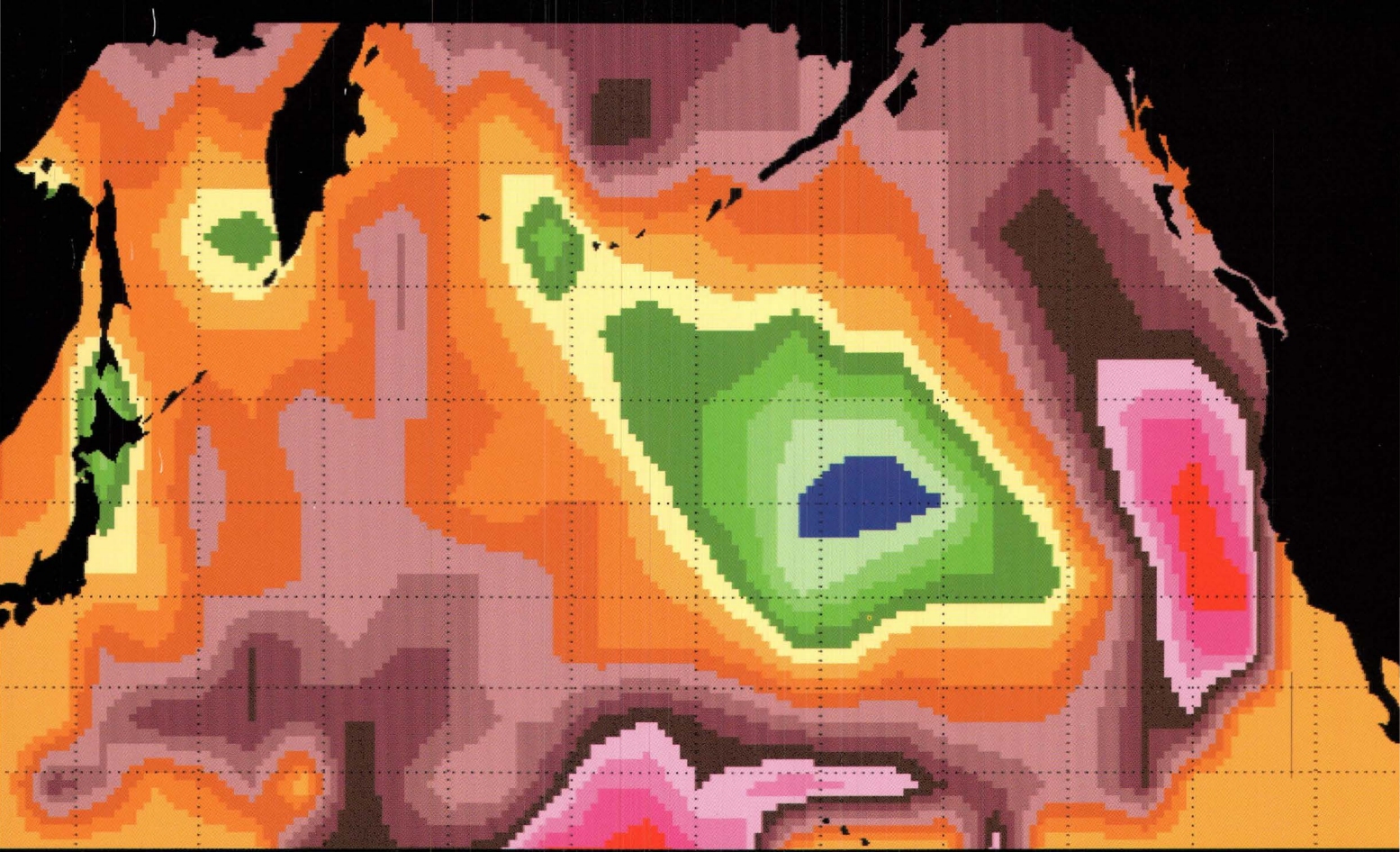
U.S. Department of the Interior
U.S. Geological Survey

Development, Testing, and Assessment of Regression Equations for Experimental Forecasts of Fall-Transition-Season Inflows to the Howard A. Hanson Reservoir, Green River, Washington

Prepared in cooperation with
Tacoma Public Utilities,
Seattle Public Utilities, and
U.S. Army Corps of Engineers



Water-Resources Investigations Report 00-4153

USGS
science for a changing world

Cover: The spatial distribution of correlation coefficients for the North Pacific between sea-surface temperatures and reservoir inflow. The front cover shows the correlation pattern from August—14 months prior to the October inflows, and the back cover shows the correlation pattern from September—13 months prior to the October inflows. The reds represent the highest positive correlation and the blues represent the highest negative correlation.

United States Department of the Interior

U.S. GEOLOGICAL SURVEY

Water Resources Division
Washington District
1201 Pacific Avenue, Suite 600
Tacoma, Washington 98402
(253) 428-3600 • FAX (253) 428-3614
http://wwwdwatcm.wa.water.usgs.gov/

November 30, 2000

MEMORANDUM

To: Publications Coordination, U.S. Geological Survey, Reston, VA

From: J.A. Wayenberg, Physical Scientist, WRD, Tacoma, WA

Subject: PUBLICATIONS -Recently Released Report

Enclosed are 2 copies of the following report.

Development, Testing, and Assessment of Regression Equations) for Experimental Forecasts of Fall-Transition-Season Inflows to the Howard A. Hanson Reservoir, Green River, Washington,, by J.J. Vaccaro, WRIR 00-4153.

Enclosures: (2)

cc:
Natural Resources Library, Washington, D.C. (3 copies)
Library of Congress, Washington, D.C. (2 copies)
USGS Library, Reston, VA (2 copies)
USGS Library, Menlo Park, CA (2 copies)
USGS Library, Spokane, WA (2 copies)
USGS Library, Lakewood, CO (2 copies)
Regional Hydrologist, Menlo Park, CA (1 copy)
Public Inquiries Office, Spokane, WA (1 copy)

Development, Testing, and Assessment of Regression Equations for Experimental Forecasts of Fall-Transition-Season Inflows to the Howard A. Hanson Reservoir, Green River, Washington

By J.J.Vaccaro

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 00-4153

Prepared in cooperation with the

TACOMA PUBLIC UTILITIES,
SEATTLE PUBLIC UTILITIES, and
U.S. ARMY CORPS OF ENGINEERS

Tacoma, Washington
2000

U.S. DEPARTMENT OF THE INTERIOR
BRUCE BABBITT, Secretary

U.S. GEOLOGICAL SURVEY
Charles G. Groat, Director

For additional information write to:

District Chief
U.S. Geological Survey
1201 Pacific Avenue – Suite 600
Tacoma, Washington 98402

Copies of this report can be purchased from:

U.S. Geological Survey
Information Services
Building 810
Box 25286, Federal Center
Denver, CO 80225-0286

CONTENTS

FIGURES

TABLES

VERTICAL DATUM

Sea Level: In this report "sea level" refers to the National Geodetic Vertical Datum of 1929 (NGVD of 1929)--a geodetic datum derived from a general adjustment of the first-order level nets of both the United States and Canada, formerly called Sea Level Datum of 1929.

Development, Testing, and Assessment of Regression Equations for Experimental Forecasts of Fall-Transition-Season Inflows to the Howard A. Hanson Reservoir, Green River, Washington

By J.J. Vaccaro

ABSTRACT

A method for forecasting inflows to reservoirs at long-lead times (14 months to 2 months prior) for the September–November fall-transition season was formulated, analyzed, tested, and applied to the Howard A. Hanson Reservoir on the Green River in western Washington. The method uses multiple linear regression to estimate the monthly mean streamflow for the combined months of September and October (a low-flow period), October (initial fall transition), and November (onset of the fall precipitation season). The predictors in the equations are monthly values and 3-month averages of the monthly values of hemispheric-to-regional variables calculated using atmospheric and sea-surface-temperature data, hydrometeorological data, and sunspot numbers.

Five equations were calibrated and tested, two for September–October inflows (14- and 13-month lead times), two for October inflows (11- and 2-month lead times), and one for November inflows (14-month lead time).

The equations were initially calibrated using 27 years of data from the 1952–96 period and were tested using the remaining 18 years of record. The calibrated equations were significant at greater than a 95-percent level. The results of the testing indicated that the equation for November had the largest r-squared value (0.80) and the equations for the other inflow periods had values that ranged from 0.43 to 0.57. For all equations, the standard error for the inflow estimates of the testing period was less than the standard deviation of the observed values. This initial testing suggested the potential future performance of the equations.

The final equations were calibrated using data for all years from the 1952–96 period. The September–October, October, and November equations had r-squared values of 0.71 and 0.70, 0.74 and 0.80, and 0.82, respectively. These values were smaller than those from the initial 27-year calibration period and larger that those from the 18-year testing period. The r-squared values, residual values, and standard errors all indicate that the equations, on the average, will perform reasonably well.

The equations provide a tool for the management of reservoir outflows, especially to help guide existing management decisions such as releasing water in the spring for flushing salmon smolt and, concurrently, storing water for fall releases for steelhead migration and spawning. The ability to estimate the inflows for the fall-transition season, well before the beginning of the fall runoff, may decrease uncertainty in the task of regulation decisions.

INTRODUCTION

The ability to forecast seasonal inflows to reservoirs at long-lead times provides an important tool for the complex task of regulating reservoir outflows. In cooperation with the Tacoma Public Utilities, the Seattle Public Utilities, and the U.S. Army Corps of Engineers, the U.S. Geological Survey developed a method for calculating forecasts of inflow of the Green River to the Howard A. Hanson Reservoir (Hanson Reservoir) for the September–November fall transition season. Hanson Reservoir is located on the west slope of the Cascade Range in Washington State (fig. 1).

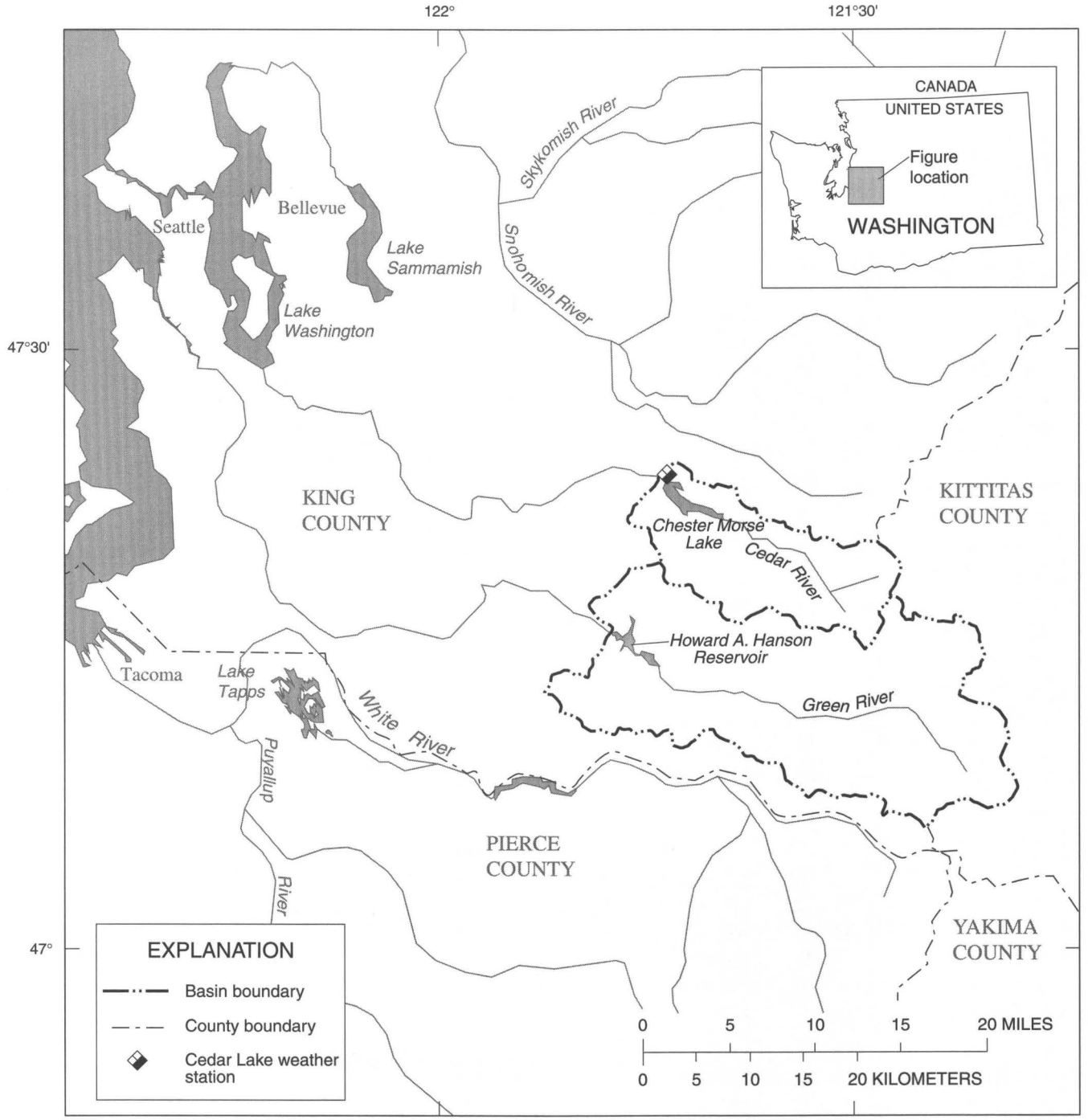


Figure 1. Location of the Howard A. Hanson Reservoir, in the Green River Basin, and Chester Morse Lake, in the Cedar River Basin, Washington.

The fall transition period was chosen because it is a critical period for water supply and fisheries and there are currently no forecasts of streamflow at long-lead times for the September–November period. The forecasts for the Green River also are applicable, in terms of either percentiles or standardized values, to the inflows of the Cedar River to Chester Morse Lake (fig. 1), which is operated by the Seattle Public Utilities. The Green and Cedar Rivers share a common basin boundary and drain the west slopes of the Cascade Range, and their highly variable inflows during the fall-transition season are strongly correlated. The correlation coefficients for the September-through-November flows between the Green and Cedar Rivers are all greater than 0.90, reflecting the fact that the shape of the outflow hydrographs for these two basins are very similar during this season.

Because each year about 26,000 acre-feet of storage is captured in the Hanson Reservoir for augmentation of summer-fall streamflow, the major issue in this study was to develop a means to forecast inflows, at long-lead times, that would indicate the potential for a wet or a dry fall-transition season–especially for October. The forecasts would help to facilitate the management of outflows, in order to meet the conflicting demands of releasing water in the spring for flushing salmon smolt and at the same time storing water for release during the fall-transition season for steelhead, coho, and chinook migration and spawning, as well as for meeting the seasonal water demand.

The method chosen to forecast inflows was to formulate multiple linear regression equations to estimate the mean monthly flow for the combined months of September and October (a low-flow period), for October (initial fall transition), and for November (onset of the fall precipitation season). This technique is the simplest of the currently used techniques, and it identifies explanatory variables that are the most related to the inflows. Initially, other, more complex techniques were also explored, and they are briefly described in Appendix 2 at the end of the report. The predictors, or explanatory variables, in the equations are monthly values, and 3-month averages of monthly values, of hemispheric-to-regional variables calculated using atmospheric and sea-surface temperature data, hydrometeorological data, and sunspot data. This approach is based on the well-documented strong link in the Pacific Northwest between natural climate-forcing mechanisms (on a hemispheric-to-regional scale) and such hydrometeorological aspects as precipitation and streamflow. Thus, the possibility of using equations to forecast streamflow was evident from the persistence and strength of these climate variables and their strong link to temperature, precipitation, snowpack, and streamflow in the Pacific Northwest. The prediction results, although not excellent, indicate that this is a promising method that should be pursued further.

Purpose and Scope

The purpose of this report is to (1) describe the selection of explanatory variables for, and the development of five regression equations that forecast inflows to the Hanson Reservoir; (2) document the calibration and testing of the equations with historical inflow data for the Hanson Reservoir; (3) assess the performance and reliability of the equations for predicting inflow for the fall transition season and for predicting a wet or a dry year; and (4) assess the performance and reliability of the equations, using the most current data, as a measure of future performance.

The explanatory variables used to formulate the equations were selected from four general categories of climate data: sea-surface temperature, regional-climate and atmospheric-circulation measures, North Pacific atmospheric-circulation measures, and hemispheric measures, including sunspots.

The inflow data used to calibrate and test the equations were for the 45-year period 1952-96 from the historical record, which includes a major regional hydrometeorological regime shift in 1967. A "partial period" of 27 years from the 1952-96 data was used to calibrate the initial equations, and the remaining 18 years of record were used to test them, including water years 1992-96 to reflect current conditions. The final equations were calibrated for the complete 1952-96 record and were tested using inflow data for water year 1997.

The seasonal inflows addressed in this report are the sum of the monthly mean flows for the September–October period, the monthly mean October flows, and the monthly mean November flows. Descriptive information for seasonal, monthly, and annual inflows for the Hanson Reservoir is given in Appendix 1; the monthly inflows were provided by Thomas Murphy (written commun., U.S. Army Corps of Engineers, 1994-97).

The inflows exhibit very low autocorrelation, with a lag-1 coefficient not exceeding 0.06. The September flow volumes exhibit little variability—the range between the 1st- and 70th-percentile flows is nearly equivalent to the standard deviation and is only about one-half of the mean.

Five regression equations were developed for the seasonal inflows. The explanatory variables for the equations use atmospheric-circulation data, sea-surface temperatures, hydrometeorological data, and, for two equations, sunspot numbers. A preliminary exploratory study indicated that monthly values and 3-month seasonal averages (or sums) of the variables were most appropriate for analysis. These two temporal domains are consistent with the seasonal progression of most atmospheric circulation patterns and with the transport of sea-surface temperatures (which also change seasonally) along oceanic currents.

Of the five equations, two are for the September–October inflows, two are for the October inflows, and one is for the November inflows. The lead times for the September–October equations are virtually the same, and the lead times for the October equations are different. Because the October flows are generally about double the September flows, the equations for the September–October inflows typically capture more of the inherent variability of the October flows, which have a standard deviation that is about one and one-half times larger than the mean September flow. In addition, for the complete period of record used in this study (1915–97), the September and October monthly flows are in phase up to 1967, and thereafter they are not in phase. (Being in phase means that the years of occurrence of higher and lower flows are generally the same.) This change in 1967 is one of the regional hydrometeorological regime shifts described by Vaccaro (1996).

The lead time of an equation used to predict reservoir inflow is the shortest time span between the beginning of the inflow season and the time represented by any of several explanatory variables of the equation. For example, one of the equations used to predict October inflows uses values of explanatory variables that precede the October inflows by anywhere from 9 to 33 months. In this case, the lead time of the equation is 9 months, because it is the shortest lead time between October and any of the explanatory variables.

Physical and Hydrometeorological Setting

The Hanson Reservoir has a storage capacity of 105,463 acre-feet and Chester Morse Lake about 38,137 acre-feet. Mean annual inflow to the Hanson Reservoir is about 1,000 cubic feet per second (ft^3s^{-1}) and to Chester Morse Lake about 400 ft^3s^{-1}, representing runoff from 220 square miles (mi^2) and 79 mi^2, respectively. Streamflow in both basins is supported by both rainfall runoff and snowmelt runoff. The outflow point of the Hanson Reservoir is about 500 ft feet lower in altitude than that of Chester Morse Lake, so not only is the Hanson Reservoir drainage area larger, but more of it lies in the transient-snow zone at lower altitude and with less relief.

Mean annual precipitation in the Cedar and Green River Basins ranges from about 60 to more than 135 inches. April 1 snow-water equivalent (1969–95) in the Green River Basin is minimal at snow-course sites below about 2,100 feet (ft) in altitude, 20 inches for sites at 3,500 ft, and 34 inches at the highest site at 4,700 ft. The snow-water equivalent at the high-altitude site has historically ranged from 6 to 66 inches. Generally, snow-water equivalent from March to April decreases for areas below about 2,800 ft and increases at higher altitudes. The Green River's mean monthly flows for November–May range from about 1,200 ft^3s^{-1} in November to 1,600 ft^3s^{-1} in December (Appendix 1). August has the lowest mean monthly flows (200 ft^3s^{-1}), followed by September (220 ft^3s^{-1}). These months of low flows reflect the overall drought-like character of the summer season in western Washington, which receives about 60 percent of its total water-year precipitation during the months November through March.

For the Green River, the October inflows have the largest coefficient of variation, followed by the November inflows. A large coefficient of variation for October inflows reflects the fact that the onset of the fall-precipitation season may or may not occur during this month. Thus, October inflows appear to consist of two very distinct populations, even more so than the other monthly inflow values. For the Cedar River, September inflows have the largest coefficient of variation, reflecting the limited ability of the smaller Cedar River Basin to store water, especially base flows, in comparison with the larger Green River Basin. October and November have the next largest coefficient of variation in the Cedar River.

DEVELOPMENT OF THE FORECASTING EQUATIONS

Numerous investigators have documented the strong linkage, at various time scales, between hemispheric-to-regional climate forcing and such aspects as precipitation and streamflow in the Pacific Northwest. See, for example, Yarnal and Diaz (1986), Speers and Mass (1986), Cayan and Peterson (1989), Redmond and Koch (1991), Ebbesmeyer and others (1989), and Cayan and Webb (1992). The persistence, strength, and interconnections of quasi-stationary circulation patterns and sea-surface temperature anomalies has also been well documented. See, for example, Bjerknes (1969), Horel and Wallace (1981), Namias (1981), Rogers (1981), Barnston and Livezey (1987), and Namias and others (1988). The persistence and strength of these climate variables and, their strong linkages to streamflow suggested the possibility of formulating atmospheric-circulation and sea-surface temperature variables that are strongly linked to temperature, precipitation, and streamflow in the Pacific Northwest. These climate-related variables resulted from analysis of regime shifts in precipitation, temperature, snowpack, and streamflow, and their linkage to climate information (Vaccaro, 1996). The variables are monthly time-series and include such previously well-documented and well-known variables as the Southern Oscillation Index (SOI), which is obtained from the Climate Analysis Center (CAC), National Oceanic and Atmospheric Administration (NOAA). The variables are spatially invariant because they are based on data from fixed points in space.

A screening process was used to obtain a small set of variables from the potentially large number of variables possible. The explanatory variables were used to develop five regression equations: two for the period September–October (at 14- and 13-month lead times), two for October (at 9- and 2-month lead times), and one for November (at a 14-month lead time).

Selection of Explanatory Variables

Generally, the explanatory variables that were chosen had not only a statistical link, but also a climatological basis for a physical link between the variables and Pacific Northwest hydrometeorological variables (Vaccaro, 1996). Some of the variables considered are not necessarily independent because they are only calculated differently using the same or nearly the same data points. For example, values from two grid points of a data set may be subtracted to obtain one variable and added to obtain another, resulting in, say, a pressure-difference variable and a pressure-height variable, which may or may not be correlated. Some of the variables formulated using 700-millibar (mb) height values (units of tens of feet minus 700) are also represented as standardized series.

Most explanatory variables were hemispheric-to-regional variables that are measures of the climate system. Although they also provide estimating ability, few site-specific variables, such as precipitation, snowpack, or temperature, were used because the intent was to make the results from the regression equations (in terms of percentiles or standardized units) more regional in nature. There is a significant relation between the inflows to Hanson Reservoir and streamflow at 22 sites in western Washington. For example, there is a correlation coefficient of 0.93 between the inflows to Hanson Reservoir and the time-series of the first principal component of the streamflow for the 22 western Washington sites. There is a correlation coefficient of 0.96 between the inflows and the average of the standardized values of streamflow for 11 of the sites that had first-principal-component values that were similar to those of the Green River. This suggests that, by using mainly large-scale variables and few basin-specific data as explanatory variables, the equations developed for the Green River may also be applicable for some other streams in western Washington. For example, a forecasted 20th-percentile inflow value for Hanson Reservoir may indicate that some other nearby streams might also have a value in the 20th-percentile range.

For three of the five equations, however, site-specific precipitation and temperature data from the Seattle Public Utilities's Cedar Lake weather station (altitude of 1,460 ft at latitude 47.4167°N and longitude 121.7333°W) and sunspot numbers were used in order to improve the robustness of the equations. The site-specific weather data are used because (1) the weather data are highly correlated to the fall-transition period inflows for both the Green and Cedar Rivers, and (2) the site is considered a key indicator of meteorological conditions for the west slope of the Cascade Range in Washington (Rasmussen and Tangborn, 1976; Ebbesmeyer and others, 1989; Vaccaro, 1996). Mean water-year precipitation at Cedar Lake is 102 inches and has ranged from 67 to 135 inches. Mean monthly precipitation ranges from

14.5 inches in January to 2.2 inches in July. The sunspot numbers are correlated to seasonal inflows, but they are poorly correlated with most of the climate-related variables; therefore, they provide additional forecasting ability that is relatively independent of the climate variables.

The atmospheric information used in this study consists of the monthly SOI (CAC, NOAA) and diamond-grid, monthly mean 700-mb heights for the Northern Hemisphere (data from 1947–95 from D. R. Cayan, written commun., U.S. Geological Survey, 1995; data from 1996 to present from CAC, NOAA—extracted from the monthly mean values contained in the climate data assimilation system (CDAS) data set). The sea-surface temperature data set used in the preliminary stages of this study consisted of 5°-gridded monthly means for the latitude range 20°-60°N and the longitude range 110°W-130°E (D. R. Cayan, written commun., U.S. Geological Survey, 1995). The final sea-surface temperature data set consisted of 4°-gridded monthly means for the same area. This data set was obtained by averaging the 1°-gridded monthly means of the optimally interpolated sea-surface temperature data for the period 1982 to present, obtained from the National Center for Environmental Prediction (NCEP), NOAA. For the period 1950–81, reconstructed sea-surface temperatures on a 2°-grid (Smith and others, 1996) were averaged to the 4°-grid. The sunspot number time-series was obtained from the National Geophysical Data Center (NGDC) operated by NOAA. The SOI, CDAS, and sea-surface temperature data can be obtained from the CAC/NCEP world wide web site, and the sunspot data can be obtained from the NGDC site.

The water years of record for the SOI were 1933–96, for the CDAS were 1950-96, and for the sea-surface temperature were 1947-96; the sea-surface temperature data set started in January 1950, and the 700-mb data set started in December 1946. Thus, most of the atmospheric and sea-surface temperature data used to formulate the variables have a common period of water years 1950-96 (47 years). The actual record length used in the analyses ranged from 44 to 48 years because of the long-lead aspect of the equations. The sunspot time-series began in 1750.

Development of Final Explanatory Variables

Each explanatory variable in an equation is a time-series measurement for some specific prior time: for example, a December value of a 4°-grid sea-surface temperature would be used in an equation to predict the following October inflow. Because of the potentially large number of possible explanatory variables, a screening process was used to obtain a small set of variables to be used in a stepwise regression analysis. As a result of this process, a lead time was selected for each equation and the understanding of the relation between inflows and the climate system was improved.

Values of monthly explanatory variables and their 3-month averages initially were screened using a correlation analysis to determine which variables correlated most strongly to each of the seasonal inflow periods. The correlation coefficients were calculated by lagging the variables 1 month at a time, starting in October of the fall-transition season and ending 3 years prior to the October. For example, for the 3-month averages, calculation of a correlation coefficient might be based on the 45-year period of record for, say, a March–April–May average of some variable and the following October inflow; the next correlation coefficient would be for the February–March–April average of the variable and the October inflow. Each monthly time-series therefore has 72 potential values (36 months plus the 3-month averages for the 36 months) for use in regression analysis. This screening analysis was repeated with cubed values of the 3-month and monthly variables because cubing emphasizes the extremes of some variables.

Several thousand correlation coefficients were examined. It was determined that, for each of the three seasonal inflows, a large number of variables were significantly correlated at various time lags. To further screen the significantly correlated variables, they were stepwise regressed against the corresponding inflows. The regression was done for several partial periods consisting of 26 years of inflows and also for the 1952–96 period. The 1952–96 period was used because there have been distinct temporal changes, including shifts and trends, in the persistence, strength, existence, and centers of major circulation features that are not fully captured by partial periods or by spatially invariant variables (Vaccaro, 1996).

The results of both the compilation of the significantly correlated variables and the regression analysis were used to select five sets of 20 to 50

variables, one set for developing each of the five regression equations. This selection was oriented to finding the fewest number of variables common between the sets, and was also aimed at obtaining sets of variables that were somewhat consistent with the seasonal progression of hemispheric-North Pacific circulation features (Barnston and Livezey, 1987) and North Pacific features (Vaccaro, 1996). Both the September–October and the October equations share common variables because most of the climate signature is represented in the highly variable October inflows.

A preliminary exploratory analysis determined that five to nine variables is a reasonable number for an equation. Using fewer than five variables allows one to two variables to be too influential. Using more than about nine variables leads to stability problems that are probably due to (1) the effects of large changes in variables that are not as strongly correlated as the first one to nine variables, (2) increasing cross correlation between variables as the number increases, and (3) increasing noise with increasing variables (because the less-dominant modes of hemispheric-to-regional circulation features contain more noise).

Function of Explanatory Variables in Equations

Nineteen variables were selected to be used in the formulation of the five regression equations and they are listed in table 1. The variables can be grouped into four general categories: sea-surface temperatures (sst), regional climate and atmospheric-circulation measures, North Pacific atmospheric-circulation measures, and hemispheric measures.

The sst variables (variables 1-4 in table 1) essentially account for the persistence of Pacific Ocean temperatures in the upper part of the water column in certain regions of the North Pacific. This persistence is associated with the development of anomalous sst patterns and affects atmospheric-circulation features. Sst1 accounts for sst persistence during the summer for a region located offshore of the Pacific Northwest that borders the eastern edge of a cold-water pool associated with a major circulation feature called the Aleutian Low. Sst2 and sst3 account for changes and persistence in sst's from summer into winter near the influential California Current. Based on the preliminary exploratory analysis, sst4 is an indicator of west-east sst gradients and the movement of water across the northern part of the Pacific.

All but one of the regional measures (variables 5-10) indicate the importance of persistence during the May through July period. The lead time for the variables also corresponds to the dominant frequencies contained in the reservoir inflows for the fall-transition season as determined by harmonic analysis. Variables 5 to 8 essentially measure the tropospheric atmospheric conditions centered over northwestern Washington and southwestern British Columbia, Canada, and are influential for the following reasons (Vaccaro, 1996): (1) the variables are indicative of the strength and location of an influential high-pressure cell (HIGH-PRESSURE) that is one of the centers of a dominant hemispheric circulation pattern; (2) the height field over northwestern Washington determines or affects the magnitude of the upper-level winds, the location of storm tracks, and the strength of the zonal (latitudinal) or meridional (longitudinal) atmospheric flow. These three aspects are accounted for by the HEIGHT, HEIGHT-STAND, and NSGEO variables. Variables 9 and 10, the data for the Cedar Lake weather station, capture the persistence in wet/dry and warm/cold patterns for part of the west slope of the Cascade Range. The climate signal in the data for this site is very strong at scales from hemispheric to North Pacific to regional. For example, the winter precipitation at this site is highly correlated to the SOI, and is strongly linked to the existence, location, strength, and persistence of the Aleutian Low and an associated cold-water pool. At the regional scale, air temperatures for the Cedar Lake site are strongly related to geostrophic flow over northwestern Washington, and the variable captures major climate-regime changes.

The North Pacific atmospheric-circulation variables (variables 11-16) measure the strength of circulation patterns that influence streamflow in western Washington during the fall-transition season. The variables WPO1, WPO4, and NPAC contain features (represented as 700-mb grid-point data) of circulation patterns that are described by Barnston and Livezey (1987). The EASTPAC circulation pattern was clearly defined as an influential dipole in the preliminary exploratory analysis completed in this study.

Table 1. Atmospheric-circulation, sea-surface temperature, hydrometeorological, and sunspot variables used in regression equations for estimating September–October, October, and November seasonal inflow

Variable name	Month	Type[1]	Lead time[2] (months)	Descriptive information
				Sea-surface temperature variable[3]
1 sst1	June	M	17	4°-grid cell, average for 48°N,138°W
2 sst2	Dec	M	11	4°-grid cell, average for 36°N,126°W
3 sst3	Aug	3M	15	4°-grid cell, average for 36°N,126°W
4 sst4	June	3M	5	4°-grid cell, average for 44°N,162°E
				Regional climate and atmospheric-circulation variables[4]
5 HEIGHT-STAND	June	3M3	17	Measures geopotential height field, centered at 50°N,125°W
6 HIGH-PRESSURE	July	M	28	Measures strength of a high-pressure cell, centered at 55°N,115°W
7 HEIGHT	July	3M	28	Measures geopotential height field, centered at 50°N,125°W
8 NSGEO	May	M3	30	Measures strength of northerly flow between 50°N,120°W–130°W, a measure of the 700-mb geostrophic flow from the north
9 Cedar L. Precip.	Nov	M	36	Measures precipitation at Cedar Lake weather station at 1,460 feet
10 Cedar L. Temp.	May	M	30	Measures monthly mean temperature at Cedar Lake weather station
				North Pacific atmospheric-circulation variables[4]
11 EASTPAC	Mar	M	20	Measures strength of north-south dipole, 25°N,135°W and 60°N,150°W
12 WPO4	Mar	M	20	Measures strength of north-south dipole, 25°N,165°W and 55°N,165°W
13 WPO1	Aug	3M	27	Measures strength of northeast-southwest dipole, 25°N,110°W and 55°N,175°E
14 LOW-PRESSURE	June	3M	29	Measures strength of Aleutian Low centered at 45°N,165°W
15 HT-PNA-STAND	Jan	3M3	22	Measures geopotential height field (variable 7), centered at 50°N,125°W multiplied by strength of pattern with centers at 55°N,115°W;45°N,165°W; and 30°N,85°W
16 NPAC	May	3M	30	Measures strength of a pattern with centers near 40°N,170°W; 70°N,180°W; and 40°N,130°E
				Hemispheric variables[4]
17 SOI	Feb	M	33	Measures east-west sea-level pressure gradient in the tropics
18 SUNSPOT	Oct	M3	25	Measures sunspot numbers
19 SUNSPOT2	June	M3	17	Measures sunspot numbers

[1]M, monthly variable; 3M, 3-month mean centered at indicated month; M3 or 3M3, value of variable is cubed.

[2]Lead time of listed month prior to the November seasonal inflow; for example, 1 month would be October.

[3]In units of degrees Fahrenheit.

[4]Except for the SOI (variable 17), all atmospheric-circulation variables are calculated from 700-millibar diamond-gridded data; the SOI is calculated using sea-level pressure at Tahiti and Darwin, Australia; STAND, standardized variable: monthly value of variable is standardized by subtracting the mean monthly value for the complete period of record and then dividing by the standard deviation; a dipole consists of 700-mb grid points, generally representing the height difference between a high and a low pressure; a pattern or height field consists of more than two grid cells used to calculate the variable.

This north-south dipole measures the relative strength of a low-pressure region near the typical location of the Aleutian Low (LOW-PRESSURE) and a high-pressure region centered well offshore of central Mexico that is consistent with the generation of storms and storm tracks. The WPO4 variable is similar to the EASTPAC except that the high-pressure region is much more to the west. Both WPO4 and EASTPAC, which have the same lead time, account for persistence in the atmospheric signal over broad areas just after the winter season. The apparent signal of these two patterns, as seen in the inflow data, changes from the EASTPAC for the September–October inflows to the WPO4 for the October inflows, suggesting the importance of a westward translation of a high-pressure region during this lead time. The NPAC variable measures a pattern that covers a reasonably large area of the northern part of the western North Pacific. The strength of the NPAC is also an indicator of the future strength of both the EASTPAC and the WPO4, and the lead times of the three suggest not a seasonal, but rather an annual progression of influential atmospheric-circulation patterns.

The last three measures (variables 17-19) are classified as hemispheric. The strong influence of the Southern Oscillation/El Niño phenomenon on global climate and hydrology is well documented. In addition, its influence on the hydrology of the Pacific Northwest also is well documented (Cayan and Peterson, 1989; Redmond and Koch, 1991; Cayan and Webb, 1992; Vaccaro, 1996). At first, the SOI was excluded from the equations because its large-scale influence on the climate system might interfere with the signal from other important variables. However, the February SOI was significantly related to both the September–October and October inflows at the 33-month lead time. As a result, the SOI was included. Moreover, because of the overall weak precursor climate signal contained in these first 2 months of the fall-transition season, some kind of explanatory variable was needed that accounted for variance in the inflow data. Sunspot numbers are known to be correlated with many geophysical phenomena, but the cause-effect relations are unknown and vigorously debated. The cube of the sunspot number provides estimating ability and is not significantly correlated to most other variables. Thus, sunspot variables 18 and 19 are included in two equations, resulting in more reliable and independent equations.

The Forecasting Equations

The five multiple linear regression equations developed to forecast inflow to the Hanson Reservoir are presented below for each inflow period and each equation lead time. The equations were developed using the inflows for the complete 1952–96 period.

Equations 1 and 2 are for September–October inflows and they use explanatory variables with values from about 3 years to 13 months prior to September. Equations 1 and 2 have lead times of 13 and 14 months, respectively. Equations 1 and 2 share five variables, and equation 2 has two additional variables, HIGH-PRESSURE and sst1. Sst1 essentially builds on the concurrent strong relation between a central North Pacific cold-water pool and the average geopotential height-field over northwestern Washington (HEIGHT-STAND), and the former apparently accounts for persistence in the high-pressure cell. Both equations calculate nearly the same inflow values.

Equations 3 and 4 are for October inflows and they use explanatory variables with values from about 3 years to 2 months prior to October. Equations 3 and 4 have lead times of 9 months and 2 months, respectively. The lead times of the variables in equation 4 (see table 1) indicate that they capture a distinct temporal progression of climate influences on subsequent October flows.

Equation 5 is for November inflows and it uses explanatory variables with values from 30 to 14 months prior to November. Equation 5 has a lead time of 14 months. All the variables but one are for the period May–August, and four of the seven variables are for the period of May–June. The clustering of explanatory variables during these times indicates that the summer transition is an important period for defining persistence in the system, which is obviously complexly related to the November inflows.

The five regression equations are as follows.

Regression equations for September–October inflows

- With a 13-month lead time:

$$\begin{aligned}\text{SEASONAL INFLOW} = {} & -22.97 \times \text{Cedar L. Precip} \\ & - 70.72 \times \text{SOI} - 8.767 \\ & \times \text{NPAC} + 5.734 \\ & \times \text{EASTPAC} + 139.04 \\ & \times \text{HEIGHT-STAND} + 2065.3 \end{aligned} \quad (1)$$

- With a 14-month lead time:

$$\begin{aligned}\text{SEASONAL INFLOW} = {} & -21.65 \times \text{Cedar L. Precip} \\ & - 64.897 \times \text{SOI} - 9.085 \\ & \times \text{NPAC} - 7.591 \\ & \times \text{HIGH-PRESSURE} \\ & + 4.795 \times \text{EASTPAC} \\ & + 119.499 \times \text{HEIGHT-STAND} \\ & + 31.85 \times \text{ss1} + 2965.3 \end{aligned} \quad (2)$$

Regression equations for October inflows.

- With a 9-month lead time:

$$\begin{aligned}\text{SEASONAL INFLOW} = {} & -47.38 \times \text{SOI} - 13.32 \\ & \times \text{NPAC} - 14.767 \\ & \times \text{HIGH-PRESSURE} \\ & - 13.165 \times \text{HEIGHT} \\ & + 3.771 \times \text{WPO4} \\ & + 130.013 \times \text{ss2} + 2607.2 \end{aligned} \quad (3)$$

- With a 2-month lead time:

$$\begin{aligned}\text{SEASONAL INFLOW} = {} & -50.58 \times \text{SOI} - 10.778 \\ & \times \text{NPAC} - 20.732 \\ & \times \text{HIGH-PRESSURE} \\ & - 1.893 \times 10^{-5} \\ & \times \text{SUNSPOT} + 3.246 \\ & \times \text{WPO4} + 77.47 \times \text{ss2} \\ & + 122.17 \times \text{sst4} - 3056.6 \end{aligned} \quad (4)$$

Regression equation for November inflows

- With a 14-month lead time:

$$\begin{aligned}\text{SEASONAL INFLOW} = {} & -281.16 \times \text{NSGEO} \\ & + 71.81 \times \text{Cedar L. Temp} \\ & - 29.46 \times \text{LOW-PRESSURE} \\ & + 29.242 \times \text{WPO1} - 428.32 \\ & \times \text{HT-PNA-STAND} + 7.088 \\ & \times 10^{-5} \times \text{SUNSPOT2} \\ & + 130.75 \times \text{sst3} + 608.4 \end{aligned} \quad (5)$$

RESULTS FROM CALIBRATION AND TESTING OF EQUATIONS

The initial equations were calibrated using stepwise regression based on 27 years of record from the 1952-96 period. To obtain the partial period, the lower, middle, and upper thirds of the years were ranked by September–October inflows, and nearly equal numbers of years were randomly selected from each third. The equations were tested using the remaining 18 years, which is a conservative 60/40-percent split of the record. Water years 1992–96 were included only in the 18-year test period in order to assess the equations under more current conditions. This split-sample procedure allowed for an assessment of potential performance of the equations under both past and future conditions. The final equations were calibrated with the complete 45-year record and were tested for water year 1997. Results from the calibration and testing are shown in table 2.

The testing results from the split-sample calibrated equations (table 2) indicate that the equations provide reasonable and somewhat independent estimates of the seasonal inflows and suggest how the final equations may perform in the forecast mode. The r-squared, standard error, and standard deviation values reflect the estimating ability of the equations. About 80 percent of all the residual values (observed minus predicted) were less than the standard error determined during calibration. That is, more than 68 percent (the expected percentage from regression analysis) of the predicted values from the testing period were less than a standard error from the observed values. The ranges and averages of the residuals of the five equations for the testing period were as follows.

Inflow Period	Residuals (ft^3s^{-1}) Range	Average
September–October (equations 1 and 2)	-345 – 557	78
October (equations 3 and 4)	-375 – 458	5
November (equation 5)	-986 – 590	-156

Table 2. Results from the initial calibration and testing of five multiple linear regression equations and the final calibration of the equations

[Equations developed using data from 1952–96. Equations initially were based on a sample size of 27 years for calibration and 18 years of testing. Final equations were calibrated using the complete 45 years; --, not applicable; all values rounded]

	Inflow, in cubic feet per second, except for r-squared, dimensionless											
	Mean				Standard deviation						Standard error of estimate	
	calibration		Testing		Calibration		Testing		r-square			
Equation[1]	Observed	Calculated	Observed	Calculated	Observed	Calculated	Observed	Calculated	Calibration	Testing	Calibration	Testing
					September–October inflow equations							
Equation 1												
Initial	688	684	765	663	453	421	392	246	0.86	0.44	205	303
Final	719	719	--	--	427	357	--	--	0.71	--	251	--
Equation 2												
Initial	688	688	765	658	453	420	392	246	0.87	0.43	194	305
Final	719	719	--	--	427	359	--	--	0.70	--	248	--
					October inflow equations							
Equation 3												
Initial	448	448	514	516	364	351	313	347	0.84	0.57	167	210
Final	474	475	--	--	343	295	--	--	0.74	--	188	--
Equation 4												
Initial	448	448	514	512	364	334	313	286	0.92	0.48	113	232
Final	474	475	--	--	343	307	--	--	0.80	--	166	--
					November inflow equation							
Equation 5												
Initial	1,229	1,208	1,429	1,585	851	769	960	675	0.81	0.80	419	440
Final	1,309	1,309	--	--	891	806	--	--	0.82	--	415	--

[1]Equations given in text in the section "Forecasting Equations."

The average residuals and the testing results in table 2 show that the September–October equations (1 and 2) have a bias toward estimating smaller-than-average values, the October equations (3 and 4) have little bias, and the November equation (5) has a bias toward predicting larger-than-observed values. Some of the bias in the split-sample testing is retained in the final calibrated equations, discussed later in the section Assessment of Performance and Reliability of Equations. The equations generally performed well under the more recent 1992-96 conditions. The testing results indicate that the performance of the final equations when applied in the forecasting mode should be reasonable.

The observed and predicted values from the final five equations are shown on figures 2-6 and the calibrated statistics are presented in table 2. The r-squared value for the final September–October equations 1 and 2 was about 0.70 with a standard error of about 250 ft^3s^{-1}, about 40 percent smaller than the standard deviation of the observed values and 65 percent smaller than the mean (table 2). Although there is a slight difference in estimating ability between equations 1 and 2 (table 2), it is not easily discerned when comparing the difference in the scatter about the diagonal between figures 2 and 3. The additional two variables used in equation 2 should help to produce somewhat different estimates from those of equation 1, but the overall results will be similar.

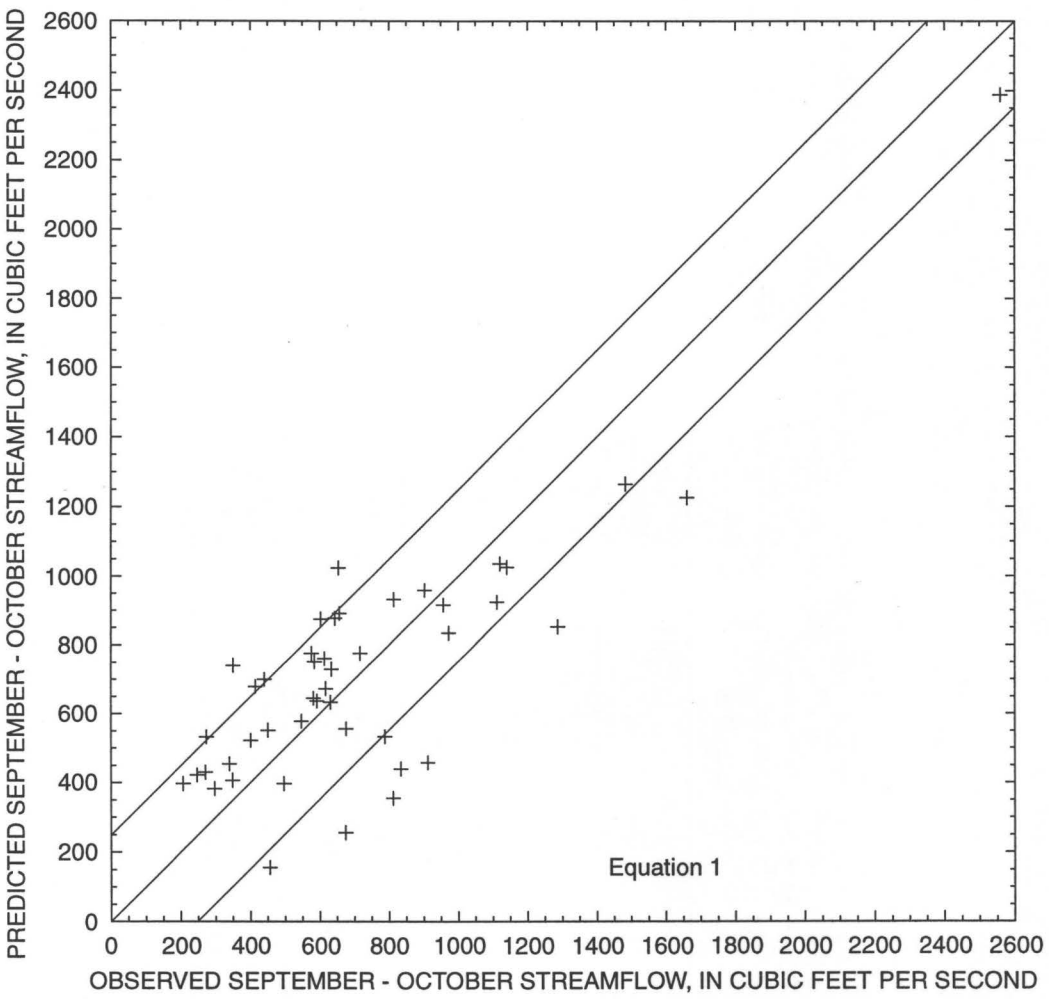


Figure 2. Observed and predicted values for the September-October inflow season, equation 1. The central diagonal line represents a theoretical line of equal observed and predicted values, and the overall scatter about the line reflects the standard deviation of the residuals. Values on the axes are the full range of observed values for the inflow period.

In practical terms, for potential future use of the equations the forecasted inflows from equations 1 and 2 may best be averaged.

The final October equations (3 and 4) had r-squared values of 0.74 and 0.80 and standard errors of 188 ft^3s^{-1} and 166 ft^3s^{-1}, respectively (table 2). Comparing figures 4 and 5 illustrates the slightly improved fit of equation 4 versus equation 3. The figures also illustrate that the equations produce estimates that generally differ, and yet retain about the same distribution.

In comparison to the other equations, equation 5 better represents the observed values (fig. 6), with an r-squared value of 0.82 and a standard error of 415 ft^3s^{-1} (table 2). Most of this improvement can be accounted for by the fact that the November inflows generally contain a much stronger climate signal than either the September or October inflows, and this signal is apparently captured by the variables in equation 5. In western Washington, the third largest precipitation of all the months happens in November, and thus November is strongly linked to the winter atmospheric-circulation regime.

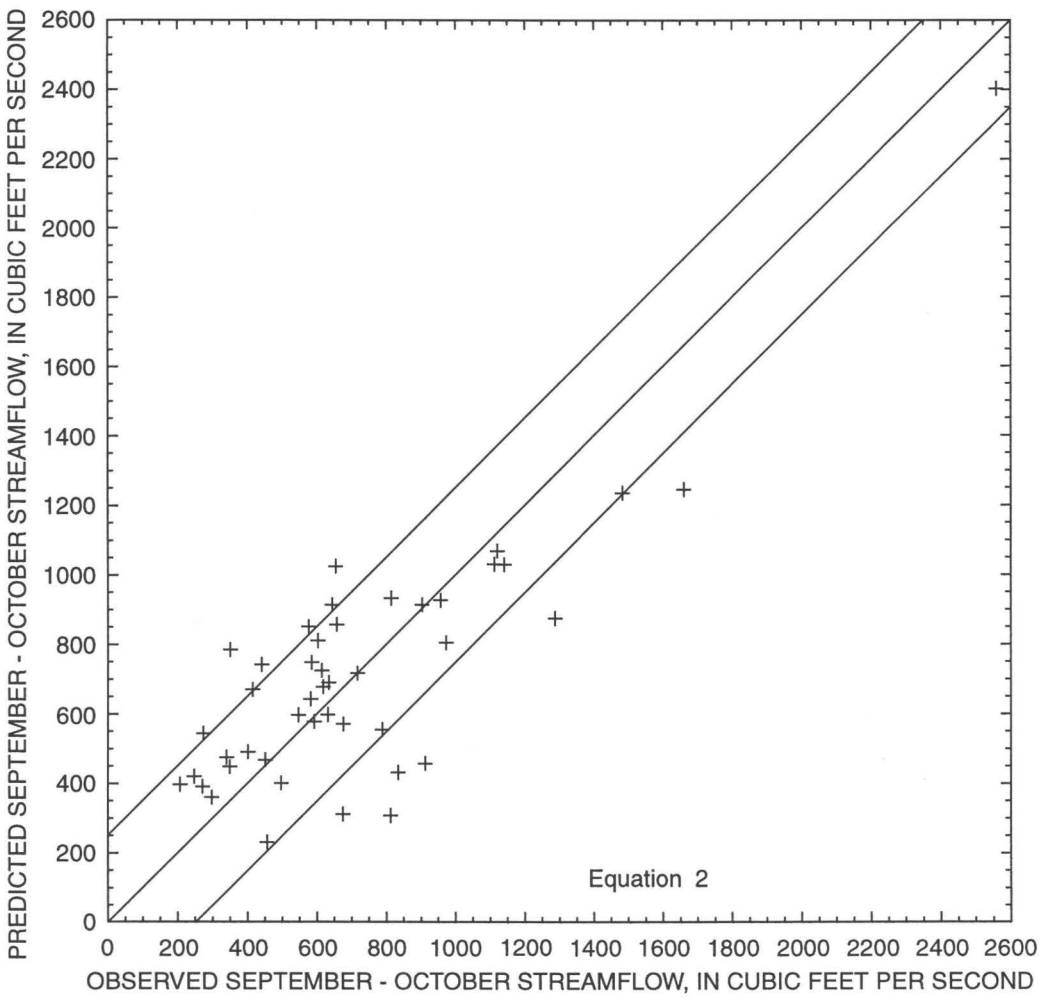


Figure 3. Observed and predicted values for the September-October inflow season, equation 2. The central diagonal line represents a theoretical line of equal observed and predicted values, and the overall scatter about the line reflects the standard deviation of the residuals. Values on the axes are the full range of observed values for the inflow period.

The persistence in the atmospheric-oceanic system (which the equations essentially attempt to account for) is reflected in the November inflows. The standard error, which reflects the standard deviation of the residuals, is considerably less than the standard deviation of the observed values, as demonstrated on figure 7. This figure also illustrates the bias to predicting larger-than-observed values (negative residuals)

A good estimate of the error in values calculated by each equation is the standard error of estimate of the final calibrated equations (table 2). The potential errors in predicted values for equations 1-5 are, respectively, 251 ft^3s^{-1}, 248 ft^3s^{-1}, 188 ft^3s^{-1}, 166 ft^3s^{-1}, and 415 ft^3s^{-1}. The smaller an estimated inflow value is, the larger its potential percent error would be. The equations will yield good results on average, but not necessarily in every individual year. The reason is that the equations are statistical; that is, forecasts for some years may be out on the distribution tail strictly by chance or because of changes in the climate regime.

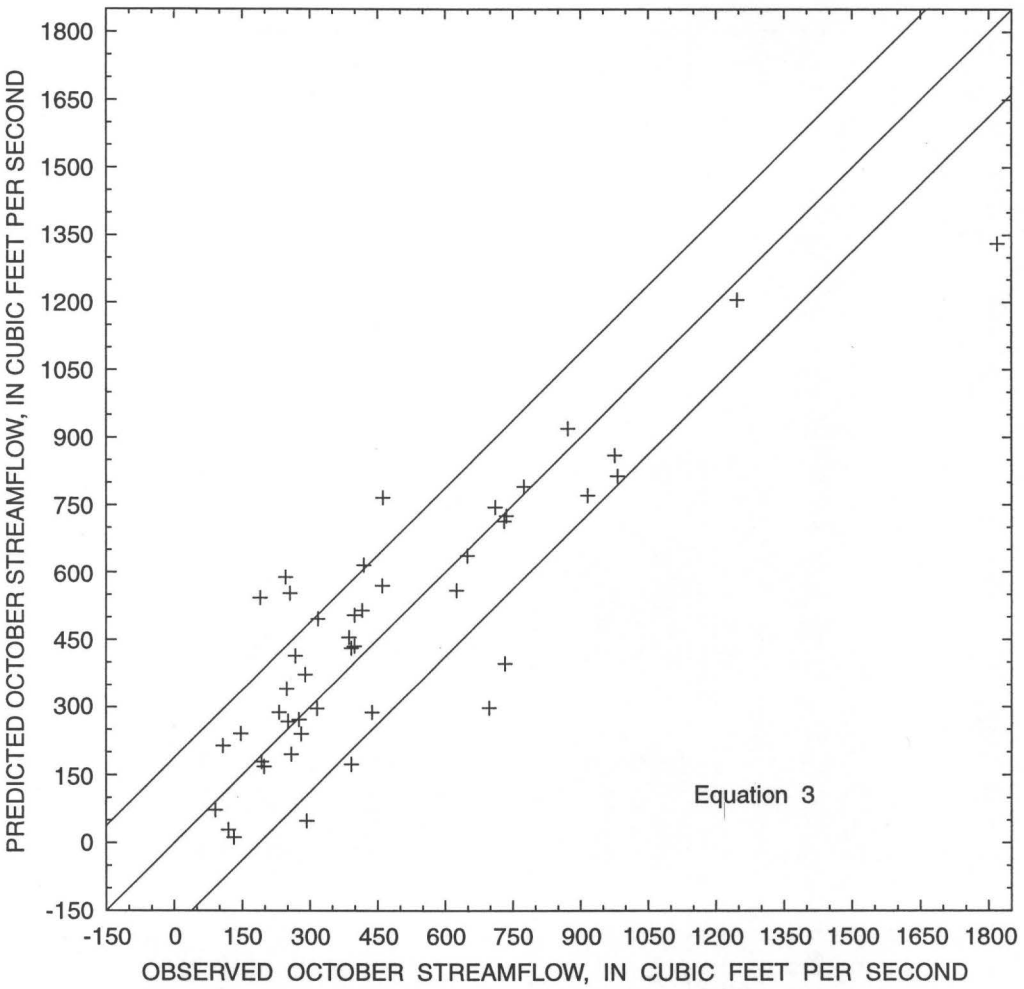


Figure 4. Observed and predicted values for the October inflow season, equation 3. The central diagonal line represents a theoretical line of equal observed and predicted values, and the overall scatter about the line reflects the standard deviation of the residuals. Values on the axes are the full range of observed values for the inflow period.

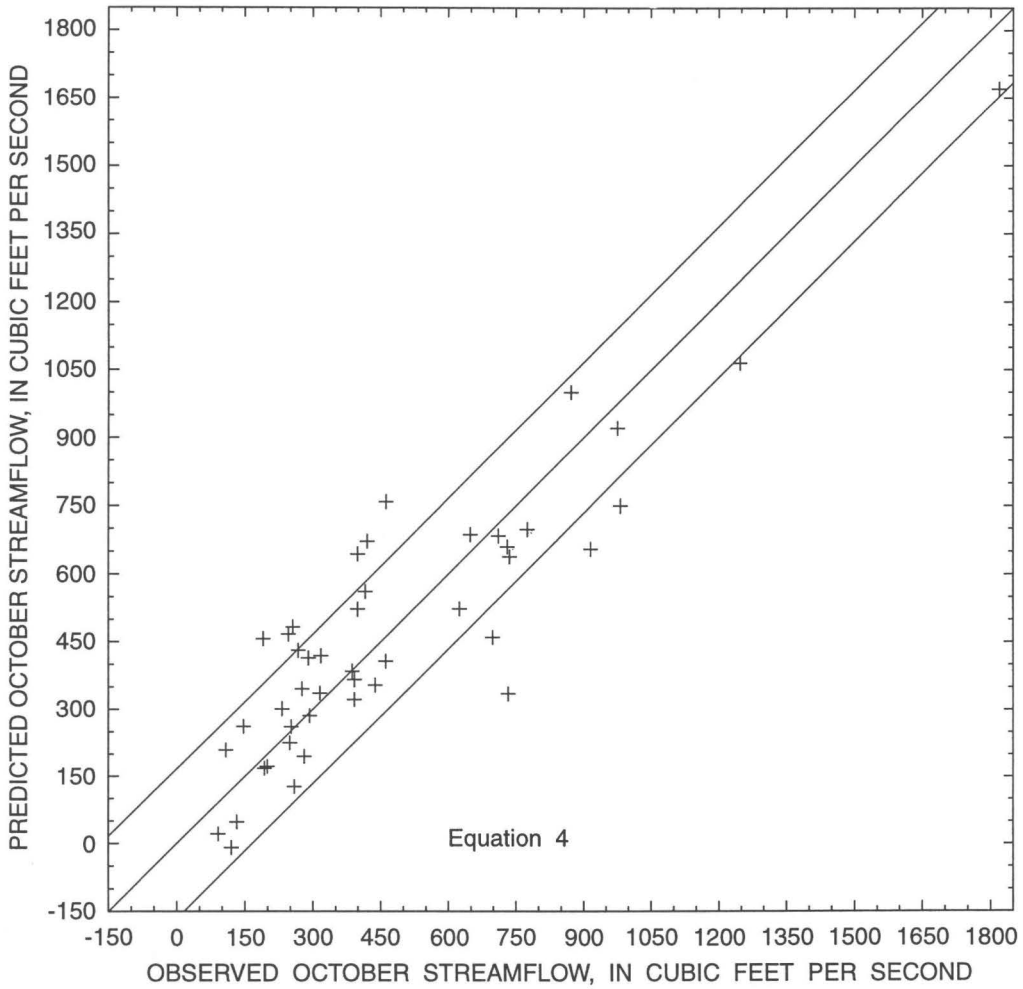


Figure 5. Observed and predicted values for the October inflow season, equation 4. The central diagonal line represents a theoretical line of equal observed and predicted values, and the overall scatter about the line reflects the standard deviation of the residuals. Values on the axes are the full range of observed values for the inflow period.

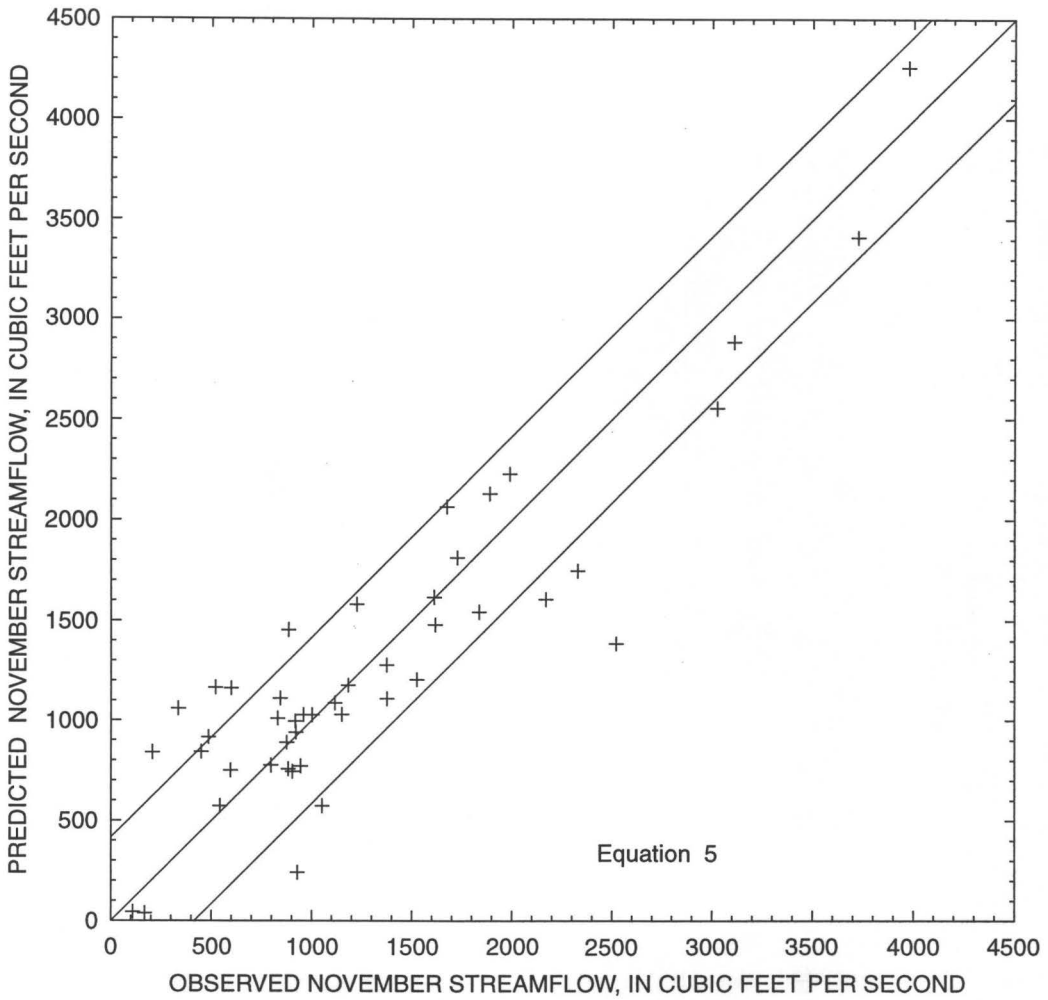


Figure 6. Observed and predicted values for the November inflow season, equation 5. The central diagonal line represents a theoretical line of equal observed and predicted values, and the overall scatter about the line reflects the standard deviation of the residuals. Values on the axes are the full range of observed values for the inflow period.

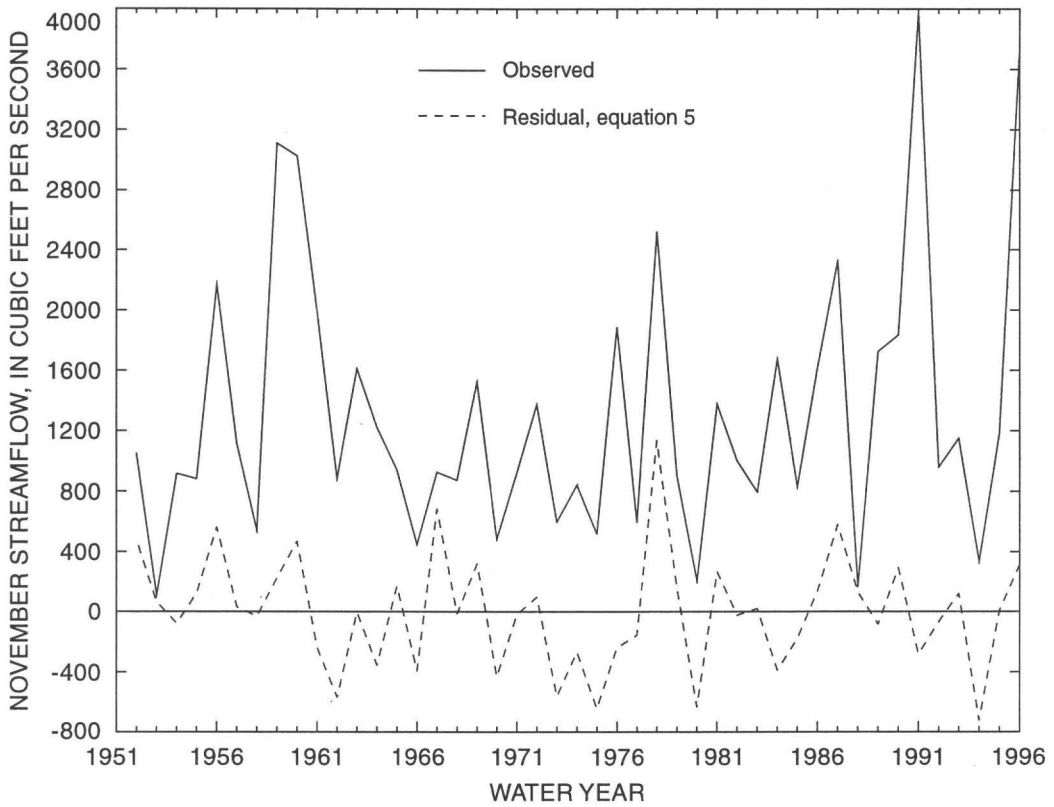


Figure 7. Observed November inflows and residual values calculated from equation 5.

ASSESSMENT OF PERFORMANCE AND RELIABILITY OF EQUATIONS

The performance and reliability of the final equations were assessed by (1) applying the equations to forecast inflow for the 1997 water year and then comparing forecasted with observed values for the three inflow periods for calendar year 1996 (September of water year 1996 and October–November of water year 1997—the fall-transition season inflows for water year 1997), and (2) analyzing the results for high (generally the upper 70th-percentile values) and low (generally the lower 30th-percentile values) inflow periods. The latter was done because the primary interest is in providing long-lead estimates for potential for wet or dry fall-transition seasons. In addition, the ability of the equations to capture the concurrency of high or low inflows for October and November was also assessed.

The observed inflows, forecasted inflows, and percentiles for the fall-transition period for water year 1997 are:

Inflow period	Observed[1] inflow (ft^3s^{-1})	Percentile	Forecasted inflow (ft^3s^{-1})	Percentile[2]
September–October				
Equation 1	1,623	96th	1,020	83rd
Equation 2	1,623	96th	959	79th
October				
Equation 3	1,464	98th	848	85th
Equation 4	1,464	98th	774	83rd
November				
Equation 5	2,771	91st	1,512	69th

[1]Thomas Murphy, written commun., U.S. Army Corps of Engineers, 1997.

[2]Based on the observed historical record.

The observed inflows for the fall-transition period for water year 1997 were some of the largest on record, and the forecasted inflows were also well-above average. Although the monthly mean flow for September was a 30th-percentile value (based on the 1952–97 period of record), both the observed and forecasted inflows for the September–October period were large. The large values demonstrate both the differences between the monthly mean September and October inflows and the fact that equations 1 and 2 typically capture the variability in October inflows. In terms of standardized series (standardized series have a mean of zero and a standard deviation of 1), the observed flows ranged from 1.63 for November to 2.66 for October. The forecasted inflows ranged from 0.26 for November to 1.16 for October, and except for November, all values were more than one-half of a standard deviation above the mean. Overall, this first application of the equations to future conditions produced reasonable estimates.

The reliability of the equations for forecasting low- and high-flow periods was assessed using standardized series for the observed and calibrated values. Using standardized data allows for the comparison of two series whose distributions or values may not be the same. For the inflow periods during the years that the observed values were more than half a standard deviation from the mean (values greater than 0.5 or less than -0.5), the observed and calibrated values (figures 8–10) are compared graphically for each inflow period. Values less than -0.5 represent low flows and values greater than 0.5 represent high flows. The calibrated values for the corresponding observed low- and high-flow inflows are shown in the table below by inflow period and by the number of occurrences in each standardized discharge range.

Equations used to compute calibrated discharge	Observed low flows[1]: The number of times that the calibrated discharge is in these ranges of standardized units				Observed high flows[1]: The number of times that the calibrated discharge is in these ranges of standardized units				Percentage of calibrated low and high flows corresponding to observed low and high flows
	Less than -0.5	-0.5 to 0	0 to 0.5	Greater than 0.5	Less than -0.5	-0.5 to 0	0 to 0.5	Greater than 0.5	
	September–October inflow period; 14 low flows and 9 high flows observed from 1953–96								
1	10	3	1	0	0	0	2	7	74
2	10	2	2	0	0	0	2	7	74
	October inflow period; 19 low flows and 13 high flows observed from 1952–96								
3	13	3	3	0	1	1	0	11	75
4	13	5	1	0	0	2	0	11	75
	November inflow period; 13 low flows and 10 high flows observed from 1952–96								
5	7	6	0	0	0	0	3	7	61

[1]Low flow is defined as a discharge of less than -0.5 standardized units; high flow is defined as a discharge of greater than 0.5 standardized units.

If the match between observed and calibrated values was extremely good—that is, the number of calibrated low- and high-flow values would be equal to the respective number of observed values—then all values would plot in either the lower left (low-flow values) or the upper right (high-flow values) quadrant on figures 8–10. For equations 1 and 2 (September–October inflows with 23 observed values—14 low flow, 9 high flow), 74 percent of the calibrated values plot in the correct quadrant (fig. 8, above table). For equations 3 and 4 (October inflows with 32 values—19 low flow, 13 high flow), 75 percent of the calibrated values plot in the correct quadrant (fig. 9).

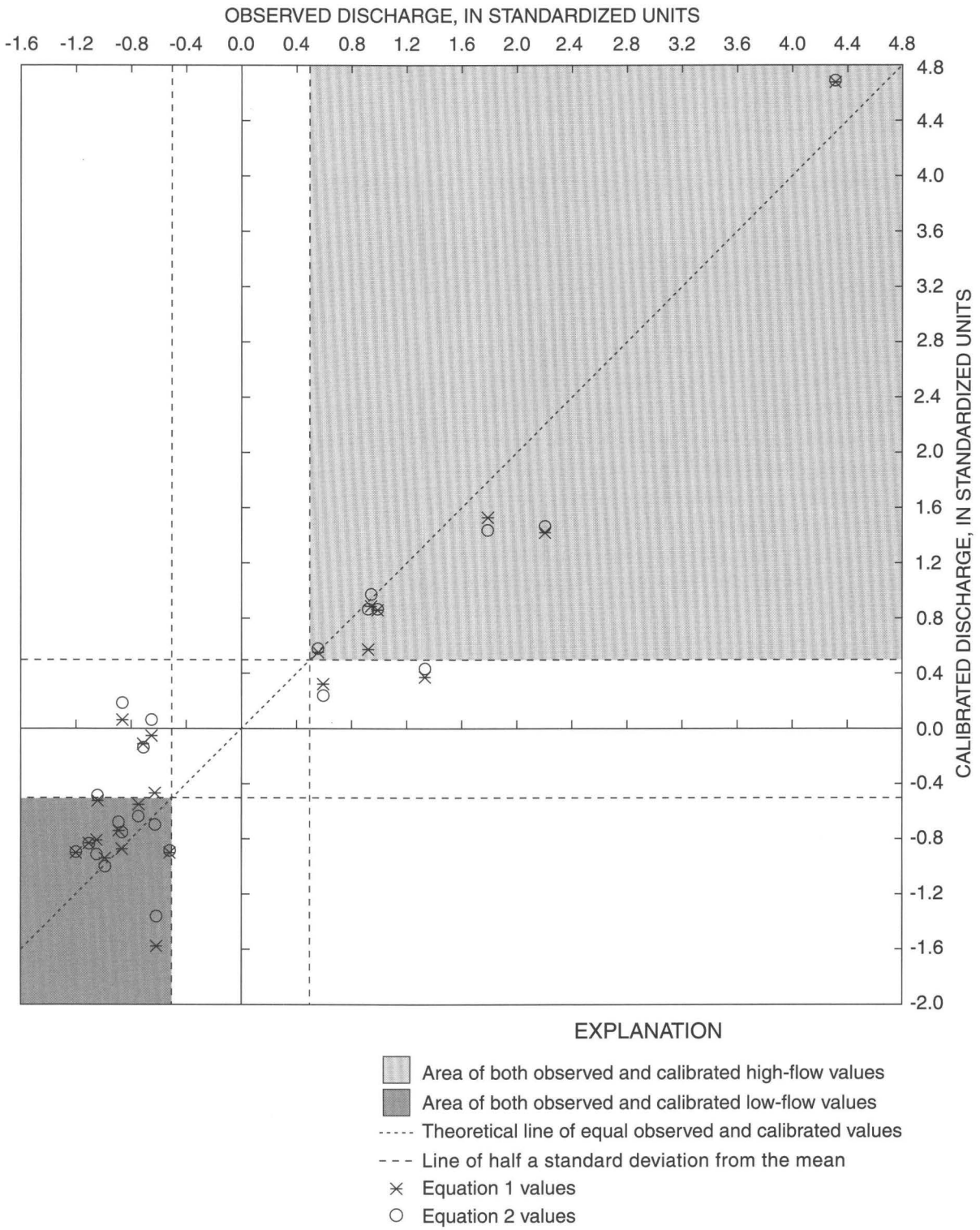


Figure 8. Standardized values of observed and calibrated inflows for the September-October inflow season, equations 1 and 2. Calibrated values are only shown for observed high flows (observed values > 0.5) and observed low flows (observed values < -0.5).

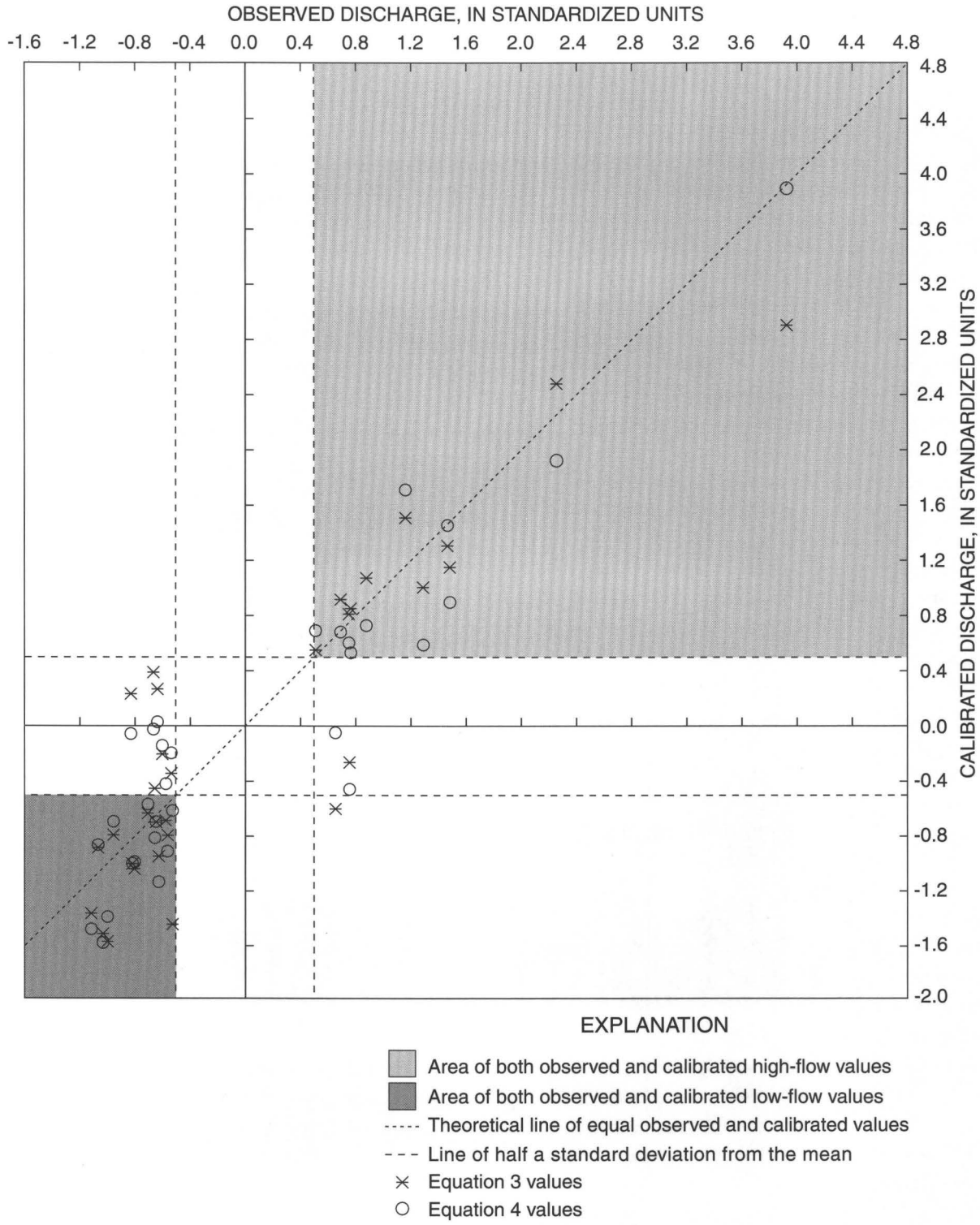


Figure 9. Standardized values of observed and calibrated inflows for the October inflow season, equations 3 and 4. Calibrated values are only shown for observed high flows (observed values > 0.5) and observed low flows (observed values < -0.5).

For October, fully 68 percent of the inflows were more than half a standard deviation from the mean, further showing that October inflows clearly represent two distinct populations. For equation 5 (November inflows with 23 observed values—13 low flow, 10 high flow), 61 percent of the calibrated values plot correctly (fig. 10). Thus, with a reasonable probability, the five equations are able to capture the occurrence of low- and high-flow discharge values. All but 7 of the 79 calibrated low-flow values were less than the mean and only 4 of the 54 possible high-flow values were less than the mean. This suggests that the equations also provide reliable estimates of below- or above-average inflows for the fall-transition season.

The closeness of values to the diagonal on figures 8-10 demonstrates the match of the calibrated values to observed values for extreme years, and grouping of values above or below the diagonal indicates the potential bias. A bias toward over-prediction—predicting larger-than-observed values for high-flow years or smaller-than-observed for low-flow years—would be better than a bias toward under-prediction. Over-prediction identifies occurrences of inflows more than half a standard deviation from the mean. In under-prediction, calculated values are not only closer to the mean but they may also be above the mean during a low-flow year and below the mean during a high-flow year.

Equations 1 and 2 have a bias to under-predict the extremes (fig. 8). The preponderance of the calibrated values lie above the diagonal for low flows (in three cases greater than the mean) and below the diagonal for high flows (no occurrences of less than the mean). Although eight values from equations 3 and 4 were of the opposite sign of the observed values (fig. 9), these equations have less overall bias than equations 1 and 2. There also is a tendency for the standardized values for low- and high-flow years from both equations 1 and 2 to be more similar to those from equation 4 than from equation 3.

The calibrated values for November (fig. 10) show little bias and a good fit for the high-flow years. For low-flow years, there is a reasonable fit but a bias for calculating larger-than-observed values (under-prediction). This bias also occurred for the low-flow, split-sample testing years. All equations generally capture the occurrences of high flows better than low flows, especially equations 3 and 4 (fig. 10, above table).

For each inflow period there were more occurrences of low-flow years than high-flow years, indicating the increased probability for the fall-transition season to be dry. Indeed, for October, there were 6 more low-flow years than high-flow years, representing about 13 percent of the 45-year period of record used in the analysis. However, the range of the low-flow values is much smaller than the range of high-flow values (figs. 8-10). For example, for the three inflow periods, the range between the upper 30th-percentile values varies from four to nine times the range between the lower 30th-percentile values. These differences, which are also exhibited by the Cedar River inflows, identify important aspects of the hydrology of the two basins and their relation to climate forcing—the basins can only get so dry, but they can get very wet. The equations are better able to capture the wet extremes because of the much larger range in the wetter years.

The climate regime differs greatly between low- and high-flow years because low and high flows generally are derived from two separate populations, each with specific characteristics. If one considers only the question of whether the fall-precipitation season will start in October or November, then these differences are quite important. For example, for the 45 years of record used in this analysis, 13 of the standardized October inflows had values greater than 0.5 and one October had a value of 0.44. In practical terms, about 30 percent of the time the precipitation season began in October and the amount of precipitation was enough to sustain higher flows. The remaining 70 percent of the time, October inflows were smaller than the mean (about 70 percent of which could be considered low-flow years). The October equations captured the occurrence of the onset of the fall-precipitation season about 85 percent of the time. The climate regime that drives the wet Octobers is apparently accounted for in equations 3 and 4. There were only six occurrences of concurrent high flows during both October and November, but all of the November standardized values during high-flow Octobers were greater than -0.5. Given that in the historical record no very-low-flow November years occurred during high-flow Octobers, the October equations should provide reasonable estimates if the fall precipitation season will not only start in October, but may continue through November.

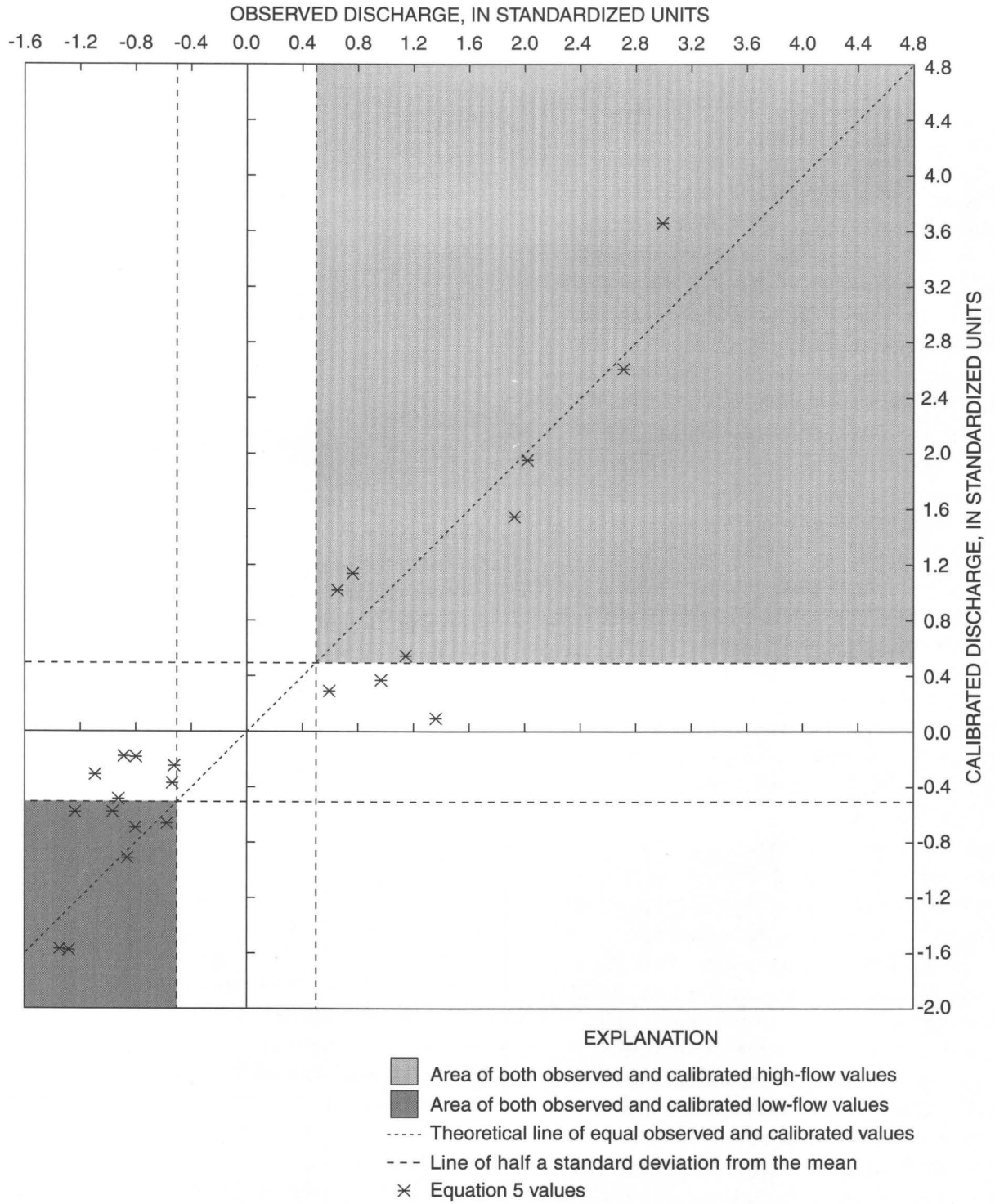


Figure 10. Standardized values of observed and calibrated inflows for the November inflow season, equation 5. Calibrated values are only shown for observed high flows (observed values > 0.5) and observed low flows (observed values < -0.5).

One reason that there were only six concurrent occurrences of high flows during both October and November is that for several of the high-flow Octobers with average November inflows, the precipitation during November was above normal but the air temperatures were such that much of the precipitation fell as snow.

Given the two distinct climatic regimes, another question of interest is whether the equations capture the concurrency of a low-flow October and a below-average (but not necessarily low-flow) November. For the 19 low-flow Octobers, 15 Novembers met this criterion; the equations calculated 14 of these 15 concurrent low-flow periods correctly. In addition, of the 28 years that November standardized inflows were less than 0.0, 13 were less than -0.5 and all but 2 of these years had concurrent low-flow Octobers; that is, low-flow Novembers are generally preceded by low-flow Octobers. The results of equations 3-5, taken together, should provide useful information on whether October and November will be wet or dry; see, for example, figure 11. The information shown on figure 11 also suggests that a standardized value provides reliable information and should be considered in conjunction with the actual forecast value.

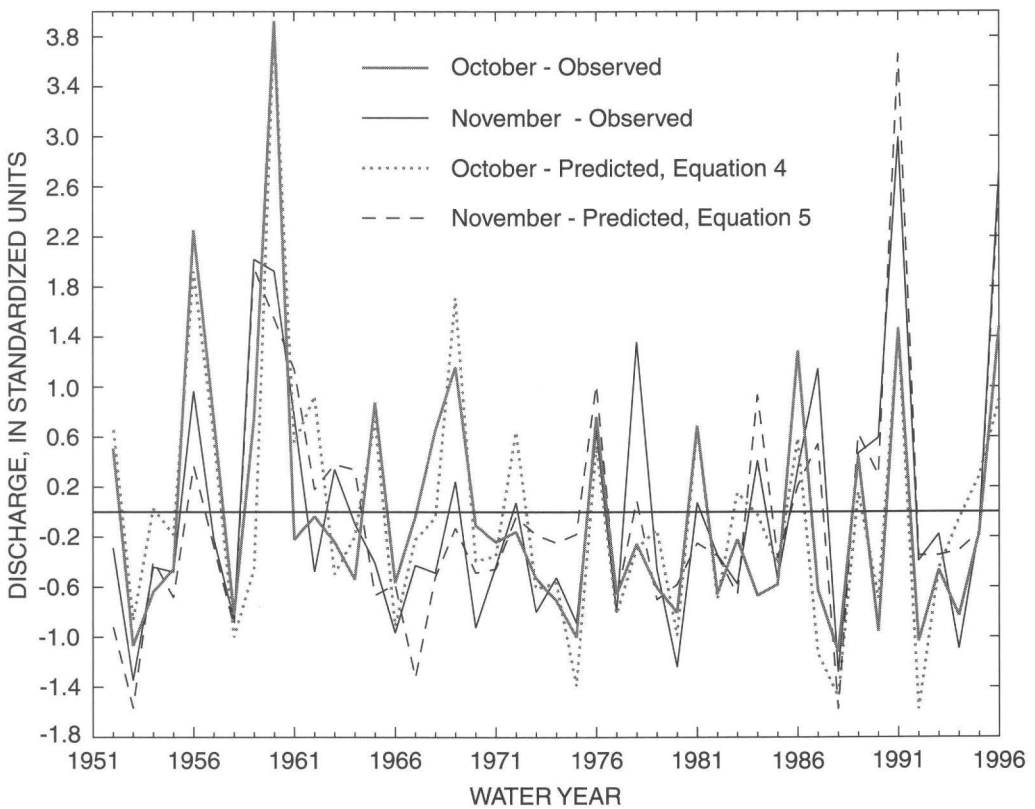


Figure 11. Standardized values of observed and predicted inflows for the October and November inflow seasons. The predicted inflows are from equations 4 and 5.

POTENTIAL USE AND PROBLEMS WITH EQUATIONS

The equations may be used for reservoir management in several ways. One way is to calculate the inflows and then account for the values when making reservoir release decisions throughout the year. Another way that they may be used is, if equations 1 and 2 (13- and 14-month lead times) calculate a 20th-percentile value for the following September–October inflow, then the observed inflows from January through October for the years that produced, say, 1st- through 40th-percentile inflows could be used in a management model to assess potential management strategies. This approach could also be used with the results of equation 3 at 9 months lead time and equation 4 at 11 months lead time. Management strategies might include hedging releases at certain times in order to augment flows for the September–October period.

The accuracy of the equations is influenced by the strength and position of atmospheric- and oceanic-circulation patterns. Indeed, at times particular circulation patterns may not even exist. This inherent problem leads not only to a loss of accuracy, but also to the fact that the equations essentially act as detectors of climate-regime shifts. The equations may no longer be valid at some future period because of an overall realignment of atmospheric circulation and sea-surface temperature anomaly patterns. Much of this problem is due to the highly nonlinear and dynamic nature of the ocean-atmosphere system and because, climatologically, the period of record used in this study is short. This problem can be partly overcome by using variables that are more dynamically based; for example, using a pressure-gradient variable rather than the currently used pressure-difference variable. Developing more dynamically based variables for the equations may increase their overall accuracy.

SUMMARY

The U.S. Geological Survey, in cooperation with the Tacoma Public Utilities, Seattle Public Utilities, and the U. S. Army Corps of Engineers, developed a method for estimating, at long-lead times, the fall-transition seasonal inflow to the Howard A. Hanson Reservoir. The method uses multiple linear regression equations to estimate inflow values. The equations were developed for three inflow periods: September–October, October, and November. Two regression equations were developed for each of the first two inflow periods and one equation for the November inflows. The equations for the September–October inflows are at 14- and 13-month lead times, the October equations are at 9- and 2-month lead times, and the November equation is at a 14-month lead time. The equations use monthly and 3-month averages of variables that are calculated mainly from atmospheric and sst data.

The equations were initially calibrated using 27 years of data from the 1952–96 period of record. Inflows for years 1992–96 were saved for testing the equations in order to assess the estimating ability of the equations for the most recent period of record used in this study. Testing the equations using 18 years of data resulted in r-squared values that ranged from 0.43 to 0.80. This initial calibration and testing procedure suggested how the equations may perform under future conditions. The final five equations were calibrated using the complete 1952–96 period of record. The final calibrated equations had r-squared values that ranged from 0.70 to 0.82. The equation for the November inflows has the largest r-squared value, and the equations for the September–October inflows have the smallest r-squared values. These results are consistent with the fact that the climate signature becomes stronger as the fall-transition season progresses.

The results and assessment of the final equations showed that the equations generally are able to capture the occurrence of low-flow and, especially, high-flow years. The October equations can provide a reasonable estimate of whether the onset of the fall-precipitation season will begin in October. In addition, the assessment indicates that standardized values should be considered in conjunction with the actual forecast values, especially for the November forecasts, due to the error and bias in the estimates.

The ability to predict reservoir inflows with long lead times can be used to help manage releases from the reservoir throughout the water year. The final equations, with results reformulated in terms of percentiles or standardized units, are also applicable to Chester Morse Lake on the nearby Cedar River. Similarly, the equations may also be applicable to other western Washington streams.

REFERENCES

Barnett, T.P., and Preisendorfer, R.W., 1978, Multifield analog prediction of short-term climate fluctuations using a climate state vector: Journal of Atmospheric Science, v. 35, p. 1771-1787.

Barnston, A.G., and Livezey, R.E., 1987, Classification, seasonality and persistence of low-frequency atmospheric circulation patterns: Monthly Weather Review, v. 115, p. 1083-1126.

Bjerknes, J.H., 1969, Atmospheric teleconnections from the equatorial Pacific: Monthly Weather Review, v. 97, p. 162-172.

Cayan, D.R., and Peterson, D.H., 1989, The influence of North Pacific atmospheric circulation on streamflow in the west, *in* Peterson, D.H., ed., Aspects of climate variability in the Pacific and Western Americas: Washington, D.C., American Geophysical Union, Geophysical Monograph 55, p. 375-397.

Cayan, D.R., and Webb, R.H., 1992, El Niño/Southern Oscillation and streamflow in the western United States, *in* Diaz, H.F. and Markgraf, V.M., eds, El Niño, Historical and paleoclimatic aspects of the Southern Oscillation: Cambridge, England, Cambridge University Press, p. 29-68.

Christensen, R.A., Eilbert, R.F., Lindgren, O.H., and Rans, L.L., 1981, Successful hydrologic forecasting for California using an information theoretic model: Journal of Applied Meteorology, v. 20, p. 706-713.

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

Ebbesmeyer, C.C., Coomes, C.A., Cannon, G.A., and Bretschneider, D.E., 1989, Linkage of ocean and fjord dynamics at decadal period, *in* Peterson, D.H., ed., Aspects of climate variability in the Pacific and Western Americas: Washington, D.C., American Geophysical Union, Geophysical Monograph 55, p. 399-417.

Gladwell, J.S., 1970, Runoff generation in western Washington as a function of precipitation and watershed characteristics: Pullman, Wash., Washington State University, College of Engineering, Research Division, R. L. Albrook Hydraulic Laboratory, Bulletin 319, 341 p.

Horel, J.D., and Wallace, J.M., 1981, Planetary-scale atmospheric phenomena associated with the Southern Oscillation: Monthly Weather Review, v. 109, p. 813-829.

Namias, J., 1981, Teleconnections of 700 mb height anomalies for the northern hemisphere: La Jolla, Calif., Scripps Institution of Oceanography, CALCOFI Atlas 29, 265 p.

Namias, J., Yuan, X., and Cayan, D.R., 1988, Persistence of North Pacific sea surface temperature and atmospheric flow patterns: Journal of Climate., v. 1, p. 682-703.

Rasmussen, L.A., and Tangborn, W.V., 1976, Hydrology of the North Cascades region, Washington 1. Runoff, precipitation, and storage characteristics: Water Resources Research, v. 12, no. 2, p. 187-202.

Redmond, K.T., and Koch, R.W., 1991, Surface climate and streamflow variability in the Western United States and their relationship to large-scale circulation indices: Water Resources Research, v. 27, no. 9, p. 2381-2399.

Rogers, J.C., 1981, The North Pacific Oscillation: Journal of Climatology, v. 1, p. 39-57.

Smith, T.M., Reynolds, R.W., Livezey, R.E., and Stokes, D.C., 1996, Reconstruction of historical sea surface temperatures using empirical orthogonal functions: Journal of Climate, v. 9, p. 1403-1420.

Speers, Pamela, and Mass, C.F., 1986, Diagnosis and prediction of precipitation in regions of complex terrain: Washington State Department of Transportation, WA-RD 91.1, 166 p.

Torranin, Padoong, 1972, Application of canonical correlation in hydrology: Fort Collins, Colo., Colorado State University, Hydrology Papers, No. 58, 30 p.

Vaccaro, J.J., 1996, Changes in the hydrometeorological regime in the Pacific Northwest, *in* Isaacs, C. M. and Tharp, V. L, eds., Proceedings of the twelfth annual Pacific climate (PACLIM) workshop: Technical Report 46 of the Interagency Ecological Program for the Sacramento-San Joaquin Estuary, April 1996, p. 143.

Yarnal, B., and Diaz, H.F., 1986, Relationships between extremes of the Southern Oscillation and the winter climate of the Anglo-American Pacific coast: Journal of Climatology, v. 6, p. 197-219.

APPENDIXES

Appendix 1. Descriptive statistics for monthly, annual, and seasonal (September–October) inflows of the Green River to the Howard A. Hanson Reservoir, Washington

[Monthly unregulated flows provided by the U.S. Army Corps of Engineers. Abbreviations used are: Station ID,U.S. Geological Survey identification number for reservoir stage gage; MEDN, median value; MAXM, maximum value; MINM, minimum value; SDEV, standard deviation; CVAR, coefficient of variation; Sept.-Oct., sum of monthly mean values for September and October]

Howard A. Hanson Reservoir Green River, Storage Capacity of 105,463 acre-feet

Station identification number	Drainage area (square miles)	Start year	End year	Number years of record	Altitude gage (feet)	Location of dam	
						Latitude°N	Longitude°W
12105800	220	1915	1996	82	1,141	47.2778	121.7856

Inflow, in cubic feet per second

	Oct.	Nov.	Dec.	Jan.	Feb.	March	Apr.	May	June	July	Aug.	Sept.	Annual	Sept.-Oct.
MEAN	500.46	1,236.07	1,615.28	1,495.36	1,342.89	1,169.55	1,504.23	1,478.25	955.98	381.25	201.02	220.75	1,008.42	723.80
MEDN	392.00	943.50	1,432.00	1,409.00	1,197.50	1,055.00	1,501.00	1,466.00	812.50	308.00	190.50	170.50	997.25	618.00
MAXM	1,818.50	3,974.50	5,293.00	4,394.00	3,445.92	3,439.00	2,490.50	2,842.00	2,802.00	1,076.00	389.50	788.50	1,567.37	2,558.50
MINM	90.50	108.50	202.50	208.00	265.00	491.50	600.00	434.00	186.00	110.00	116.00	97.00	488.08	207.00
SDEV	368.79	857.77	948.56	826.28	689.92	496.33	413.03	564.92	574.96	199.06	60.22	124.17	231.41	440.48
SKEW	1.56	1.19	1.51	0.65	0.91	2.02	-0.08	0.39	1.20	1.43	1.08	2.45	0.15	1.77
CVAR	0.74	0.69	0.59	0.55	0.51	0.42	0.27	0.38	0.60	0.52	0.30	0.56	0.23	0.61

Appendix 1. Descriptive statistics for monthly, annual, and seasonal (September–October) inflows of the Green River to the Howard A. Hanson Reservoir, Washington–Continued

Inflow, in cubic feet per second

	Oct.	Nov.	Dec.	Jan.	Feb.	March	Apr.	May	June	July	Aug.	Sept.	Annual	Sept.-Oct.
Quartile														
5	119.77	120.00	424.00	383.00	426.00	542.00	754.50	517.00	252.00	164.00	124.00	115.00	637.04	257.00
10	155.50	334.69	601.50	481.00	540.00	677.50	878.50	745.00	301.00	197.00	132.00	120.00	688.29	288.00
15	190.48	421.00	749.50	555.00	636.00	738.00	1,004.50	873.72	429.00	214.20	142.00	131.00	744.25	328.50
20	207.50	520.00	807.50	633.50	690.00	797.50	1,069.00	920.00	444.50	225.50	150.00	139.00	785.96	348.76
25	232.00	606.00	879.50	707.50	779.00	833.00	1,179.00	1,031.00	513.50	237.00	154.50	147.50	834.08	400.50
30	252.00	691.00	975.00	907.50	879.00	877.50	1,254.00	1,050.00	538.02	257.00	159.50	153.50	848.25	425.00
35	258.50	796.50	1,120.50	1,006.00	967.00	926.50	1,335.00	1,199.00	686.50	263.00	161.50	158.50	885.08	455.50
40	280.50	882.00	1,314.00	1,093.00	1,050.82	946.50	1,382.50	1,250.50	730.00	276.50	172.50	161.50	911.83	571.00
45	318.00	917.00	1,390.50	1,286.00	1,131.50	983.07	1,483.50	1,341.50	758.00	284.50	181.00	165.00	933.54	591.50
50	392.00	943.50	1,432.00	1,409.00	1,197.50	1,055.00	1,501.00	1,466.00	812.50	308.00	190.50	170.50	997.25	618.00
55	412.00	1,002.00	1,524.00	1,477.00	1,271.50	1,081.50	1,555.00	1,509.00	861.00	329.50	193.50	175.00	1023.00	644.00
60	461.00	1,116.00	1,641.00	1,563.50	1,342.50	1,166.00	1,607.00	1,599.50	915.50	350.00	202.00	199.50	1052.79	664.00
65	502.00	1,225.00	1,687.00	1,683.00	1,422.50	1,193.50	1,680.00	1,613.00	968.00	390.00	209.00	206.00	1061.79	731.00
70	538.00	1,447.50	1,799.00	1,830.00	1,517.50	1,257.00	1,713.00	1,655.00	1,057.50	416.00	211.00	234.50	1135.33	788.00
75	648.50	1,617.00	1,958.50	2,030.00	1,618.50	1,380.00	1,795.00	1,788.00	1,131.00	445.00	221.00	238.00	1188.71	834.00
80	730.50	1,763.50	2,154.50	2,242.00	1,734.00	1,462.50	1,861.00	1,878.00	1,418.50	474.50	231.00	246.00	1214.08	941.00
85	783.00	1,988.50	2,162.00	2,482.00	2,021.50	1,561.00	1,908.00	2,169.50	1,464.50	579.00	269.00	335.00	1244.75	1037.00
90	916.00	2,462.00	2,481.00	2,613.00	2,292.00	1,610.00	1,939.00	2,233.00	1,641.50	686.00	280.00	344.00	1346.00	1169.50
95	1,247.00	2,786.00	3,231.50	2,696.00	2,537.00	1,732.50	2,072.50	2,416.00	2,092.00	754.50	299.50	424.00	1375.83	1513.00

Appendix 1. Descriptive statistics for monthly, annual, and seasonal (September–October) inflows of the Green River to the Howard A. Hanson Reservoir, Washington–Continued

Autocorrelation coefficients, number years = 82; number of lags = 29

Lag	Oct.	Nov.	Dec.	Jan.	Feb.	March	Apr.	May	June	July	Aug.	Sept.	Annual	Sept.-Oct.
1	0.04	0.06	0.03	0.19	-0.03	-0.08	0.06	0.22	0.10	0.24	0.13	0.01	0.13	0.01
2	-0.06	-0.08	0.04	0.03	-0.13	0.06	0.01	0.00	-0.03	0.05	0.05	-0.11	0.07	-0.10
3	-0.12	-0.14	0.00	-0.13	-0.06	-0.04	-0.01	0.15	-0.07	-0.01	-0.09	-0.09	-0.02	-0.13
4	-0.04	-0.02	-0.11	-0.26	0.12	0.08	-0.10	0.17	0.11	0.08	0.25	0.26	0.05	0.01
5	-0.01	0.20	-0.08	-0.04	-0.11	-0.11	0.04	0.15	0.11	0.08	0.10	0.13	-0.03	-0.04
6	0.03	-0.12	-0.09	0.04	-0.07	-0.16	0.01	-0.01	0.07	-0.07	0.02	0.04	-0.13	0.02
7	0.07	-0.21	-0.08	0.06	0.15	0.11	0.02	-0.04	-0.10	-0.03	0.11	-0.07	-0.01	-0.03
8	0.13	0.05	0.04	0.18	-0.05	-0.14	-0.02	0.00	-0.04	0.02	0.22	0.00	-0.05	0.06
9	0.07	0.01	0.01	-0.02	0.05	-0.21	-0.07	-0.16	-0.04	0.05	0.10	0.33	-0.07	0.12
10	-0.07	-0.05	-0.18	-0.03	0.21	-0.17	-0.07	0.09	0.18	0.05	-0.08	0.00	-0.04	-0.07
11	-0.16	-0.13	-0.14	-0.10	0.04	0.10	-0.01	0.04	-0.04	-0.09	-0.03	-0.09	0.00	-0.17
12	0.03	-0.09	-0.15	-0.08	-0.21	-0.01	-0.07	0.01	-0.08	-0.04	0.11	-0.02	0.00	0.07
13	0.20	0.35	0.04	-0.06	-0.10	-0.17	0.13	-0.01	-0.08	-0.14	0.03	0.21	0.03	0.25
14	0.19	-0.07	-0.02	0.09	0.18	0.11	-0.05	-0.04	0.00	-0.12	-0.01	0.15	0.12	0.14
15	-0.13	0.02	0.05	0.15	-0.10	0.07	-0.09	0.08	-0.04	-0.17	-0.16	-0.05	0.27	-0.17
16	-0.11	-0.21	0.25	0.00	-0.02	0.18	-0.09	-0.01	0.11	0.09	0.09	-0.15	0.01	-0.13
17	-0.08	-0.06	0.00	-0.01	-0.08	-0.13	-0.03	-0.14	-0.01	0.13	-0.07	-0.04	0.07	-0.01
18	-0.07	0.15	-0.17	-0.15	0.01	0.31	-0.02	-0.01	-0.04	-0.04	0.00	0.11	-0.09	-0.02
19	-0.16	-0.08	-0.03	-0.19	0.20	-0.08	0.03	0.04	0.07	0.10	-0.07	0.00	-0.18	-0.15
20	-0.06	-0.24	0.03	-0.03	0.01	-0.12	0.12	-0.05	-0.05	-0.06	0.00	-0.10	-0.26	-0.16
21	0.16	-0.05	0.03	-0.14	0.29	-0.02	-0.25	0.14	0.07	0.00	0.07	-0.08	-0.09	0.12
22	0.01	0.06	0.24	-0.11	-0.25	0.20	-0.12	-0.04	0.01	0.07	-0.02	-0.05	0.12	0.05
23	-0.05	0.03	-0.07	-0.14	0.12	-0.03	-0.05	-0.09	-0.14	-0.24	-0.23	0.05	-0.17	-0.02
24	-0.17	-0.12	0.13	-0.27	-0.15	-0.06	0.06	0.04	-0.12	-0.23	-0.31	0.08	-0.14	0.17
25	-0.02	-0.20	0.02	0.02	-0.10	0.05	-0.01	0.02	-0.05	-0.13	0.05	-0.19	0.03	-0.03
26	0.17	0.22	-0.07	-0.11	-0.16	-0.01	-0.07	0.16	0.11	-0.09	-0.02	-0.14	0.17	0.17
27	0.07	0.20	-0.20	0.00	0.09	-0.16	-0.07	0.06	-0.03	-0.09	0.01	0.11	-0.05	0.13
28	-0.05	0.25	-0.09	0.01	0.04	-0.24	-0.01	-0.12	-0.07	-0.17	-0.12	-0.06	-0.06	-0.11
29	0.01	-0.17	0.14	0.03	0.05	0.16	0.02	-0.09	0.05	0.04	0.03	-0.21	-0.04	-0.06

Appendix 2. Other Techniques Considered For Developing Equations To Forecast Inflows For The Fall-Transition Season

Initially, several techniques were explored for developing equations for forecasting inflows, from which the regression-equation method was chosen. One technique was to develop variables using principal-component analysis (PCA), a tool that has been used by many investigators in a large range of atmospheric and hydrologic studies. One of the early uses of PCA in western Washington was reported by Gladwell (1970), who applied PCA to climatological and physiographic variables to form explanatory variables for regression analysis in order to determine which ones were the most influential for explaining streamflow variability in western Washington. For the PCA used in the early stages of this study, the first three to four principal component (PC) time-series of about 25 variables were calculated using a common period of record (1947–94) for all variables. The PC time-series were then used as the explanatory variables in regression analysis. Conceptually, in using such a regression technique, current values of the time-series would be updated on the basis of the current values of the variables and the historical values of the loadings for each variable. This method assumes that the loadings would not need to be updated using PCA because they should be relatively stable due to the length (48 years) of the time-series used to develop them. Although initial results from this technique proved promising, it was not pursued because the additional data and work needed to calculate the many variables used in the PCA, and the formulation of the PC series, would tend to make a monthly, operational technique more complex.

Another technique was tested that also used PCA. In this technique, prior to calculating the variables for the atmospheric-circulation features, a time-series representing the first PC of the North Pacific sea-surface temperatures was used as a weighting factor for multiplying the 700-mb data. This time-series and the streamflow series in western Washington are coherent on the decadal scale. This type of weighting was further tested by standardizing each of the 5°–gridded sea-surface temperature time-series, and then multiplying each of the monthly 700-mb gridded time-series by the corresponding month of the nearest gridded value of the sea-surface temperature data. For both cases, the resulting 700-mb grid was reduced to the spatial extent of the sea-surface temperature data; thus, variables based on the 700-mb atmospheric-circulation features that extended beyond the spatial domain of the sea-surface temperature data set were not calculated. The weighted 700-mb data were then used to calculate variables for multiple linear regression equations. Results from this testing were not as good as other techniques. Use of only major hemispheric variables was also explored. For this case, only variables such as the SOI were analyzed in conjunction with the seasonal inflows. It was found that, for the fall-transition season, the signal of large-scale variables was both highly variable and not very strong.

Another technique explored was that of Barnett and Preisendorfer (1978), which they called multivariate analog prediction. This technique identifies regions where data, such as pressure heights and sea-surface temperatures, are highly related to the dependent variable. Data from these regions are then used to formulate independent series for analysis. The entropy minimax nonparametric-technique (Christensen and others, 1981; Christensen and Eilbert, 1985), canonical correlation analysis (CCA), and non-linear multiple regression were also briefly explored. These three techniques were felt to be either too complex for operational use or there was not enough time to fully explore them. Previously, Torranin (1972) tested CCA for "long-range prediction" using gridded sea-surface temperatures as independent variables and coastal precipitation in the Pacific Northwest (which is highly correlated to the streamflow in western Washington and precipitation at Cedar Lake) as a dependent variable; he found the results to be unreliable. However, the use of CCA by National Oceanic and Atmospheric Association's Climate Analysis Center for producing long-lead forecasts over much of the continental United States suggests that the spatial scale of the explanatory data used in this study may be such that CCA may be worth pursuing in conjunction with the already completed work.

☆ U.S. GOVERNMENT PRINTING OFFICE: 2000 — 589-121 / 03022 Region No. 10

USGS

J.J. Vaccaro

Development, Testing, and Assessment of Regression Equations for Experimental Forecasts of Fall-Transition-Season Inflows to the Howard A. Hanson Reservoir, Green River, Washington

USGS-WRIR 00-4153