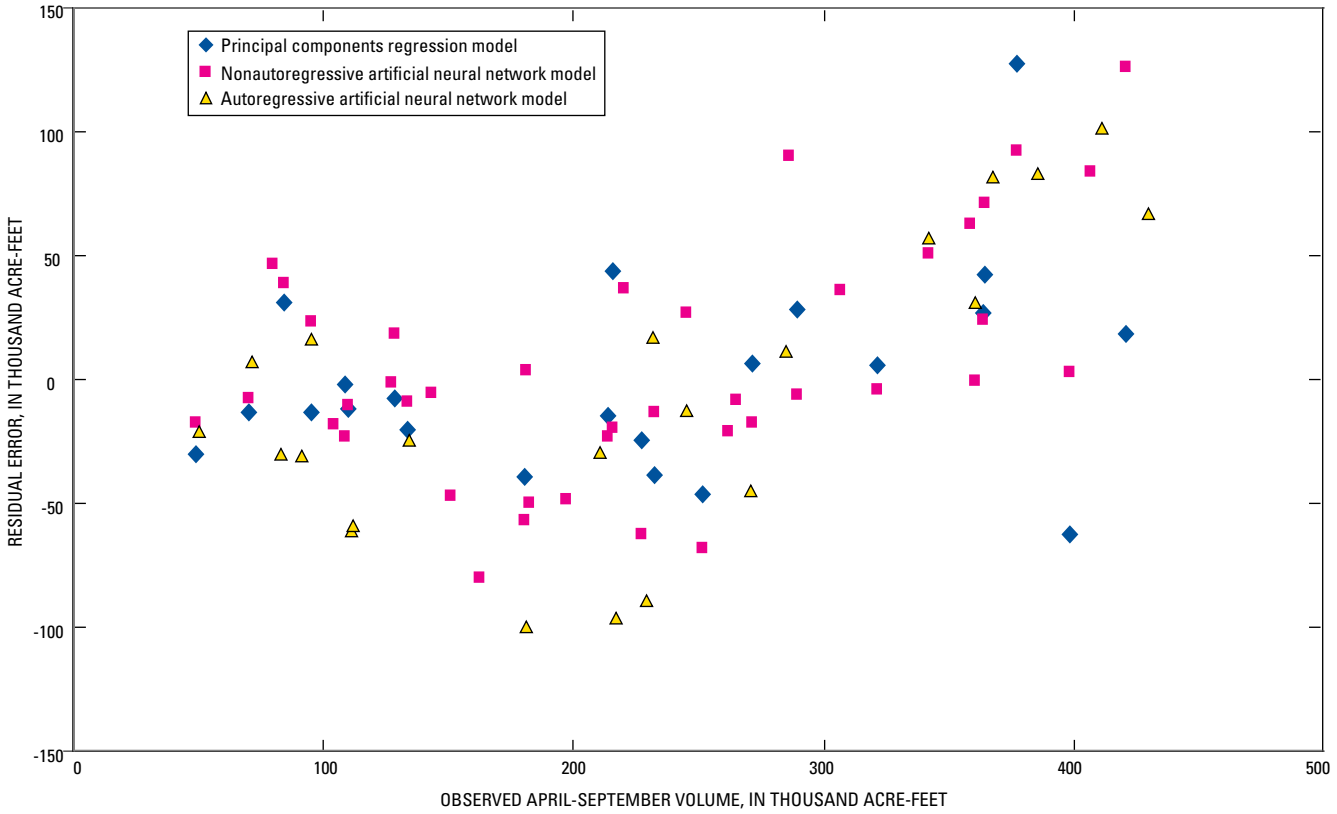


Figure 9. Comparison of residuals from three models predicting April through September Sprague River flow volume on April 1.

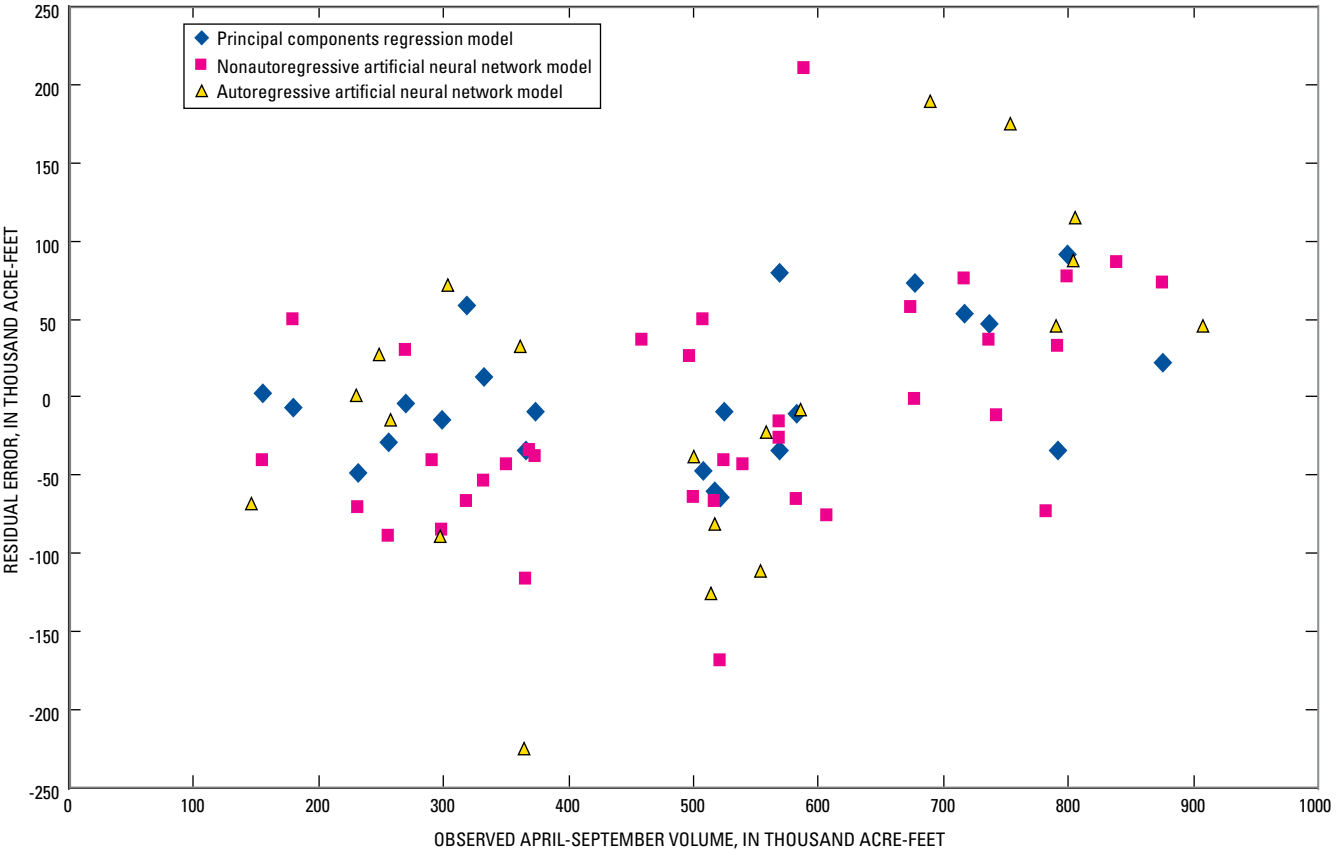


Figure 10. Comparison of residuals from three models predicting April through September Upper Klamath Lake net inflow volume on April 1.

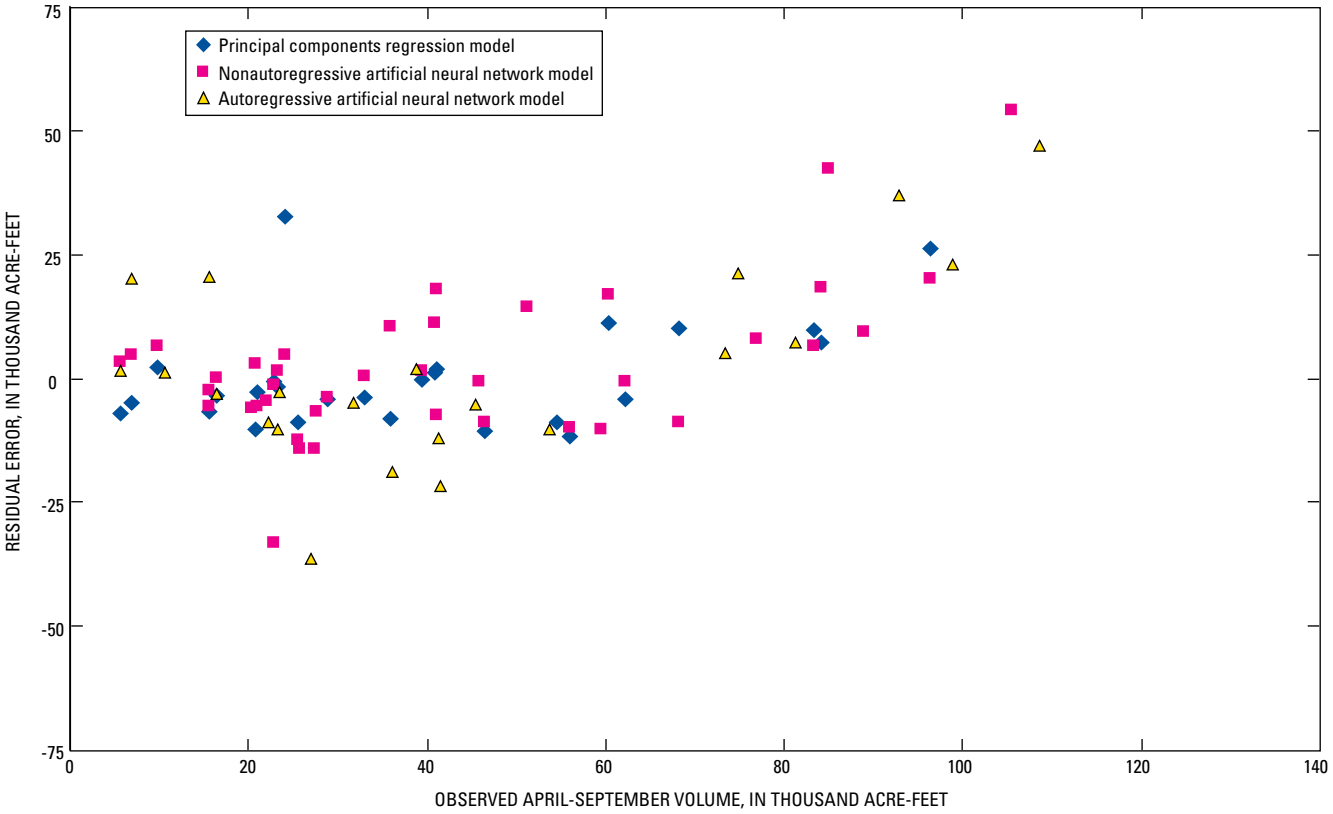


Figure 11. Comparison of residuals from three models predicting April through September Clear Lake Reservoir net inflow volume on April 1.

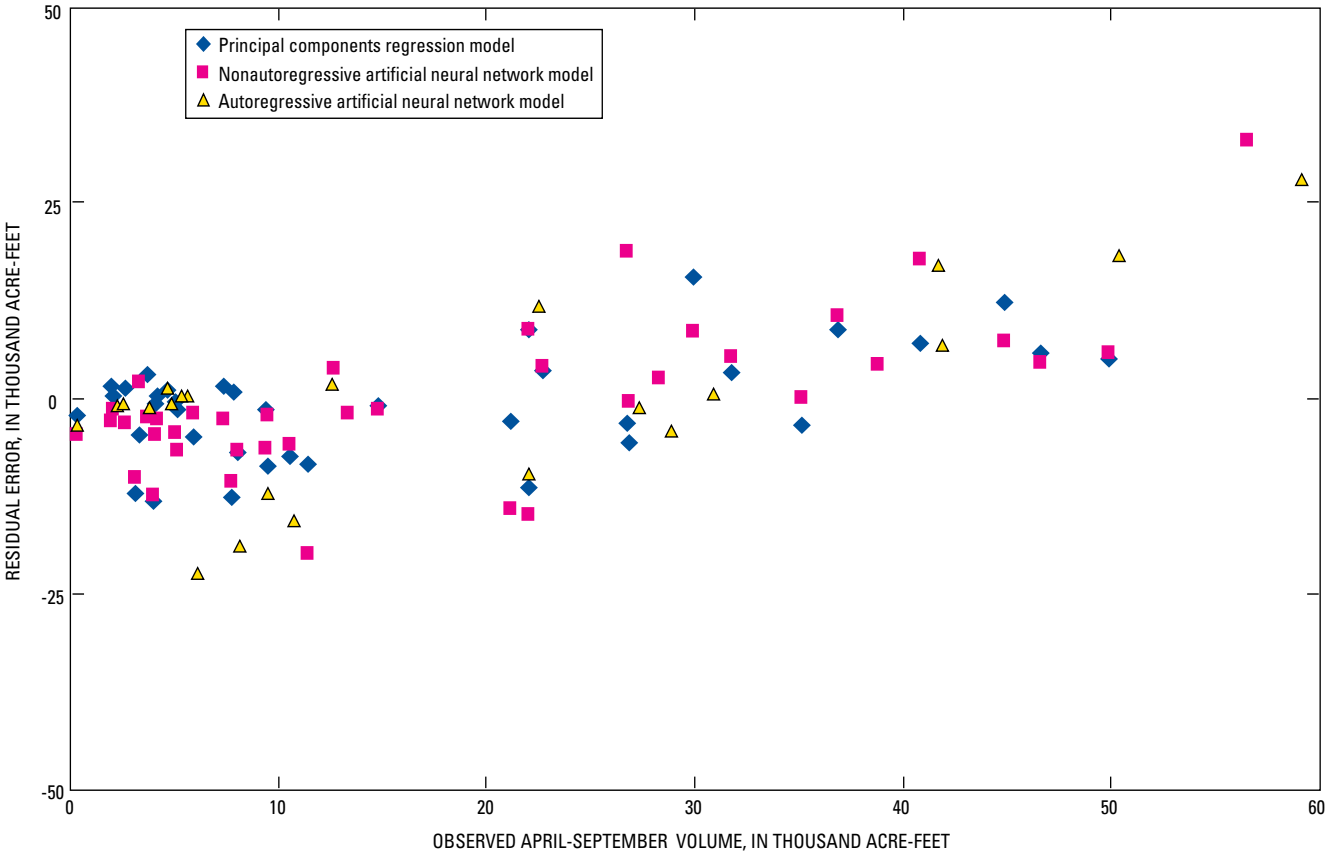


Figure 12. Comparison of residuals from three models predicting April through September Gerber Reservoir net inflow volume on April 1.

Forecast Date Variations

As the period in time between the date of a flow forecast and the actual forecast period is narrowed, generally (but not always) there is a reduction in model error. For most instances, this can be seen in *figures 13–17*, which compare the mean absolute error of the historical (1961–2002) April–September volumes forecasted by the models for each forecast date from January 1 to April 1.

For the Williamson River and Gerber Reservoir forecast sites, the principal components regression models generally, but not always, performed better than the other models (figs. 13 and 17). However, for the Sprague River forecast site, the nonautoregressive ANN models did better than the other models for January, February, and March forecasts. The principal components regression models performed better for the April forecasts (fig. 14). Similarly, for the Upper Klamath Lake forecast site, the nonautoregressive ANN models had less error than the other models for the January, February, and March forecasts. The principal components regression model performed better for the April forecast (fig. 15). For the Clear Lake Reservoir forecast site, the nonautoregressive ANN models performed better than the other models for the months of January, February, and March. However, the autoregressive ANN model performed better than the other models for the month of April (fig 16).

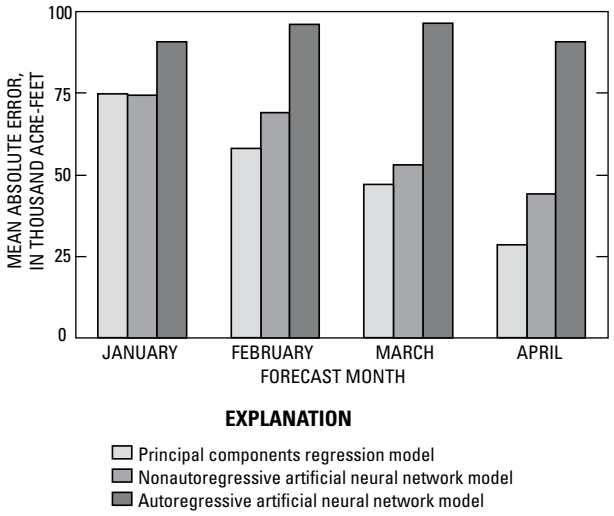


Figure 13. Model comparisons of the mean absolute error of April through September Williamson River flow volumes, 1960–2003.

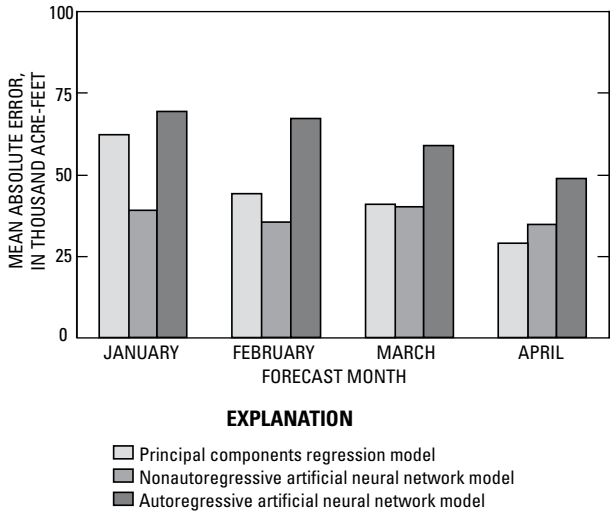


Figure 14. Model comparisons of the mean absolute error of April through September Sprague River flow volumes, 1960–2003.

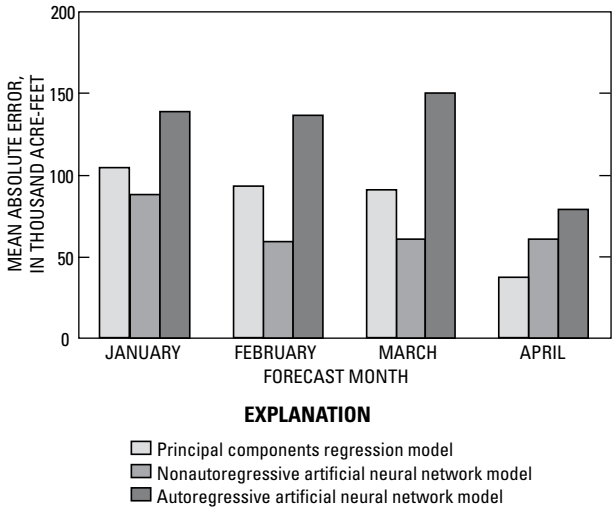


Figure 15. Model comparisons of the mean absolute error of April through September Upper Klamath Lake net inflow volumes, 1961–2003.

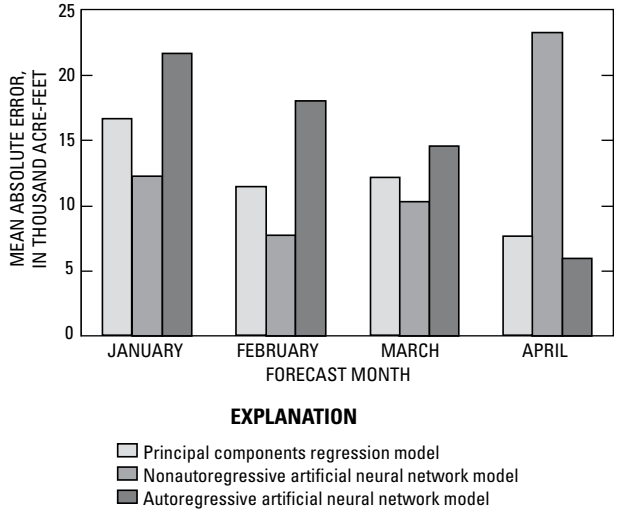


Figure 16. Model comparisons of the mean absolute error of April through September Clear Lake Reservoir net inflow volumes, 1961–2003.

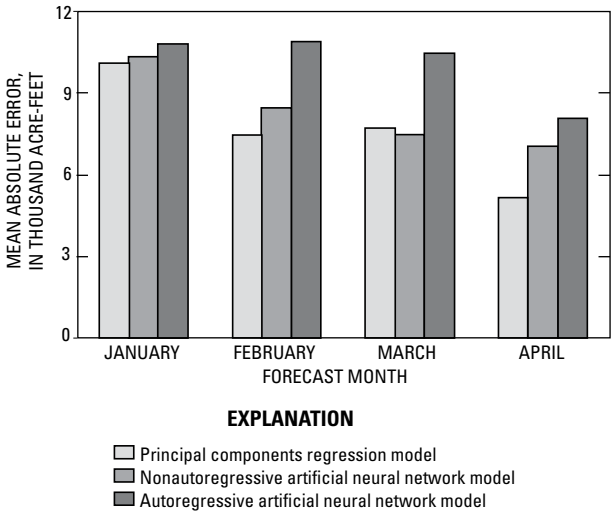


Figure 17. Model comparisons of the mean absolute error of April through September Gerber Reservoir net inflow volumes, 1960–2003.

Temporal Variations

The performance of the forecast models also can be evaluated throughout their periods of record. *Figures 18–22* compare the residual errors from the three models over time for each forecast month from January through May for the Upper Klamath Lake. Residual error is computed as observed minus predicted flow. Similar to figures 13–17, figures 18–22 also show a general reduction in error as the forecast date moves from January to May.

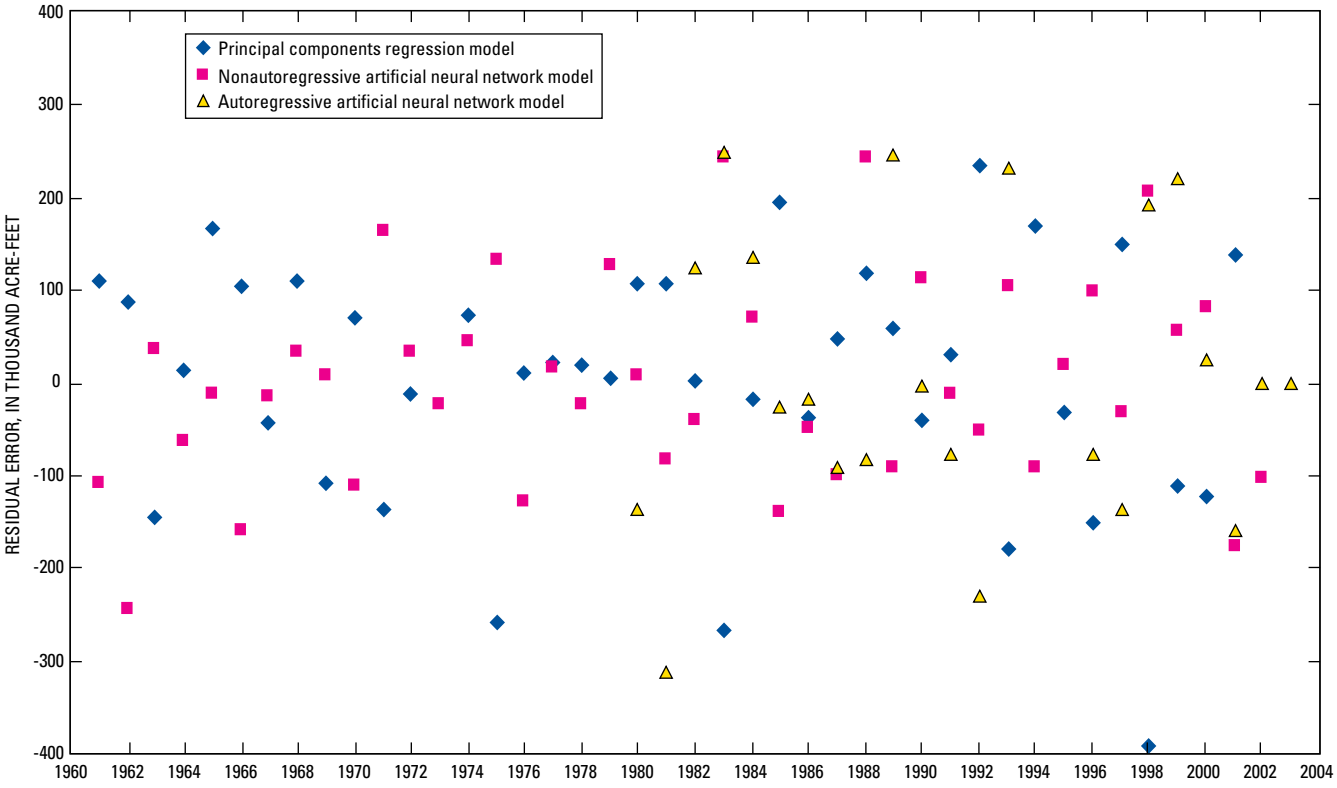


Figure 18. Comparison of model residuals for January 1 forecasts of April through September Upper Klamath Lake net inflow from 1961 to 2003.

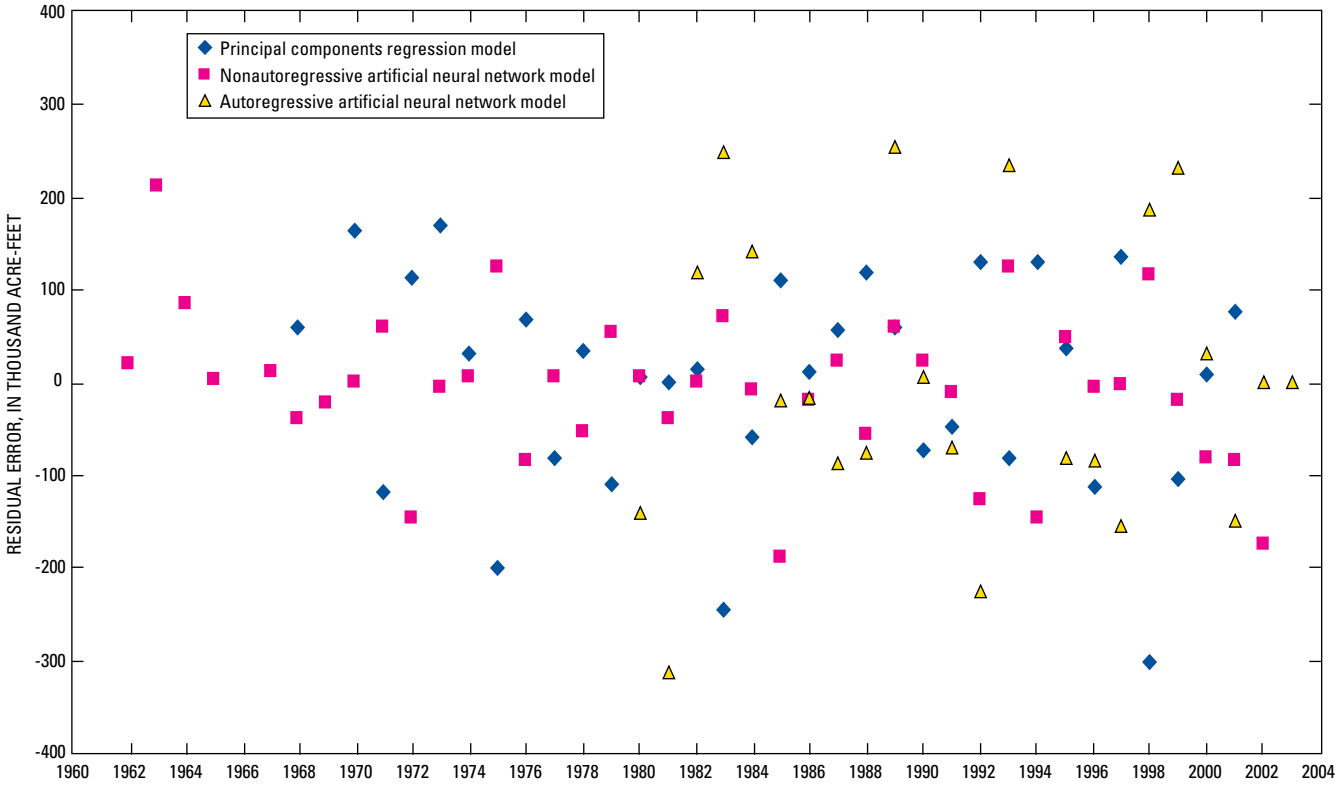


Figure 19. Comparison of model residuals for February 1 forecasts of April through September Upper Klamath Lake net inflow from 1961 to 2003.

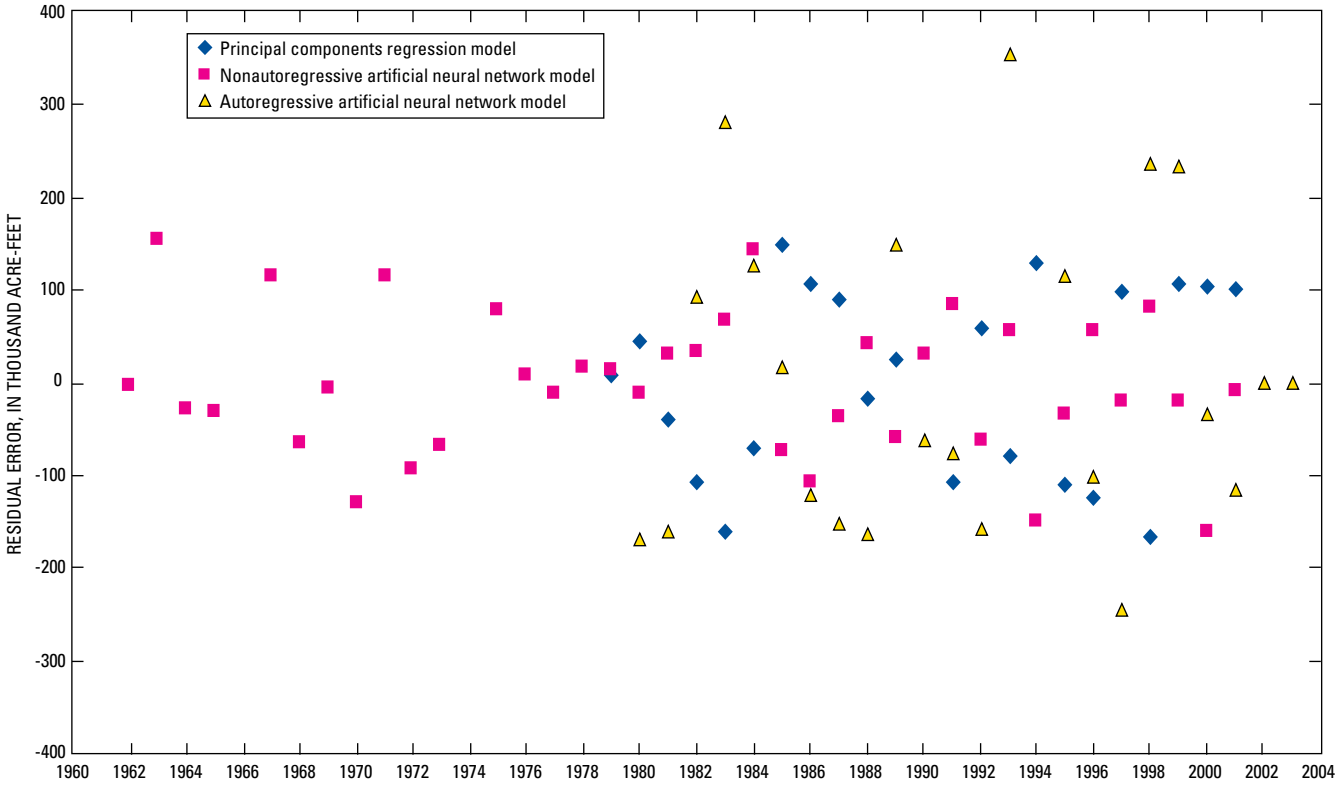


Figure 20. Comparison of model residuals for March 1 forecasts of April through September Upper Klamath Lake net inflow from 1961 to 2003.

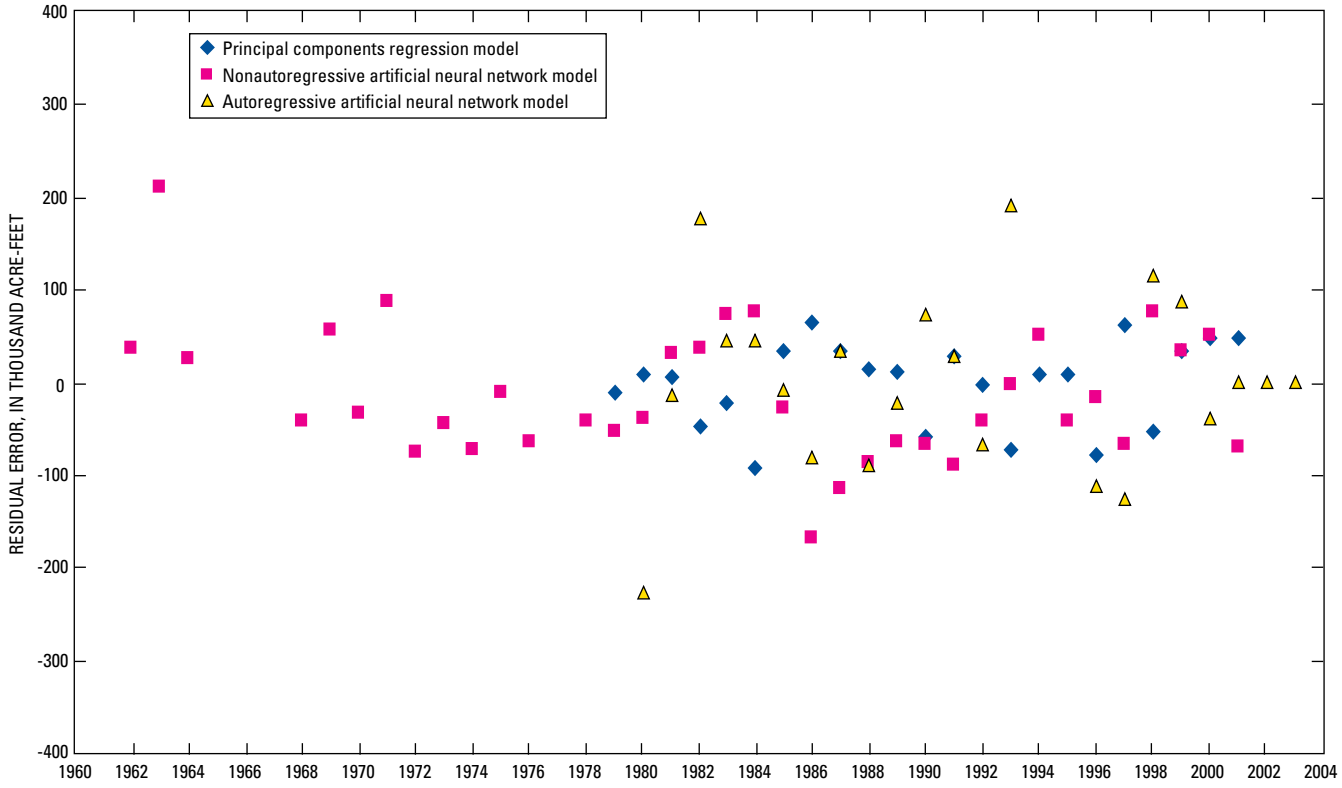


Figure 21. Comparison of model residuals for April 1 forecasts of April through September Upper Klamath Lake net inflow from 1961 to 2003.

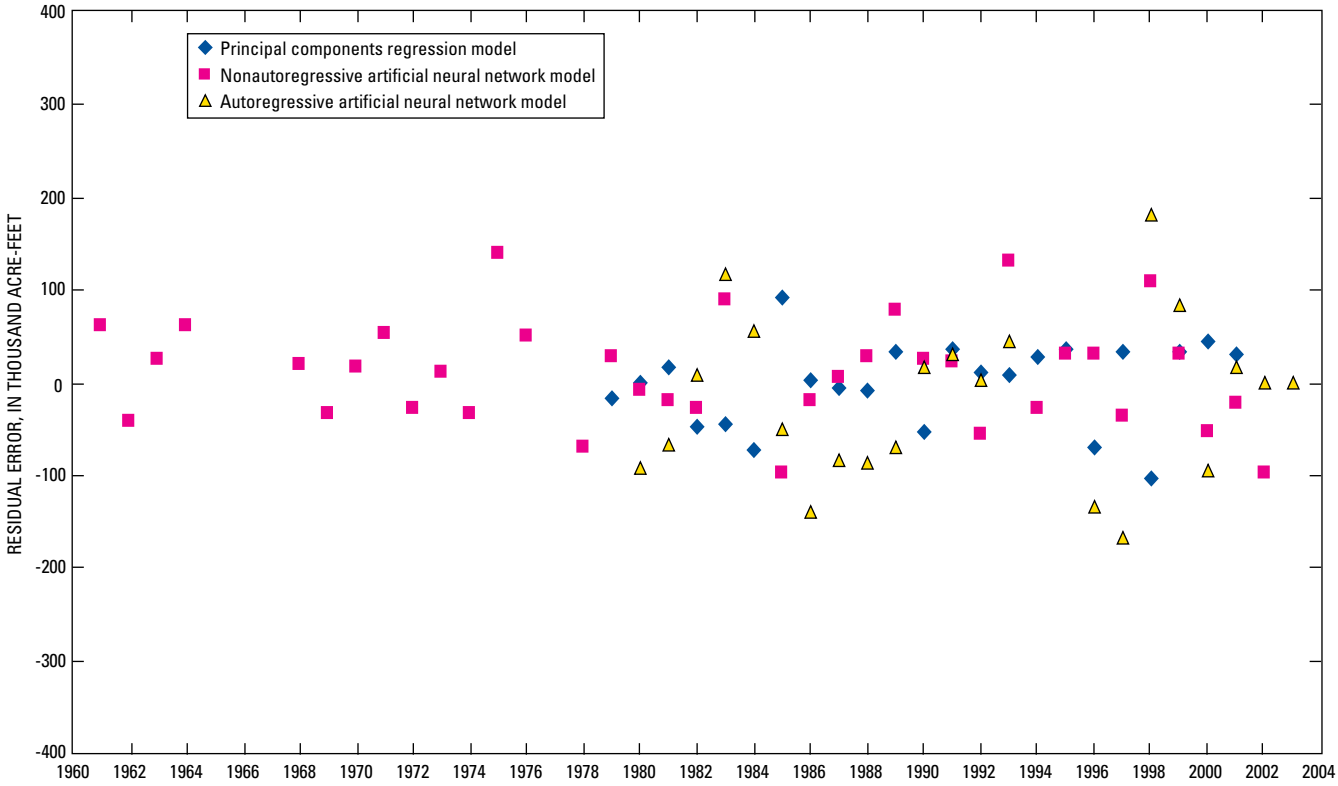


Figure 22. Comparison of model residuals for May 1 forecasts of May through September Upper Klamath Lake net inflow from 1961 to 2003.

Effect of Long-Term Climate-Trend Variables

A variety of different long-term climate-trend variables were evaluated using the principal components regression and nonautoregressive ANN models. These variables included well-water elevation, precipitation, ground-water dominated flow, and oceanic-climate indicator time series. It was not possible to evaluate these variables on the autoregressive ANN models, because those models used synthetic input variables that were derived from characteristics of the forecast flow time series. Ultimately, the only long-term climate-trend variables that were found effective in reducing model error included the ground-water dominated Fall River flow and well-water elevation time series.

Principal Components Regression Models

As described earlier, the principal components regression equations were created using a semiautomated approach. First, an optimization program creates a list of the top 20 or more optimal equations containing the highest correlation coefficients and the lowest standard errors. Then the finalized equation is manually selected from the list. The most optimal equation from the list is often, but not always, selected. An effort is made to select equations for each month during the forecast season (from January to June) that use similar climate sites for their input variables. This is done to reduce the possibility of forecasted flow volumes varying abruptly from month to month.

To adequately understand the effectiveness of using the long-term climate-trend variables (Fall River flow and water-elevation data), both the optimized and finalized equation sets were evaluated. The highest ranking optimized equations with and without the climate-trend variables were compared. Also, the finalized equations with and without the climate-trend variables were compared.

To create two sets of optimized equations, the climate-trend variables were first included in the pool of available SWE and precipitation variables used by the optimization program. The optimization program was then run a second time with the climate-trend variables removed from the pool of variables. To create two sets of finalized equations, the finalized equations (already containing the climate-trend variables) were simply recomputed with just the climate-trend variable removed.

The climate-trend variables were used in the Upper Klamath Lake and Williamson River forecast equations. *Tables 18 and 19* show the equation comparisons for these two forecast sites. Similar information is also shown in *figures 23–26*.

Table 18. Comparison of Williamson River Principal Components Regression flow forecast models with and without climate-trend variables.

[JR^2, jackknife coefficient of determination; JSE, jackknife standard error in thousand acre-feet]

April–September models								
	First ranked models from the 20 best models				Finalized models			
	With climate-trend variables		Without climate-trend variables		With climate-trend variables		Without climate-trend variables	
Forecast date	JR^2	JSE	JR^2	JSE	JR^2	JSE	JR^2	JSE
Jan. 1	0.59	89.22	0.54	94.09	0.54	94.81	0.51	98.00
Feb. 1	0.80	71.09	0.72	82.28	0.75	79.49	0.67	92.64
March 1	0.87	57.01	0.82	66.47	0.85	62.24	0.78	74.52
April 1	0.95	33.18	0.92	43.71	0.94	37.54	0.92	46.18
May 1	0.86	38.10	0.77	47.26	0.82	40.67	0.75	47.42
June 1	0.85	15.92	0.83	17.31	0.82	20.30	0.78	22.16
February-July models								
	First ranked models from the 20 best models				Finalized models			
	With climate-trend variables		Without climate-trend variables		With climate-trend variables		Without climate-trend variables	
Forecast date	JR^2	JSE	JR^2	JSE	JR^2	JSE	JR^2	JSE
Jan. 1	0.55	126.04	0.53	129.80	0.54	130.40	0.50	135.22
Feb. 1	0.77	106.34	0.64	114.28	0.70	125.00	0.62	140.40
March 1	0.88	68.16	0.84	76.42	0.85	77.18	0.80	87.96
April 1	0.95	32.28	0.93	39.51	0.94	37.97	0.92	41.80
May 1	0.83	35.73	0.77	38.28	0.78	40.56	0.74	44.09
June 1	0.88	12.07	0.88	12.07	0.78	18.35	0.78	18.43

Table 19. Comparison of Upper Klamath Lake Principal Components Regression net inflow forecast models with and without climate-trend variables.

[JR^2, jackknife coefficient of determination; JSE, jackknife standard error in thousand acre-feet]

April-September models								
	First ranked models from the 20 best models				Finalized models			
Forecast date	With climate-trend variables		Without climate-trend variables		With climate-trend variables		Without climate-trend variables	
	JR^2	JSE	JR^2	JSE	JR^2	JSE	JR^2	JSE
Jan. 1	0.53	135.59	0.52	137.14	0.52	137.30	0.48	142.17
Feb. 1	0.74	114.37	0.64	119.54	0.70	118.45	0.67	124.30
March 1	0.86	83.55	0.85	86.86	0.78	105.67	0.82	96.34
April 1	0.97	39.50	0.96	45.77	0.95	47.02	0.94	55.12
May 1	0.94	39.41	0.91	45.48	0.91	46.86	0.85	58.53
June 1	0.96	19.83	0.91	26.99	0.92	25.98	0.71	48.42
February–July models								
	First ranked models from the 20 best models				Finalized models			
Forecast date	With climate-trend variables		Without climate-trend variables		With climate-trend variables		Without climate-trend variables	
	JR^2	JSE	JR^2	JSE	JR^2	JSE	JR^2	JSE
Jan. 1	0.52	193.64	0.51	195.65	0.48	200.75	0.48	202.46
Feb. 1	0.75	152.50	0.67	159.16	0.71	162.44	0.71	164.25
March 1	0.86	99.14	0.86	99.14	0.72	139.55	0.80	117.53
April 1	0.96	37.97	0.95	41.88	0.94	49.28	0.95	44.77
May 1	0.91	38.45	0.89	43.69	0.87	46.59	0.86	48.43
June 1	0.87	20.34	0.86	21.18	0.79	25.85	0.85	21.82

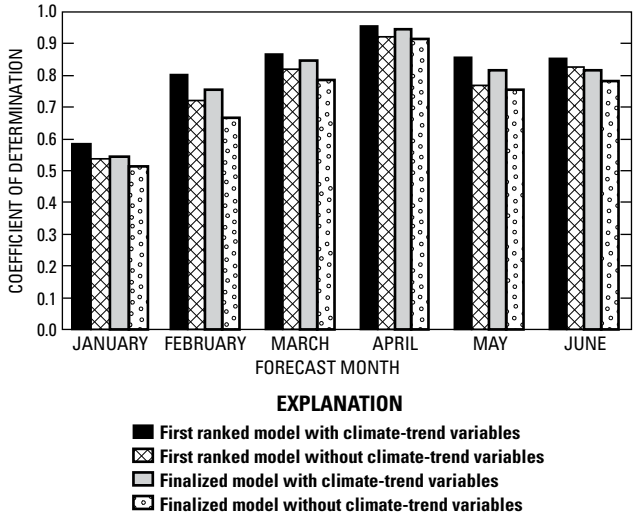


Figure 23. Comparison of finalized Williamson River principal components regression April–September flow forecast models with and without climate-trend variables.

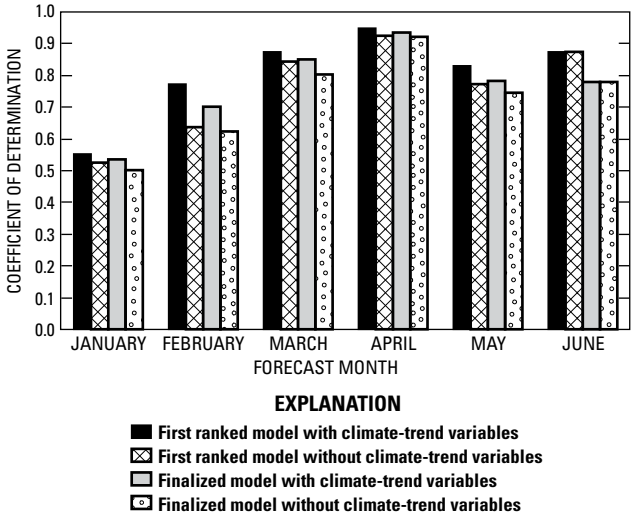


Figure 24. Comparison of finalized Williamson River principal components regression February–July flow forecast models with and without climate-trend variables.

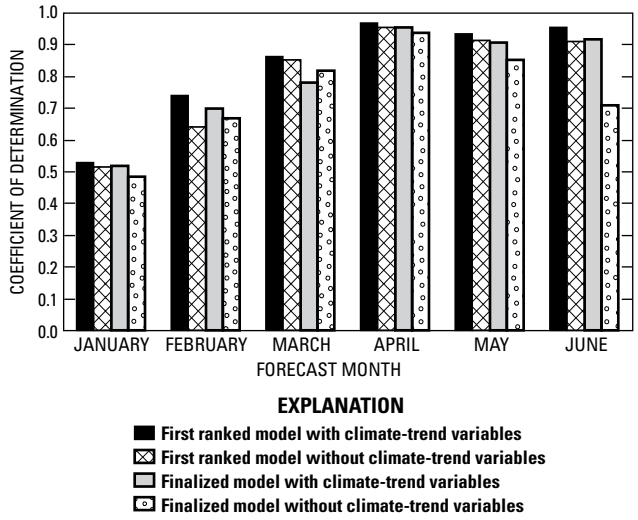


Figure 25. Comparison of finalized Upper Klamath Lake principal components regression April–September flow forecast models with and without climate-trend variables.

For the Williamson River equations, the inclusion of the climate-trend variables resulted in slightly higher correlation coefficients and lower standard errors for all comparisons for every month except the June equation in the February–July equation set. The greatest improvement was seen in the optimized February equation in the February–July equation set (*fig. 24*).

The Upper Klamath Lake equations also resulted in higher correlation coefficients and lower standard errors for most instances when the climate-trend variables were included (*table 19* and figs. 25–26). All of the optimized equations were consistently improved. However, the finalized March equation from the April–September set and the finalized March, April, and June equations from the February–July set were improved when the climate-trend variables were removed.

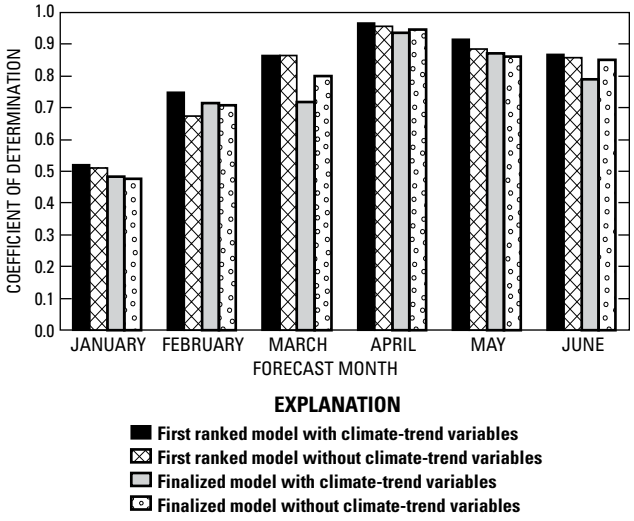


Figure 26. Comparison of finalized Upper Klamath Lake principal components regression February–July flow forecast models with and without climate-trend variables.

Nonautoregressive Artificial Neural Network Models

The nonautoregressive ANN models were created through a training process. In the initial training run, for each model, a wide pool of available input variables was used. In subsequent training runs less important variables were eliminated until the final model, containing only 5 or 6 variables, was created. For all of the models and for all five forecast sites, the climate-trend variables (Fall River flow and well-water elevation data) were included in their initial pool of input variables. However, only 13 models (all from the Williamson River, Upper Klamath Lake, and Gerber Reservoir sets) contained climate-trend variables in their final version (*table 20*). To evaluate the effectiveness of the climate-trend variables, these 13 equations were simply retrained without the climate-trend variables. Differences between the coefficients of determination (R^2) of the models with and without the climate-trend variables also are shown in *table 20*. For most instances, the climate-trend variables provided some error reduction. However, for the Upper Klamath Lake April and May models, the resulting R^2 values were higher without including the climate-trend variable.

Data Limitations in Forecast Modeling

During the development of the ANN models used in the study, the relationships between the upper Klamath Basin historical climate and flow data were analyzed to determine the ability and the extent to which current climate conditions, such as SWE and precipitation, can be used to forecast future flow conditions. To better understand the strength of climate data in predicting flow volumes, it is possible to decompose both the climate and flow time series into their annual periodic, long-term (or decadal), and chaotic components. The periodic and long-term components in a time series can be determined through spectral and statistical analyses. The chaotic component is more difficult to determine, because by its own definition it lacks a predictable cycle or pattern. It represents the ever-changing daily or weekly weather patterns embedded in the climate data. The strength of a forecast model can be measured by how well the chaotic residual component of the snow and precipitation input data can predict the future chaotic residual component of the output flow data.

For the analysis, 25 years of upper Klamath Basin daily flow, SWE, and precipitation data, collected from 1979 to 2003, were smoothed to 90-day moving average daily time series. The analysis was done for all five forecast sites; data from the Williamson River Basin are shown in *figures 27–34* as an example. SWE and precipitation data from six sites are shown in figures 27 and 28. As expected, the SWE and precipitation time series show similar timing. However, the magnitude of SWE or precipitation varies from site to site as a function of its elevation and its east-west location.

Table 20. Comparison of nonautoregressive artificial neural network forecast models with and without climate-trend variables.

[R^2, coefficient of determination]

April–September					
		With climate-trend variables		Without climate-trend variables	
Forecast site	Forecast date	Training R^2	Testing R^2	Training R^2	Testing R^2
Williamson River	March 1	0.80	0.80	0.74	0.64
Williamson River	April 1	0.84	0.84	0.77	0.81
Williamson River	May 1	0.76	0.75	0.71	0.71
Upper Klamath Lake	January 1	0.65	0.64	0.49	0.49
Upper Klamath Lake	March 1	0.84	0.83	0.81	0.85
Gerber Reservoir	January 1	0.39	0.39	0.23	0.22
February-July models					
		With climate-trend variables		Without climate-trend variables	
Forecast site	Forecast date	Training R^2	Testing R^2	Training R^2	Testing R^2
Williamson River	February 1	0.73	0.72	0.43	0.21
Williamson River	March 1	0.87	0.86	0.80	0.80
Upper Klamath Lake	February 1	0.62	0.62	0.61	0.58
Upper Klamath Lake	April 1	0.77	0.77	0.82	0.82
Upper Klamath Lake	May 1	0.70	0.71	0.76	0.76
Upper Klamath Lake	June 1	0.80	0.79	0.78	0.73
Gerber Reservoir	January 1	0.53	0.52	0.41	0.36

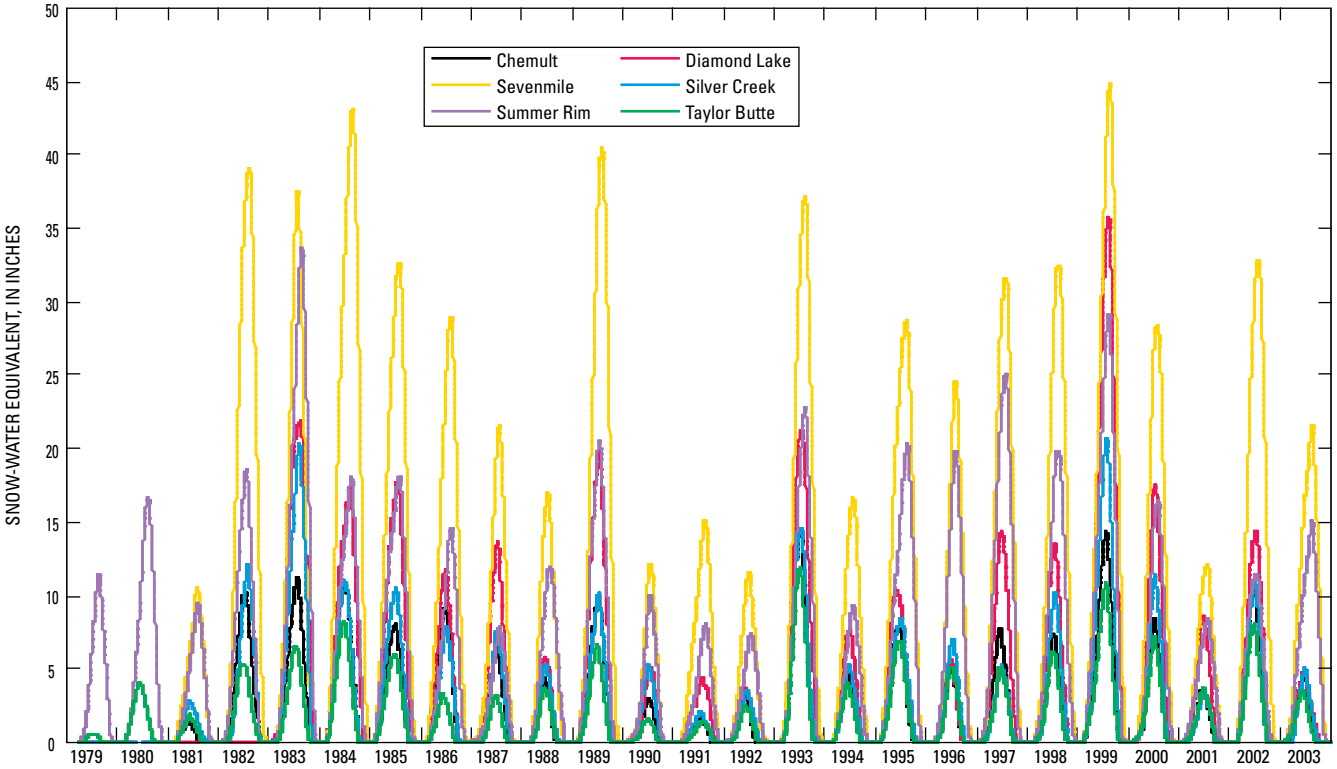


Figure 27. Williamson River Basin 90-day moving average daily snow-water equivalent for water years 1979–2003.

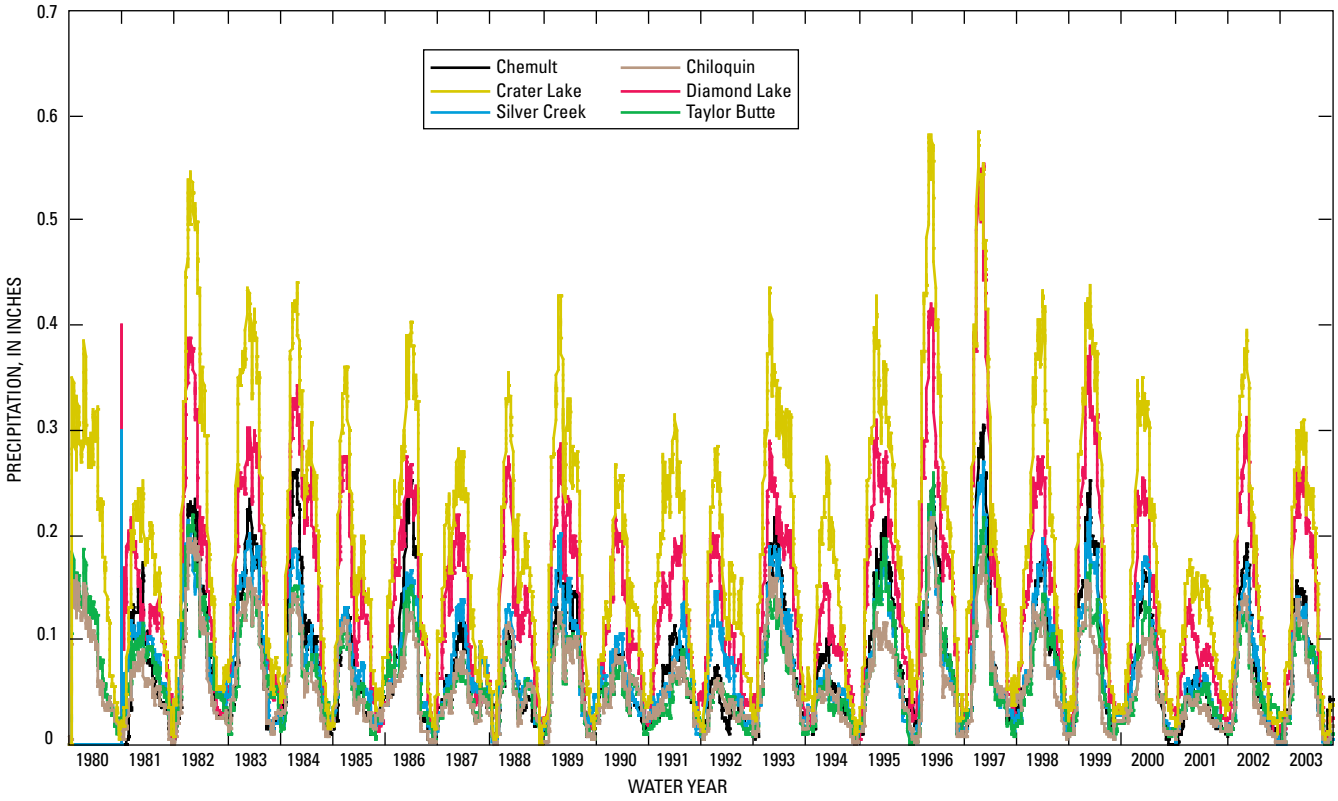


Figure 28. Williamson River Basin 90-day moving average daily precipitation for water years 1980–2003.

A Pearson correlation matrix of the 90-day moving average flow, SWE, and precipitation time series is shown in *table 21*. SWE is more highly correlated with flow than precipitation is with flow. SWE sites are highly correlated with other SWE sites. Precipitation sites are also highly correlated with other precipitation sites. However, SWE and precipitation sites are weakly correlated.

More information about the relationship between climate and flow can be seen in lagged cross-correlations between SWE and flow, and precipitation and flow, as shown in *figures 29 and 30*, respectively. In these plots, SWE (or precipitation) is lagged in time with respect to flow. This is also the same as leading flow in time with respect to SWE and precipitation. Since it is assumed that flow trends respond to SWE and precipitation trends and not the other way around, correlations between current climate and future flows are of interest. From figure 29, all of the SWE sites have their highest correlation (between approximately 0.75 and 0.85) with Williamson River flow within approximately 60 days in the future. Sevenmile SWE and Summer Rim SWE have the highest correlation coefficients, which occurs only 20 days in the future. (The physical reason for this is not apparent. These two SWE sites are not closer to the centroid of the Williamson River or lower in elevation than the other SWE sites.) With precipitation, Williamson River flow has a longer lead in time than it does with SWE, as shown in figure 30. Most of the precipitation sites have their highest correlation (approximately 0.75) with flow about 75 days in the future. After approximately 250 days, both figures 29 and 30 show some increase in their correlations. This is an artifact of the annual periodic component of SWE, precipitation, and flow time series.

Using the Williamson River flow time series as an example, *figures 31 and 32* show the steps used to remove the annual periodic and long-term (decadal) components from each time series and to reduce them to their chaotic residual components. In addition to daily mean flow and 90-day moving average flow, the 90-day moving average flow standard is shown in figure 31. The flow standard time series contains the same value for every day of the year. It is computed by taking the 25-year average of the 90-day moving average flow time series for each calendar day of the year. The 90-day moving average flow standard is the periodic (or annual) component of the time series. Figure 31 also contains a synthetic sine wave that is a representation of the long-term (decadal) component of the time series. It was computed by optimizing the best long-term frequency and phase shift fit to the daily mean Williamson River flow time series.

To compute a chaotic residual component of the time series, a simple ANN model was created. The 90-day moving average Williamson River flow time series was used as the output variable to the model. However, only the 90-day moving average flow standard and the synthetic sine wave time series were used as input variables. The predicted 90-day moving average flow time series is shown and compared with its observed time series in *figure 32*. The difference between the observed and predicted time series can be considered the chaotic residual component of the flow time series.

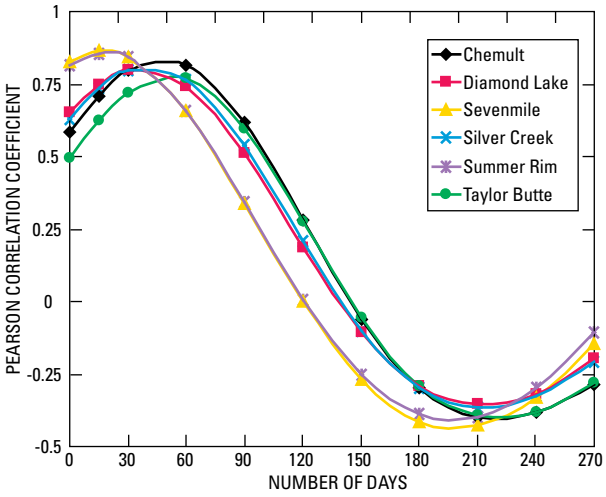


Figure 29. Lagged cross-correlation between the 90-day moving average Williamson River flow and 90-day moving average snow-water equivalent.

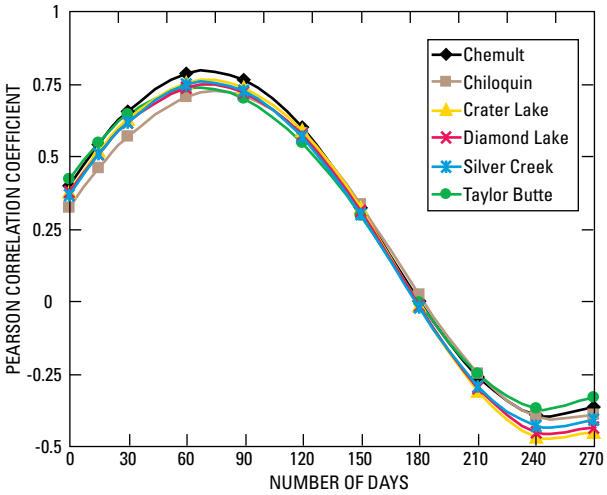


Figure 30. Lagged cross-correlation between the 90-day moving average Williamson River flow and 90-day moving average precipitation.

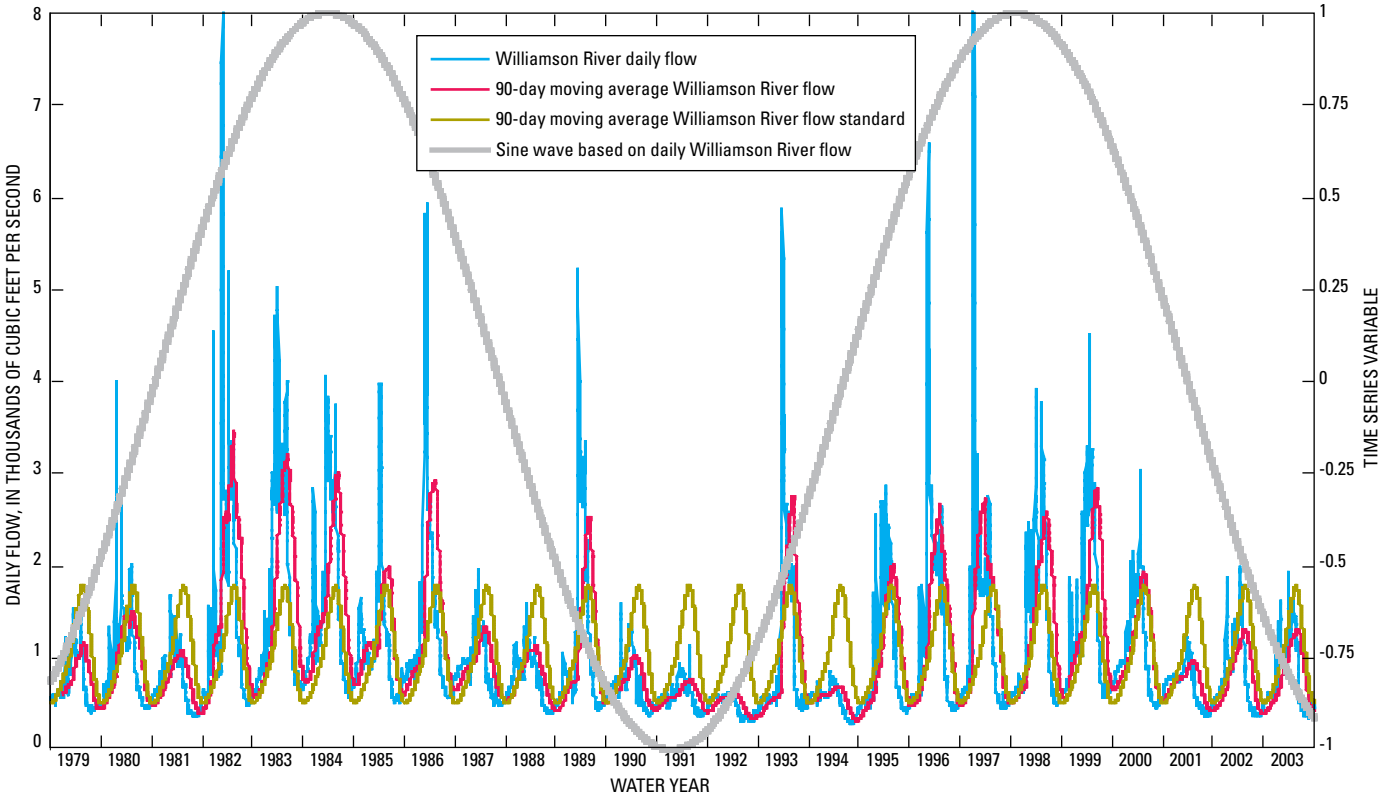


Figure 31. Williamson River daily flow and flow time series variables for water years 1979–2003.

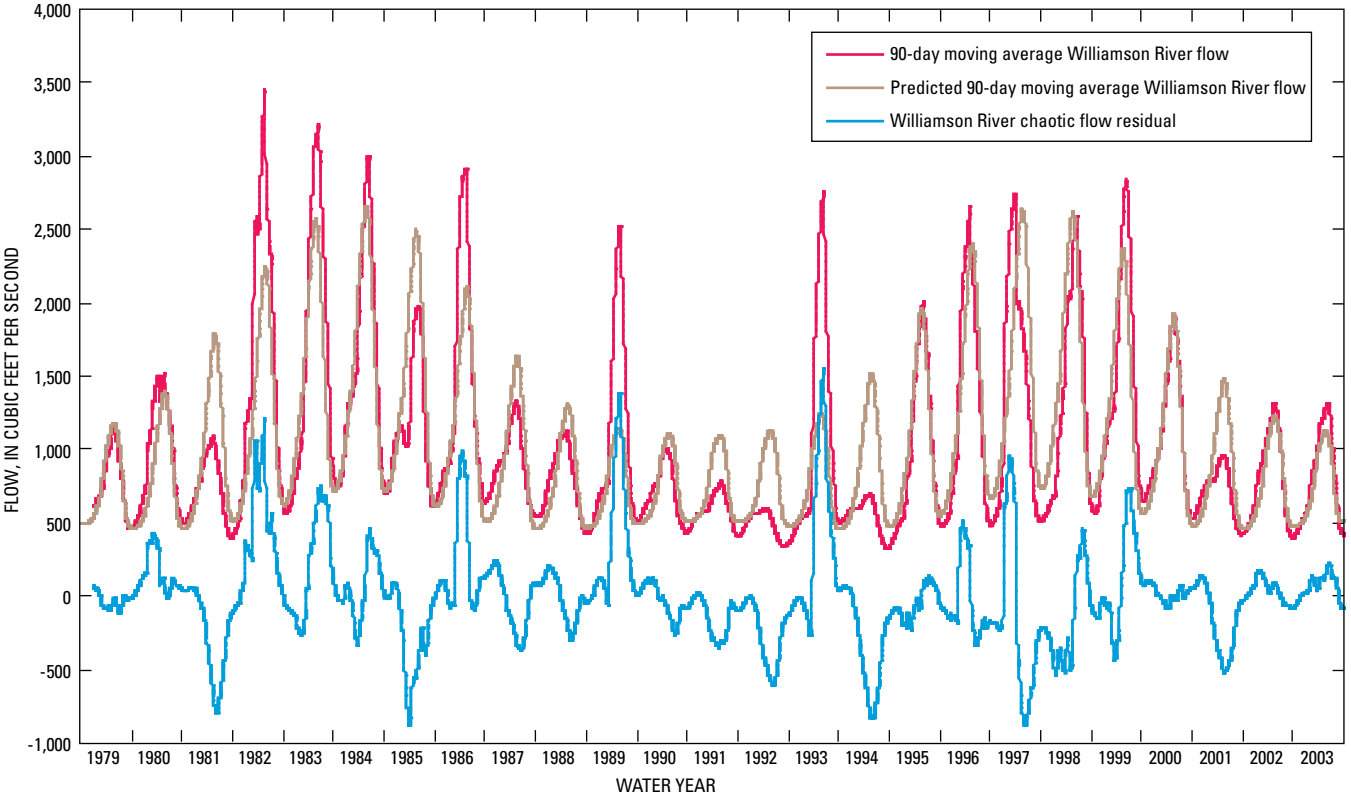


Figure 32. Williamson River observed and predicted flow time series for water years 1979–2003.

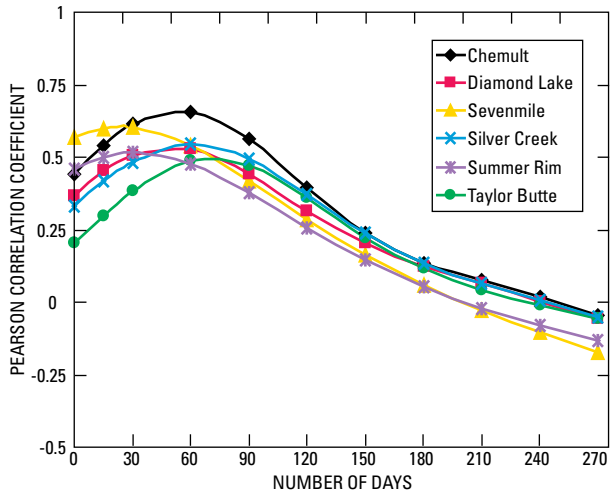


Figure 33. Lagged cross-correlation between the 90-day moving average Williamson River chaotic flow residual and 90-day moving average chaotic snow-water equivalent residual.

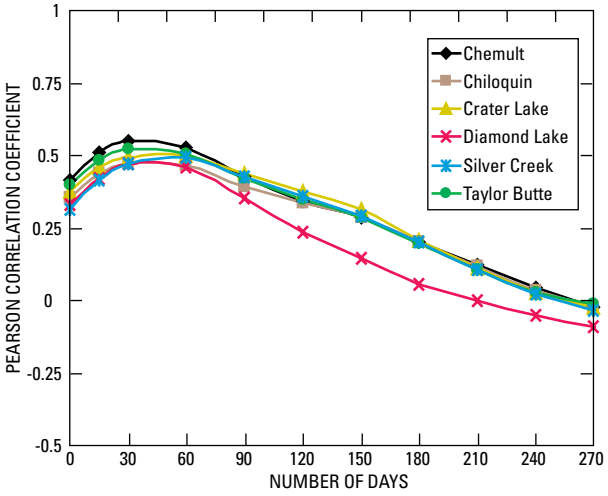


Figure 34. Lagged cross-correlation between the 90-day moving average Williamson River chaotic flow residual and 90-day moving average chaotic precipitation residual.

Table 21. Pearson correlation coefficient matrix of 90-day moving averaged Williamson River Basin flow, snow-water equivalent, and precipitation data, 1980–2002.

[SWE, snow water equivalent; WILLQ, Williamson River flow; CHES, Chemult SWE; DIAS, Diamond Lake SWE; SEVS, Sevenmile SWE; SILS, Silver Creek SWE; SUMS, Summer Rim SWE; TAYS, Taylor Butte SWE; CHEP, Chemult precipitation; CHIP, Chiloquin precipitation; CLKP, Crater Lake precipitation; DIAP, Diamond Lake precipitation; SILP, Silver Creek precipitation; TAYP, Taylor Butte precipitation.]

MATRIX	WILLQ	CHES	DIAS	SEVS	SILS	SUMS	TAYS	CHEP	CHIP	CLKP	DIAP	SILP	TAYP
WILLQ	**1.00**	0.58	0.65	0.83	0.63	0.82	0.50	0.40	0.32	0.38	0.38	0.37	0.42
CHES	0.58	**1.00**	0.93	0.83	0.95	0.79	0.96	0.67	0.58	0.63	0.60	0.64	0.60
DIAS	0.65	0.93	**1.00**	0.87	0.95	0.85	0.89	0.53	0.44	0.51	0.50	0.54	0.49
SEVS	0.83	0.83	0.87	**1.00**	0.83	0.91	0.79	0.48	0.40	0.47	0.45	0.46	0.47
SILS	0.63	0.95	0.95	0.83	**1.00**	0.85	0.91	0.57	0.50	0.55	0.52	0.57	0.53
SUMS	0.82	0.79	0.85	0.91	0.85	**1.00**	0.76	0.49	0.41	0.49	0.48	0.49	0.48
TAYS	0.50	0.96	0.89	0.79	0.91	0.76	**1.00**	0.64	0.55	0.60	0.58	0.62	0.56
CHEP	0.40	0.67	0.53	0.48	0.57	0.49	0.64	**1.00**	0.93	0.94	0.92	0.94	0.94
CHIP	0.32	0.58	0.44	0.40	0.50	0.41	0.55	0.93	**1.00**	0.94	0.91	0.93	0.94
CLKP	0.38	0.63	0.51	0.47	0.55	0.49	0.60	0.94	0.94	**1.00**	0.96	0.94	0.92
DIAP	0.38	0.60	0.50	0.45	0.52	0.48	0.58	0.92	0.91	0.96	**1.00**	0.94	0.90
SILP	0.37	0.64	0.54	0.46	0.57	0.49	0.62	0.94	0.93	0.94	0.94	**1.00**	0.93
TAYP	0.42	0.60	0.49	0.47	0.53	0.48	0.56	0.94	0.94	0.92	0.90	0.93	**1.00**

The final step in the analysis evaluated how well chaotic climate and chaotic flow residual time series were correlated. Similar to *figure 29*, *figure 33* shows the lagged cross-correlation between the chaotic SWE residual time series and the chaotic flow residual time series. The chaotic SWE residual time series all have their highest correlation (approximately 0.5 to 0.67) with the chaotic flow residual time series within 75 days in the future. After their peaks, all of the SWE site correlations decrease in time, which indicates that much of the predictable annual periodicity was removed from the data.

Correlations between chaotic precipitation residual time series and chaotic flow residual time series are shown in *figure 34*. Unlike the correlations with the predecomposed time series (*figure 31*), the chaotic precipitation residual time series have their highest correlation with flow only 30 to 50 days in the future (approximately 0.47 to 0.56). After their peaks, all of the precipitation site correlations decrease in time.

Results shown in both *figures 33 and 34* underscore some of the limitations of using SWE and precipitation data in long-range water supply forecasting of 6-month periods such as February–July and April–September. Most of the peak correlation coefficients are less than 0.6. After 120 days (approximately 4 months) all of the SWE and precipitation correlation coefficients are less than 0.4. A forecast of April–September flow forecasts made on January 1 are based almost entirely on the predictable annual periodic component within the flow time series itself. SWE data measured on January 1 have limited strength in predicting flows that will occur between approximately 90 and 270 days (3 and 9 months) in the future.

Summary

In the upper Klamath Basin, located in south-central Oregon and northeastern California, Bureau of Reclamation (BOR) water managers rely on accurate forecasts of spring and summer streamflow to optimally allocate increasingly limited water supplies for various demands. These demands include (1) maintaining required flows for threatened Chinook and coho salmon in the Lower Klamath River, (2) water deliveries to irrigate about 185,000 acres, (3) retention of water in the Upper Klamath Lake to protect water-quality habitat for two endangered sucker species, (4) hydroelectric power production, and (5) providing water to wildlife refuges.

Cumulative flow forecasts for separate 6-month periods, from April–September and February–July, are made for five upper Klamath Basin sites at the beginning of each month from January through June by the U.S. Department of Agriculture, Natural Resources Conservation Service (NRCS). The upper Klamath Basin forecast sites include flow at the Williamson and Sprague River U.S. Geological Survey (USGS) gages near Chiloquin, Oregon, and net inflows into the Upper Klamath Lake, Clear Lake Reservoir, and Gerber Reservoir. The flow forecasts are made using statistical models which use current snow and precipitation data collected at nearby monitoring sites as input.

Typically, the upper Klamath Basin flow forecasts have been less accurate than most NRCS flow forecasts made for river basins in the Western United States. This is in part because of the geologic conditions in the basin. The basin is extensively composed of permeable late Tertiary to Quaternary volcanic deposits. The youngest and most permeable deposits occur in the Cascade Range where most of the precipitation occurs. As a consequence, many upper Klamath Basin streamflows are highly ground-water dominated and influenced by the climatic conditions of proceeding years. Until recently, the NRCS flow forecast equations did not incorporate variables that represented the ground-water or long-term climate-trend conditions.

In 2003, the USGS, NRCS, and BOR began a collaborative study to reduce some of the uncertainty and errors in seasonal flow forecasting in the upper Klamath Basin. The main objectives included (1) evaluating nonregression statistical modeling approaches, such as artificial neural networks (ANN), for their efficacy in reducing model error, (2) finding and evaluating potential model variables that better described upper Klamath Basin ground-water and long-term climate-trend conditions, and (3) analyzing the limitations of upper Klamath Basin snow-water equivalent (SWE) and precipitation data in forecast models.

Three statistical modeling approaches for flow forecasting were evaluated and compared. These modeling approaches included principal components regression (currently used by the NRCS), nonautoregressive ANN, and autoregressive ANN. Sets of models to forecast April–September and February–July cumulative flows for each monthly forecast from January to June for the five forecast sites were created using the three approaches.

The principal components regression and the nonautoregressive ANN models used SWE and precipitation data, collected at sites within the basins of the forecast sites, as model input variables. However, the autoregressive models used input variables that were derived from characteristics of the flow time series that the models were based upon.

Using the coefficient of determination and residual error (observed minus predicted flow volumes) as measurements of model performance, no single modeling approach always performed better than the others. However, the principal components regression and the nonautoregressive ANN models generally performed better than the autoregressive ANN models. For the Upper Klamath Lake forecast site, the nonautoregressive ANN models had lower error than the other models for the January, February, and March forecasts. However, the principal components regression model performed better for the April forecast. Both models performed roughly the same for the May and June forecasts. For the Sprague River forecast site, the nonautoregressive

ANN models did better than the other models for January, February, March, and June forecasts. However, the principal components regression models performed better for the April and May forecasts. For the Williamson River and Gerber Reservoir forecast sites, the principal components regression models generally, but not always, performed better than the other models. For the Clear Lake Reservoir forecast site, the nonautoregressive ANN performed far better than the other models for the months of January, February, and March. However, the Clear Lake Reservoir autoregressive ANN model performed better than the other models for the month of April.

The effect of long-term climate-trend variables, which included ground-water dominated Fall River flow data and water-elevation data from a well located in the upper Sprague River Basin, were evaluated using some of the principal components regression models and the nonautoregressive ANN models. It was not possible to evaluate these variables using the autoregressive models, because those models only used inputs derived from characteristics of the flow time series. A reasonably satisfactory method of evaluating the climate-trend variables was achieved by removing the variables from the original models and then recreating the models.

For the principal components regression models, the climate-trend variables were evaluated using the Williamson River and Upper Klamath Lake models. For the Williamson River, inclusion of the climate-trend variables consistently improved all models for all the forecast dates except the June model from the February–July model set. The climate-trend variables also improved many, although not all, of the Upper Klamath Lake models.

For the nonautoregressive ANN models, the climate-trend variables were evaluated using some of the Upper Klamath Lake, Williamson River, and Gerber Reservoir models. The climate-trend variables improved most of the April–September flow models for these forecast sites. However, the variables did not improve the Upper Klamath Lake April and May February–July flow models.

During the development of the ANN models, the relationships between the upper Klamath Basin SWE, precipitation, and flow data were analyzed to determine the ability and the extent to which current SWE and precipitation conditions can be used to forecast future flow conditions. The analyses were made by decomposing the flow time series into annual periodic, long-term climatic, and chaotic components. After creating a synthetic time series of the chaotic flow component for each of the five forecast sites, it was possible to lag correlate them with SWE and precipitation time series data collected at nearby sites within their basins. Analysis results underscored some the limitations of using SWE and precipitation data in long-range water supply forecasting of 6-month periods such as February–July and April–September. After 120 days (approximately 4 months) all of the SWE and precipitation correlation coefficients are less than 0.4.

References Cited

Bureau of Reclamation, 2005, Project history: *http://www.usbr.gov/mp/kbao/news.html*, accessed on Feb. 22, 2005.

Cannon, A.J., and Whitfield, P.H., 2001, Modeling transient pH depressions in coastal streams of British Columbia using neural networks: Journal of the American Water Resources Association, v. 37, no. 1, p. 73–89.

Conrads, P.A., and Roehl, E.A., 1999, Comparing physics-based and neural network models for simulating salinity, temperature, and dissolved oxygen in a complex tidally affected river basin: South Carolina Environmental Conference, Myrtle Beach, March 15–16, 1999, Proceedings, 7 p.

Conrads, P.A., and Roehl, E.A., 2000, Real-time control of the salt front in a complex, tidally affected river basin: Artificial Neural Networks *in* Engineering Conference, St. Louis, Missouri, November 5–8, 2000, 7 p.

Conrads, P.A., Roehl, E.A., and Cook, J.B., 2002, Estimation of tidal marsh loading effects in a complex estuary: American Water Resources Association Annual Conference, New Orleans, Louisiana, Proceedings, May 2002, p. 307–312.

Conrads, P.A., Roehl, E.A., and Martello, W.P., 2002, Estimating point-source impacts on the Beaufort River using neural network models: American Water Resources Association Annual Conference, New Orleans, Louisiana, Proceedings, May 2002, p. 289–294.

Daly, C., Nielson, R.P., and Phillips, D.L., 1994, A statistical-topographic model for mapping climatological precipitation over mountainous terrain: Journal of Applied Meteorology, v. 33, p. 140–158.

Daly, C., Taylor, G.H., and Gibson, W.P., 1997, The PRISM approach to mapping precipitation and temperature: Conference on Applied Climatology, 10th, Reno, Nevada, October 1997, Proceedings (reprint): American Meteorological Society, p. 10–12.

Dicken, E.F., and Dicken, S.N., 1985, The legacy of Lake Modoc—A historical geography of the Klamath Lakes Basin: Eugene, University of Oregon, 166 p.

Gonthier, J.B., 1984, A description of aquifer units in eastern Oregon: U.S. Geological Survey Water-Resources Investigations Report 84–4095, p. 14–17.

Garen, D.C., 1992. Improved techniques in regression-based streamflow volume forecasting: Journal of Water Resources Planning and Management, 118, no. 6, 654–670.

Hinton, G.E., 1992, How neural networks learn from experience: Scientific American, September 1992, p. 145–151.

Hsu, K., Gupta, H.V., and Sorooshian, S., 1995, Artificial neural network modeling of the rainfall-runoff process: Water Resources Research, v. 31, no. 10, p. 2517–2530.

Hsu, K., Gupta, H.V., and Sorooshian, S., 1998, Streamflow forecasting using artificial neural networks, *in* Proceedings of the International Water Resources Engineering Conference, Memphis, Tennessee, August 3–7, 1998: Reston, Virginia, American Society of Civil Engineering, v. 2, p. 1362–1367.

Karunanithi, N., Grenney, W.J., Whitley, D., and Bovee, K., 1994, Neural networks for river flow prediction: Journal of Computing in Civil Engineering, v. 8, no. 2, p. 201–220.

Kuligowski, R.J., and Barros, A.P., 1998, Using artificial neural networks to estimate missing rainfall data: Water Resources Bulletin, v. 34, no. 6, p. 1437–1447.

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: Bulletin of the American Meteorological Society, v. 78, p. 1069–1079.

Morshed, J., and Kaluarachchi, J.J., 1998, Application of artificial neural network and genetic algorithm in flow and transport simulations: Advances in Water Resources, v. 22, no. 2, p. 145–158.

Risley, J.C., Roehl, E.A., Conrads, P.A., 2003, Estimating water temperatures in small streams in western Oregon using neural network models: U.S. Geological Survey Water-Resources Investigations Report 02–4218, 60 p.

Shamseldin, A.Y., 1997, Application of a neural network technique to rainfall-runoff modeling: Journal of Hydrology, v. 199, p. 272–294.

Taylor, G.H., and Hannan, C., 1999, The climate of Oregon—From rain forest to desert: Corvallis, Oregon State University Press, 224 p.

Thirumalaiah, K., and Deo, M.C., 1998, River stage forecasting using artificial neural networks: Journal of Hydrologic Engineering, January 1998, p. 26–32.

U.S. Geological Survey, 1999, National land cover data set: *http://www.geodata.gov/gos* (Go to: "Environment and conservation," then "Land cover"), accessed on Jan. 4, 2005.