Title
Seasonal streamflow prediction in Northern California basins using climate indices and Principal Component Regression

Author
Su, Xin

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Hydrology and Water Resources

THESIS

Submitted in partial satisfaction of the requirements
For the degree of

MASTER OF SCIENCE

In Civil Engineering

By

Xin Su

Thesis Committee:
Professor Kuolin Hsu, Chair
Associate Professor Xiaogang Gao
Assistant Professor Amir AghaKouchak

2015

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my appreciation to my committee chair, Professor Kuolin Hsu. His insightful questioning of concepts related to my research truly helped my knowledge, comprehension and understanding of the research process.

I would like to thank my committee members, Professor Amir AghaKouchak and Professor Xiaogang Gao, whose work demonstrates the concern for global affairs. Their support of an engagement in literature and modern technology should always transcend academia and provide a quest for our times.

In addition, I would like to thank Andrea Thorstensen for her great support in data collection as well as helping me understand the statistical models in this study.

Abstract of the thesis

Seasonal streamflow prediction in Northern California basins using climate indices and Principal Component Regression

By

Xin Su

Master of Science in Civil and Environmental Engineering

University of California, Irvine, 2015

Professor Kuolin Hsu, Chair

It has become more important in recent years to have a general estimate on how much water storage there will be during the spring season to support urban, industrial, and agricultural usage and forecasting drought. In this thesis, ground-based data and climate indices were applied to Principal Component Regression to build a prediction model to estimate seasonal streamflow in spring with winter temperature, precipitation and other climate indices. With data gathered from PRISM, NOAA, CEDC, and USGS and then testing all possible variables and combinations, it was found that using precipitation and temperature with climate indices such as SOI and PNA can provide useful information to forcast streamflow in spring season.

1. Introduction

Michael H. Glantzs illustrated that, "Drought is when the government sends you a report telling you there's no water," in his journal Consequences and Responsibilities in Drought Forecasting. Estimating seasonal water supply is an important task for almost all water supply services. Climate indices are calculated to describe the changes of the climate system, as streamflow is closely related to such factors as climate change, annual precipitation level, and the individual differences of reservoirs. In order to find out how strong climate indices are related to streamflow prediction, with a more adequate forecasting model the spring season streamflow accumulation can be estimated more accurately.

The climate indices are impelled by an assortment of parameters, such as sea surface temperature, precipitation, and air pressure. This information determines which climate variables in the winter season correlate to streamflow in the spring season, and indicate if climate indices like the Southern Oscillation Index would provide information in estimating spring season streamflow. The model will also test how different the simulation results will be by using varied numbers of climate indices. With the correlation coefficient between simulated result and observed result, it can clearly show which climate index is more effective for forecasting of streamflow in the spring season [Randall and Wood, 2007]. In this research, global climate index can be used as a variable to simulate seasonal streamflow for a specific basin that is to be tested.

There have been extensive streamflow prediction works in the past, such as the HEPEX prediction system and hybrid forecast. For many previous works, a gridded runoff dataset was used as a basic component [Rosenberg, 2011]. The goal of this study is to investigate the potential of providing an improved streamflow forecast in spring season using information from the hydrological variables in addition to major climate indices in winter season [Wei, 2010]. Climate variables such as precipitation and temperature, and other climate indices including Southern Oscillation Index (SOI), Pattern Northern-American index (PNA), and Pacific Decadal Oscillation (PDO) are selected. Those variables are then integrated in the regression analysis, mainly based on Principal Component Regression (PCR), to test how those variables are relevant to improving streamflow forecasting of spring season.

2. Climatological variables and Indices

2.1 Precipitation and Temperature

The basic data used in the research were collected from PRISM among the six elements, which are precipitation (ppt), minimum temperature (tmin), maximum temperature (tmax), dew point (tdmean), minimum vapor pressure deficit (vdpmin), and maximum vapor pressure deficit (vpdmax) [Descriptions of PRISM Spatial Climate Datasets for the Conterminous United States, 2013]. From these six PRISM climate elements, precipitation and mean temperature were selected for use in this study. The long-term average datasets are modeled with PRISM using a DEM as the predictor grid, and they used climatologically aided interpolation (CAI) to model time series. In CAI, the long-term average datasets serve as the predictor grids [Descriptions of PRISM Spatial Climate Datasets for the Conterminous United States, 2013]. The idea behind CAI is that the best first guess of the spatial pattern of climatic conditions for a given month or day is the long-term average pattern. CAI is robust to wide variations in station data density, which is necessary when modeling century-long time series [Descriptions of PRISM Spatial Climate Datasets for the Conterminous United States, 2013].

Fig. 1 illustrates monthly precipitation pattern time series for the time period from 1981-2013. The pattern shows that precipitation accumulated during winter time, especially from December to February. During the spring season when snow-melt water is released to the watershed is the major source of streamflow. Monthly temperature for the time period from 1981-2013 in Fig. 2, shows that the annual pattern of temperature is more constant than precipitation.

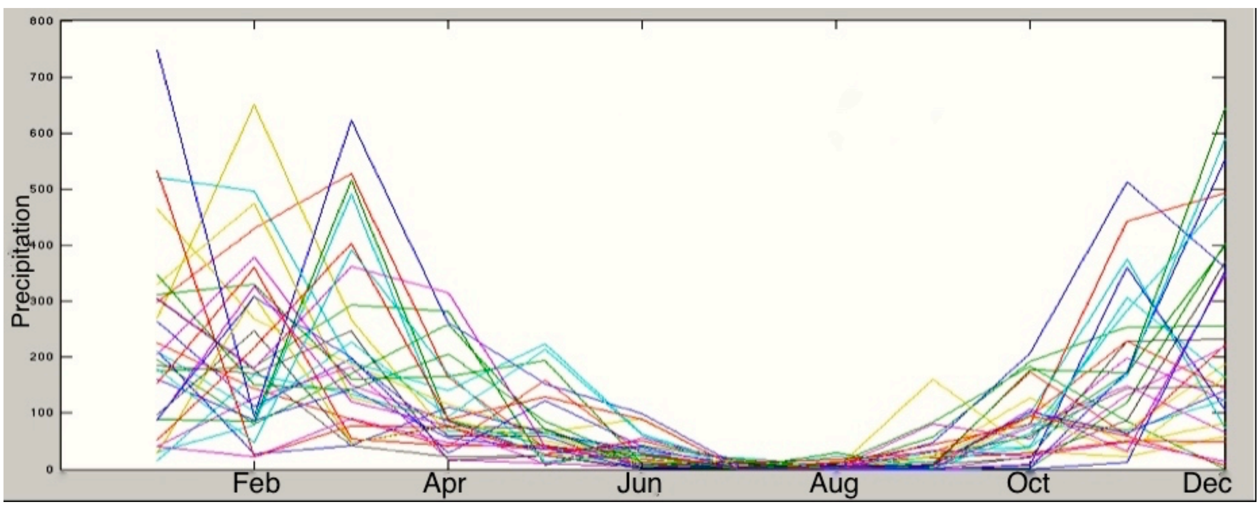


Figure 1 Monthly precipitation pattern from 1981-2103(mm)

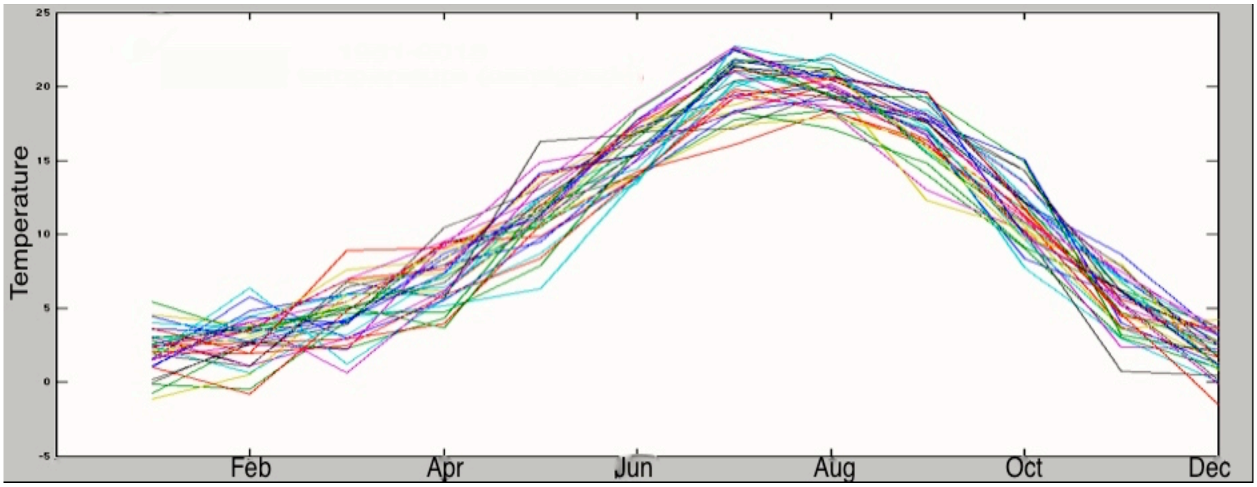


Figure 2 Monthly temperature pattern from 1981-2013(centigrade)

By plotting Feather River Mask, and apply to downloaded dataset from PRISM, regional average temperature and precipitation can be measured. The mask can be plotted by Arc-GIS. The resolution of PRISM dataset is 8km. For other usage, data of maximum temperature and minimum temperature are also offered.

2.2 Climate Indices

Climate indices are calculated to delineate the state and the changes in the climate system. Each climate index describes the certain climate phenomenon. Atmosphere parameters such as air pressure, air temperature, precipitation and sea surface temperature are all measurable parameters that influence the climate system *(Integrated Climate Data Center – ICDC).*

The classic maritime climate that occurs in the Pacific Coast is caused by mid-latitude weather systems moving from west to east [Kelly and Roy, 1991]. Kelly also claimed that, "Climate sensitivity is further amplified by growing populations in urban areas of the west, which are taxing available water for municipal and industrial uses."

2.2.1 SOI

The Southern Oscillation Index is a standardized monthly index, which is calculated from the observed sea level pressure differences between Tahiti and Darwin, Australia [Kelly and Roy, 1991; Können et al., 1998]. For a long period of record, SOI has been used as one measure of large-scale fluctuations in air pressure occurring between the western and eastern tropical Pacific during El Nino and La Nina episodes [Wikipedia; NOAA Climate Prediction Center]. SOI can be calculated with the functions below.

$$SOI = \frac{Standardized\ \ Tahiti - Standardized\ \ Darwin}{MSD}$$

$$Standardized\ \ Tahiti = \frac{Actual\ \ Tahiti\ \ SLP - Mean\ \ Tahiti\ \ SLP}{Standard\ \ Deviation\ \ Tahiti}$$

$$Standard\ \ Deviation\ \ Tahiti = \sqrt{\sum (actual\ \ Tahiti\ \ SLP - mean\ \ Tahiti\ \ SLP)^2 / N}$$

$$Standardized\ \ Darwin = \frac{Actual\ \ Darwin\ \ SLP - Mean\ \ Darwin\ \ SLP}{Standard\ \ Deviation\ \ Darwin}$$

$$Standard\ \ Deviation\ \ Darwin = \sqrt{\sum (actual\ \ Darwin\ \ SLP - mean\ \ Darwin\ \ SLP)^2 / N}$$

$$MSD = Monthly\ \ Standard\ \ Deviation = \sqrt{\sum (standardized\ \ Tahiti - Standardized\ \ Darwin)^2 / N}$$

$$N = number\ \ of\ \ months$$

Where SLP is sea level pressure, and MSD is monthly standard Deviation, while standard deviation of sea level pressure.

Kelly (1991) shows the largest correlation between SOI and sea surface temperature in Pacific Northwest. In addition, there will are tremendous differences in precipitation during the extremes of SOI [Kelly and Roy, 1991].

In Fig. 3, value above zero (positive), is abnormally cold ocean water of El Nino episodes, while negative shows the abnormally warm water of La Nina episodes [Chowdhury, 2003; Syed et al., 2006].

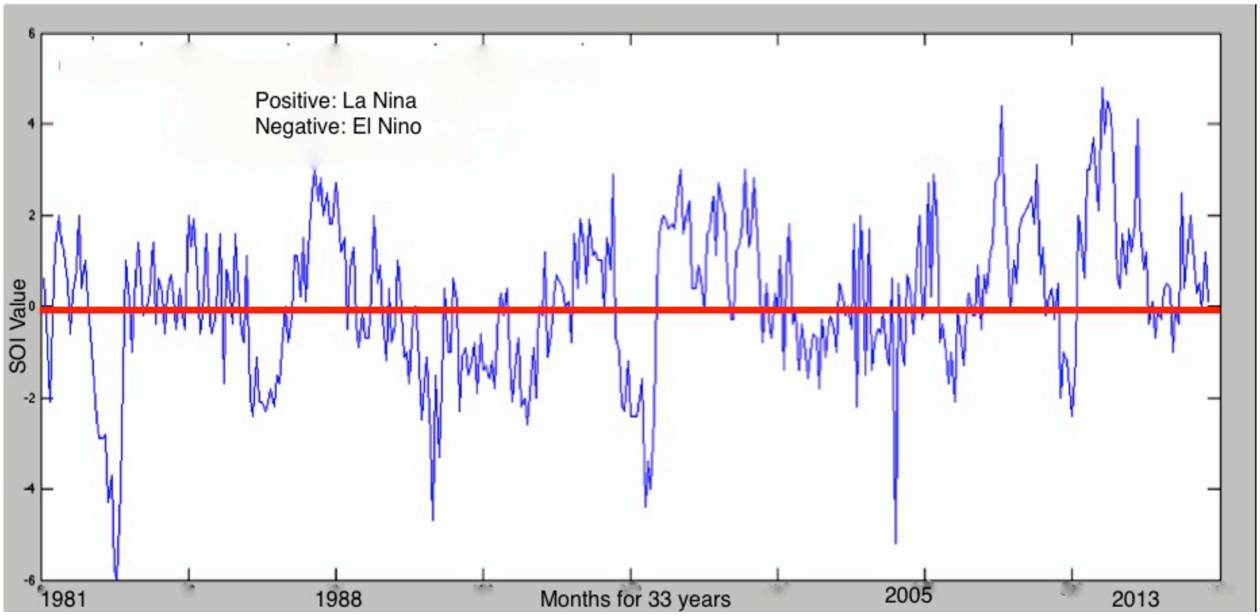


Figure 3 monthly Southern Oscillation index over 33 years (1981-2013)

2.2.2 PNA

The Pacific North America pattern is a quantitative index. PNA pattern is related to presence of warm water in the central Pacific [Kelly and Roy, 1991]. PNA index was defined by using a linear combination of standardized height anomalies near the three nominal centers of this pattern [Wallace and Gutzler, 1981; Yarnal and Diaz, 1986; Sergei R., Raymond A., 2001]. PNA was slightly modified by:

$$PNA=[-Z(50N,170W)+Z(50N,110W)-Z(30N,90W)]^3$$

Where Z (a,b) is 700 mbar monthly height departure from long-term monthly averages at latitude and longitude b [Kelly and Roy,1991; Sergei and Raymond, 2001]. When PNA index is positive, it means that there is an intensified western ridge and deeper troughs in the Aleutians and the U.S. southeast [Kelly and Roy, 1991; Sergei and Raymond, 2001]. The dryness appears in Southwestern America, which includes California, and is often associated with PNA, as in Fig. 4, positive value results in dryness.

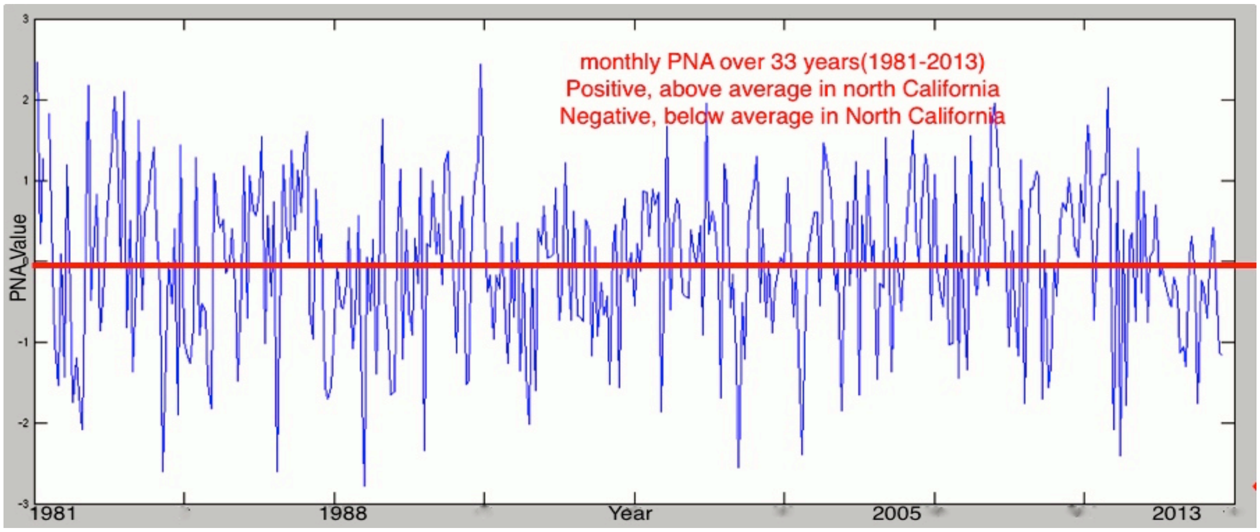


Figure 4 Monthly Pacific North California pattern index (PNA) over 33 years (1981-2013)

It shows high correlations between PNA index and temperature during winter season, however, correlations are also shown in differnent regions in the country during the spring and autumn season [Leathers et al., 1991].

2.2.3 PDO

A PDO index developed by Hare (1996) and Zhang (1996), and also used by Mantua et al. (1997), is the dominant PC from an un-rotated empirical orthogonal function (EOF) analysis of monthly residual North Pacific sea surface temperature (SST) anomalies, for the 1900-1993 period of record, poleward of 20°N. Residuals are the difference between observed anomalies and the monthly mean global average SST anomaly [Zhang, et al., 1997; Mantua and Hare, 2002]. The index shows noteworthy tendency for multi-year and multi-decadal consistency [Mantua and Hare, 2002].

Other studies also contribute evidence that PDO variations had considerable impact on climate-sensitive natural resources in the Pacific and over parts of North America in the 20th century [Mantua and Hare, 2002].

As PDO show positive, or warm, the west Pacific becomes cool and part of the eastern ocean warms. While it shows negative, or cool, the opposite pattern occurs. The inter-decadal time scale is usually about 20 to 30 years [Haggag et al., 2010]. During winter season, there is more precipitation than usual in southwestern United States. Moisture for California and the West are provided by the Pacific Ocean. When the waters are warm, the high temperature would heat the air above, and it would cause rain and snow over the cooler lands. While the ocean is cool, it shows the opposite [Reynolds et al., 1997].

Sea surface temperature (SST) tends to be abnormally cool in the central North Pacific, and coincides with atypically warm SSTs along the west coast of the Americas during warm PDO phases. Warm PDO sea level pressure (SLP) anomalies have low pressures over the North Pacific for November through March averages, which causes enhanced counter-clockwise winds. Fluctuations were most dynamic at periods in the 15-to-25 year and 50-to-70 year bands as found by Minobe (1999, 2000) when Wavelet analysis was applied to indices for boreal winter and spring North Pacific SST and SLP [Mantua and Hare, 2002].

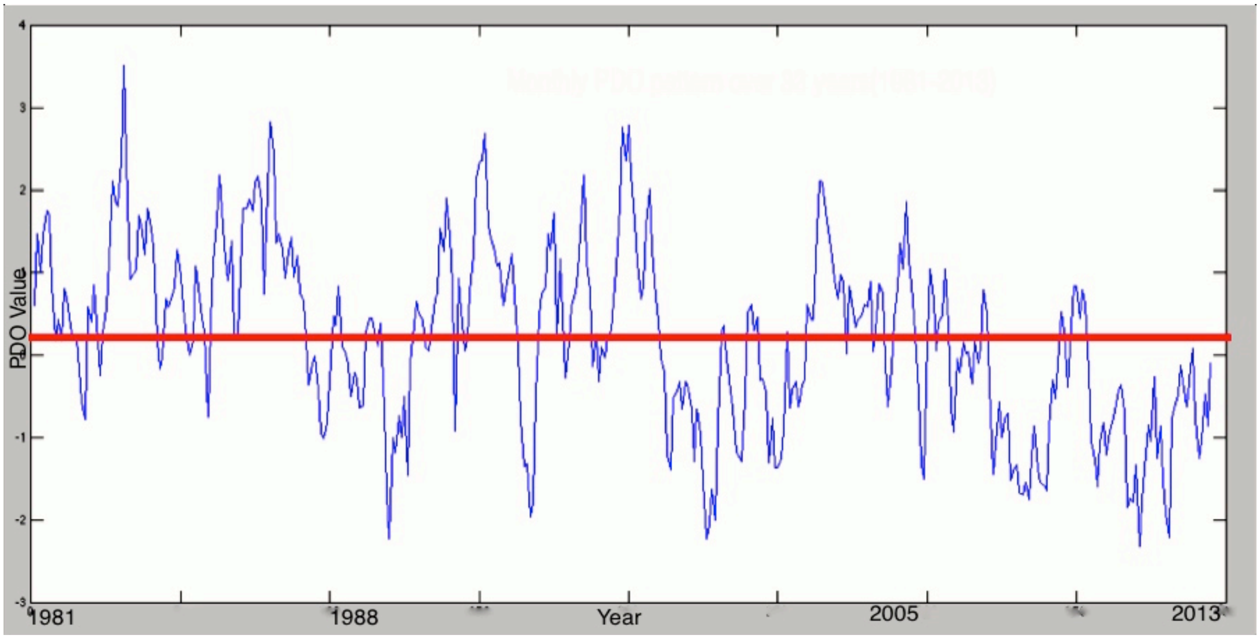


Figure 5 Monthly Pacific Decadal Oscillation index (PDO) over 33 year (1981-2013)

Precipitation and temperature anomalies is correlated to PDO. The warm phase of PDO Fig. 5 represents anomalously dry periods in much of Central America and northern South America. Patterns of November through April temperature anomalies and warm phases Fig. 5 tend to coincide with warm temperature in northwestern North America [Mantua and Hare, 2002].

3. Methodologies

3.1 Previous Work Review

3.1.1 Hydrological Ensemble Prediction Experiment (HEPEX)

After 2004, the development, testing, and operational management of forecasting systems, the Hydrologic Ensemble Prediction Experiment (HEPEX) has been widely introduced to other hydrologists or people working on hydrology predictions [Schaake et al., 2007]. Currently, HEPEX is co-chaired by NOAA, the European Centre for Medium-Range Weather Forecast and the European Commission Joint Research Centre [Schaake et al., 2007]. HEPEX is focused on the measurement of soil moisture, snow depth and water equivalent, and includes satellite and radar data by utilizing an ensemble hydrological system [Schaake et al., 2007]. In the streamflow prediction system, HEPEX is within a realistic spatial correlation of errors [Schaake et al., 2007]. HEPEX can also help educate users by adapting their existing practices (discussed in a 2002 Bulletin article by Zhu et al) [Schaake et al., 2007]. The main components and information that are in an ensemble hydrologic prediction system can be separated into two parts, the weather-climate ensembles and the land-surface state observations.

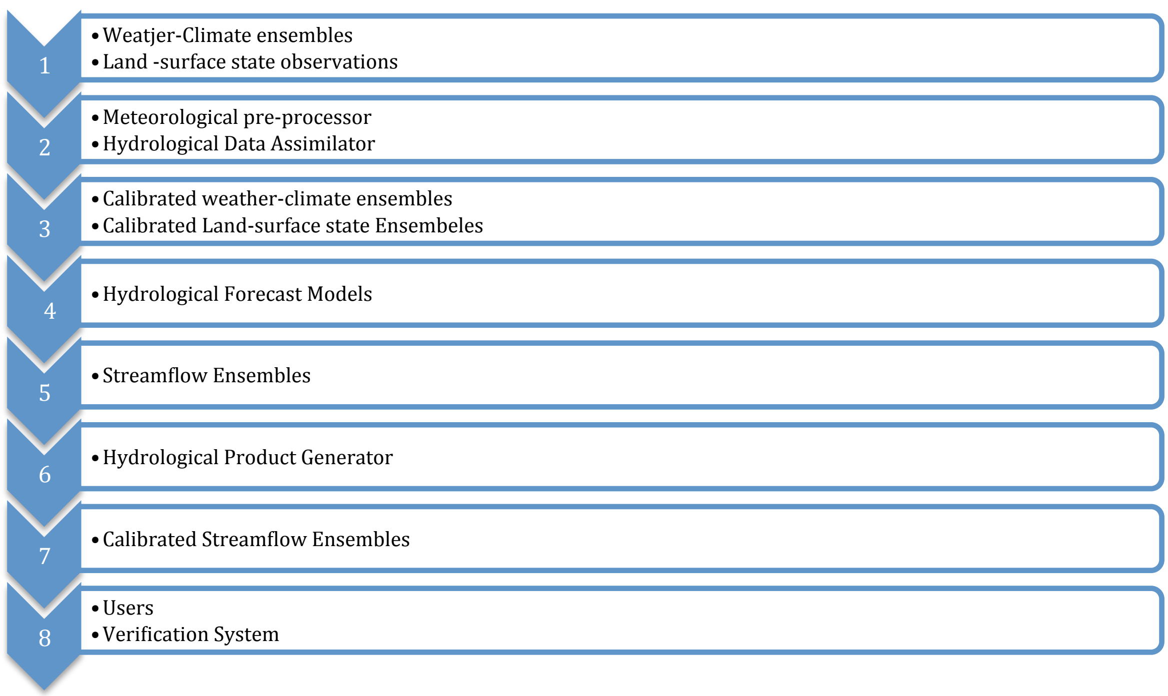


Figure 6 Schematic of the component sand information flow in an ensemble hydrologic prediction system [Schaake et al., 2007].

The process shown in Fig. 6 illustrates how predictions procedure in an ensemble hydrologic system.

The procedure for prediction of the experiment is to combine calibrated weather-climate ensembles and a meteorological pre-processor from weather-climate ensembles, such as climate indices. By using both components in hydrologic model, it structures the ensembles for streamflow. Another part is to calibrate land-surface state ensembles by a hydrological data assimilator. With the hydrological forecast models, the streamflow ensembles can be calculated in addition to the hydrological product generator and calibrated streamflow ensembles.

Forecasting uncertainty, extending predictability, better analysis and prediction elements are considered in HEPEX. Model resolution cannot be infinitely improved due to the limitation of computer resources. The major problem in HEPEX is how to link the model resolution to ensemble prediction systems. In order to create a high resolution ensemble forecast, HEPEX relies on a single mode deterministic forecast to keep the balance between low-resolution ensemble forecasting.

3.1.2 Hybrid Forecast

Hybrid forecast is the forecast skill that uses a physically based hydrologic model with the regression-based methods applied with combined initial conditions [Rosenberg, 2011]. Gridded observed precipitation and model-simulated snow water equivalent (SWE) are utilized in a hybrid framework as predictors in a regression model, transferred from an operational forecasting environment [Rosenberg, 2011; He et al, 2011]. In hybrid models, the forecast is based on the comparison of grid point simulated data as surrogates with ground-based observing stations data on observed counterparts [Rosenberg, 2011; He et al, 2011]. The hybrid model's forecasting has a better performance with a larger selection of grid points. Physically based hydrologic models provide more detailed information than the point observations alone [Li et al., 2009; Wood and Lettenmaier, 2006; Rosenberg, 2011; He et al, 2011].

In order to estimate the target period of streamflow for a specific study area, both DWR and NRCS use the snow water equivalent (SWE), runoff (RO), and accumulated precipitation (P) as three general components [Rosenberg, 2011; He et al., 2011]. The DWR and NRCS statistical forecasting models can be represented as:

$$Q = f(SWE, P, RO)$$

In hybrid models, unlike DWR's methodology, many predictor variables were missing for some time period, and the useable ones would be at the time of forecast. In the study, for example, the forecast on 1 April was thus for the entire April-July season, however, for 1 May was made just for May-July [Rosenberg, 2011; He et al, 2011]. VIC-simulated SWE, gridded precipitation forcing data and runoff predictors collected from California Data Exchange Center (CDEC) were used in the hybrid model. By expanding the watershed boundaries by either ¼^{o}, ½^{o}, or ¾^{o} in latitude or longitude, the catchment of watershed would cover more points, in case there exists unmonitored points of the whole basin (DWR Division of Flood Management, personal communication, 2008) [Rosenberg, 2011; He et al., 2011].

3.2 Principal Component Regression (PCR)

Principal component regression is the method that was built to select principal components that could estimates regression coefficients with low mean squared error [Gene et al., 2002]. The major component of principal component regression method is eigenvalues.

The regression model is of the form $y=X\beta+\varepsilon$, as y represents the m*1 vector of dependent variables, X represents the m*p matrix of independent variables, β is a p*1 vector of regression parameters, and ε is the m*1 vector of error terms [Jon et al., 1992; Tapani et al., 2010]. X can be described by three matrices, U, Σ, and V, such that $X=U\Sigma V^T$. U symbolizes an m*m orthogonal matrix, $\Sigma=\text{diag}\ (\sigma_1 \dots \sigma_n)$ is an m*p matrix, and V represents the p*p orthogonal matrix [Jon et al., 1992; Tapani et al., 2010].

The following equation shows how the basic relationship of y and X can be solved:

$$\hat{\beta} = \sum_{i \in M}^{i} \frac{1}{{\delta_i}^2} V_i {V_i}^T X^T y$$

The vector $X^T y$ belongs to the column space of X^T, and therefore is a linear combination of the vector V_i. As the covariance of $\hat{\beta}$ can be computed by:

$$cov(\hat{\beta}) = \delta^2 \sum_{i \in M}^{i} \frac{1}{{\delta_i}^2} V_i {V_i}^T$$

The equation implies that small singular values are not desired since they can increase the variance of $\hat{\beta}$[Gene et al., 2002].

By comparing with other regression models, principal components of the explanatory variables are used as repressors instead of other models regressing the dependent variables on the explanatory variables directly. Often, the principal components with higher variances (the ones based on eigenvectors corresponding to the higher eigenvalues of the sample variance-covariance matrix of the explanatory variables) are selected for regression. However, in predicting models, the principal components with low variances may also be important.

In addition, PCR can effectively result in dimension reduction through substantially lowering the effective number of parameters characterizing the underlying model. Usually regression is performed on a subset of all the principal components [Sutter et al., 1992]. In hydrologic prediction models, selecting major parameters is the most important, as there are a lot of uncertainties based on various climate phenomenon, making reduction of parameters very important. In previous work and until now, DWR used a PCR- based model to do streamflow prediction with VIC-simulated SWE and gridded precipitation dataset [Rosenberg, 2011].

According to the basic theory of PCR, the major work that should be done is to select all possible variables. By applying the variables to PCR, the regression will reorder all variables (the top variables perform almost all the prediction), which allows the most important variables to represent the observed data. Before all steps, the dataset should be normalized. Since the result shows a logarithmic curve, the normalized observation data should be exponentiated before doing the fitting.

3.3 Case Study

In the thesis, ground-based observation station data and climate indices collected by CDEC were used. Both datasets have a strong correlation with seasonal accumulated streamflow, as accumulated precipitation has a direct influence on streamflow change as it can be presented in the monthly pattern. According to previous work, streamflow in Northern California basins is influenced by snow water equivalent (SWE) [Rosenberg, 2011], while temperature is one key variable highly relevant to snow accumulation and melt.

Climate indices, such as El Nino Southern Oscillation (ENSO) indices, Southern Oscillation Index and ENSO precipitation index are frequently used to describe the variability and changes of climate system [Stenseth et al., 2003]. These climate indices may have an indirect connection to the available fresh water for streamflow response in Spring Season over the Northern California region [Huntington, 2006].

3.3.1 Study Area

In California, the most notable water supply systems are the State Water Project (SWP) operated by DWR, and the Central Valley Project (CVP) operated by the U.S. Bureau of Reclamation [Rosenberg, 2011].

North Fork Feather River is the watercourse of the northern Sierra Nevada in the U.S. state of California [Lindgren and Knowlton, 1991]. North Fork Feather River flows southward from its headwaters near Lassen Peak to Lake Oroville, the reservoir formed by Oroville Dam in the foothills of the Sierra where it runs into the Feather River. The river drains about 2,100 square miles (5,400 km^2) of the western slope of the Sierras [Wikipedia]. It is the largest tributary of the Feather River by discharge [Wikipedia].

For global hydrological and climatological data such as precipitation and temperature, the Water Resources Data System (WRDS) is a clearinghouse [Wilson et al., 2000; Brekke, 2009]. The Wyoming State Climate Office (SCO) is a branch of WRDS, and together they provide a variety of services ranging from the development of enhanced drought-monitoring products to the online dissemination of water resource publications. WRDS and the SCO are also helping to coordinate long-term monitoring efforts throughout the region. In both upstream and downstream, there are two dams, which have a huge effect on streamflow. For some stations along the north Fork, the record can be dated back to early 1900s, but there are some missing data for current years.

The monthly streamflow shows a random pattern for 33 years (1981-2013) in Fig. 7, with peak values indicating flood. In the study, the goal is to catch the peak and the pattern of seasonal trend. Fig. 8 illustrates Feather River North Fork Basin, both upstream and downstream. The station in the case study is located in the upstream area of the catchment, with the downstream data collected being unreliable, or the records were incomplete. Station data that was used as case study in this research is shown in Fig. 8.

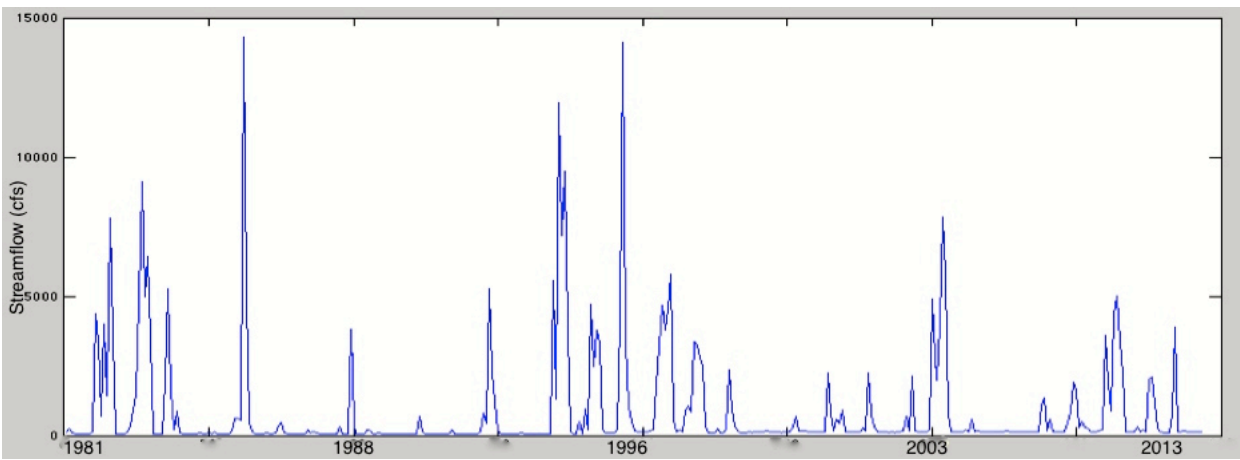


Figure 7: monthly streamflow for Feather River north fork the station number of 11404500 from USGS ground data base with the unit of (cfs)

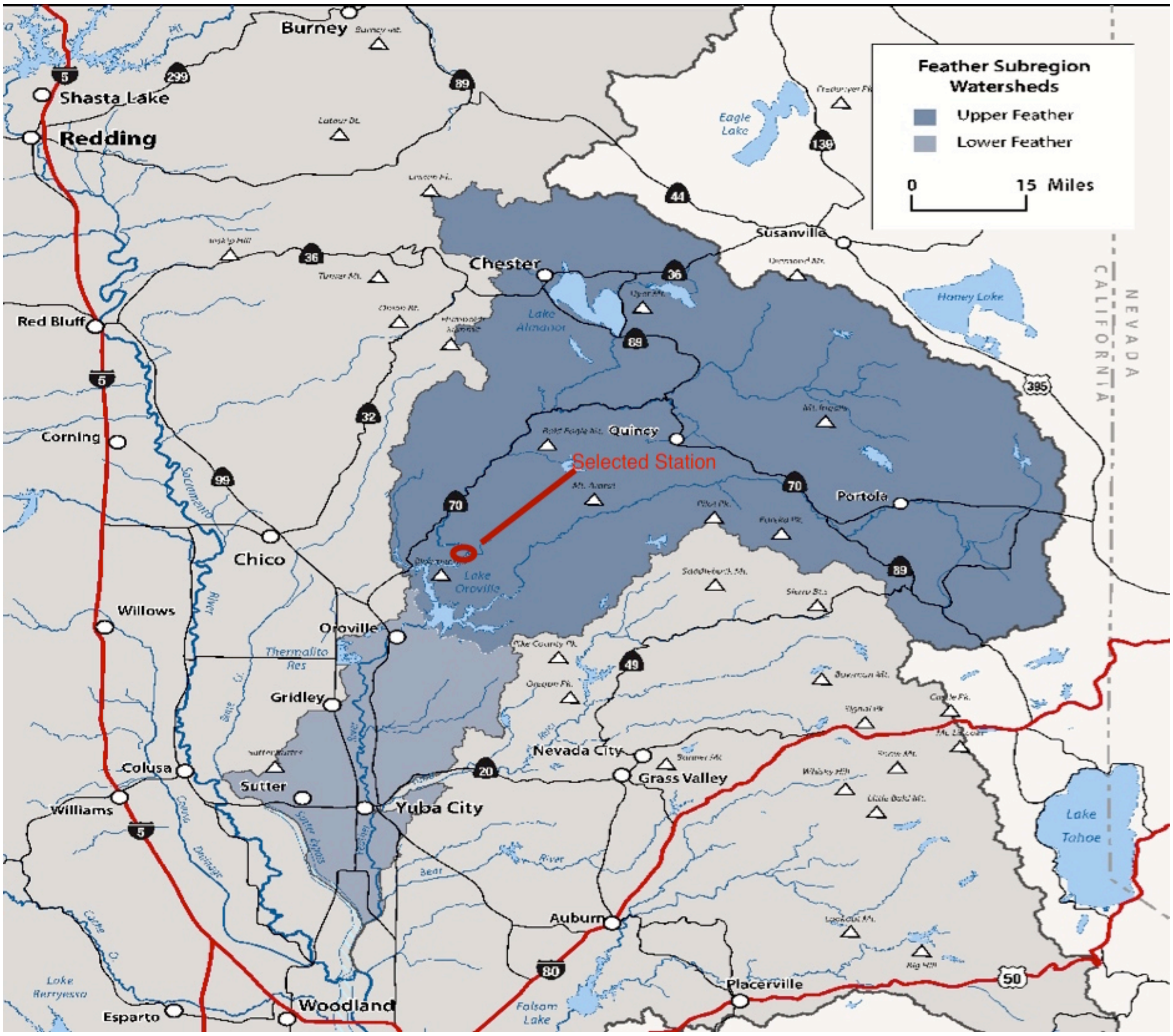


Figure 8 North Fork Feather River Basin, upstream and downstream [Sacramento watershed program]

3.3.2 Basin conditions

According to the integrated regional water management plan, Upper Feather Watershed supplies approximately 3.2 million acre-feet per year for downstream water users (urban, industrial, and agricultural). North Fork Feather River Basin provides over 60% drainage of the whole basin, with average daily flow into Oroville of 3,200 cfs [Seckler, 1971].

3.3.3 Data used in this study

To achieve PCR approach in the Feather River region and to combine the ground-based climate indices and statistical model, variables need to be selected during prediction [Rosenberg et al., 2011]. Gridded precipitation and temperature dataset from PRISM climate group were selected. Climate indices as additional variables were selected in simulation, including Southern Oscillation Index (SOI), Pattern of Northern America Index (PNA), and Pacific Decadal Oscillation (PDO). The target simulation is spring season streamflow, with monthly precipitation of the months of December, January and February along with the accumulated precipitation of December through February used as variables. [Jeong and Kim, 2005; Nijssen et al., 2001]. In addition to monthly data for December, January, and February, average temperature for winter season (December, January, and February) was also used. The climate indices, along with the monthly average data for December, January and February are utilized as variables [Kennedy et al., 2009]. Variables from 1981-2005 are used to do model calibration and for validation, with variables used for the simulation from 2006-2013.

From principal component analysis (PCA), principal components are transferred variables. As physical variables, precipitation and temperature are not transferred variables. Independent variables in regression model can be defined as:

$X=[X_{1,} X_{2,} X_{3 \ldots\ldots} X_{11}]$

Where X_1 is accumulated precipitation for winter season (December, January and February (DJF) (mm), X_2 is average temperature for winter season (DJF) (C^o). X_3 is monthly precipitation for December, X_4 is monthly precipitation for January, X_5 is monthly precipitation for February, and X_6 is monthly average temperature for December, X_7 is monthly temperature for January, X_8 is monthly average temperature for February. As for climate indices, X_9 is average value of SOI (DJF), X_{10} is average value of PNA (DJF), and X_{11} equals to average value of PDO (DJF). The above variables constitute the independent variable in regression model. While dependent, which is also predicted, variable, is Y=average streamflow for March, April, and June (cfs). Basic variables in regression model can be defined as:

$X_b=[X_1,X_2, X_3, X_6]$

The variables include accumulated precipitation for DJF, average temperature for DJF, and monthly precipitation and temperature for December. To compare different contribution of other climate indices in PCR, it can be achieved by adding it on to basic variables.

$X_a=[X_b, X_i, X_j, X_{k\ \ldots\ \text{etc.}}]$

3.3.4 Evaluation Statistics

variables	Notation	Correlation	Bias	RMSE	NSE
Basic Variables	$[X_b]$	0.6875	1.6023e-16	0.5138	0.4726
Basic Variables with SOI	$[X_b$, SOI]	0.7105	1.1824e-16	0.4979	0.5048
Basic Variables with PNA	$[X_b$, PNA]	0.7009	2.2358e-17	0.5047	0.4912
Basic Varibales with PDO	$[X_b$, PDO]	0.6880	1.9377e-16	0.5135	0.4734
Basic Variables with PNA and PDO	$[X_b$, PNA, PDO]	0.7134	-4.099e-17	0.4958	0.5090
Basic Variables with PNA and SOI	$[X_b$, PNA, SOI]	0.7382	1.4533e-16	0.4773	0.5449
Basic Variables with SOI and PDO	$[X_b$, SOI, PDO]	0.7106	1.0706e-16	0.4978	0.5050
Basic Variables with SOI and PNA and PDO	$[X_b$, Soi, PNA, PDO]	0.7474	2.5339e-16	0.4701	0.5586
All Variables	$[x_1, x_2, \ldots, x_{11}]$	0.8171	1.6396e-16	0.4079	0.6677

Table 1 different combination of climate indices used in the model, the correlation, bias, root-mean-square error, and Nash-Sutcliffe error

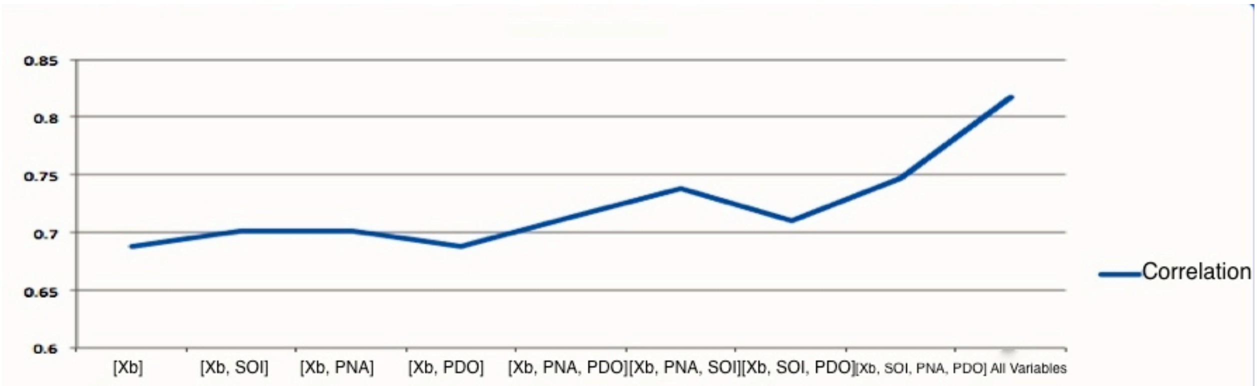


Figure 9 Correlation curve of using different climate indices

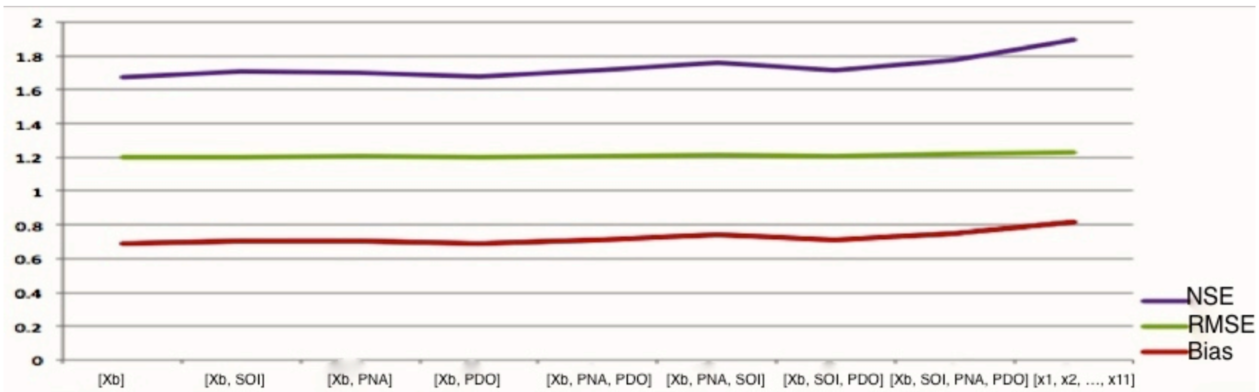


Figure 10 NSE RMSE Bias curves of different indices used in the model

Table 1 summarizes the correlation between the simulated and observed variable, showing the linear relationship between the two variables. Based on Table 1, Fig. 9 gives us a visualized image of how climate indices in the model correlate. With more variables in the model, the result is much more accurate. In this model, there are six major variables used, with the SOI & PNA showing the best result. Fig. 10 visualizes bias target variable and observed variable values, and the root-mean-square-error (RMSE) shows the difference between the estimator and what is estimated. The RMSE is a good measure of accuracy, with a smaller RMSE the veracity of the model. From Table 1, by adding climate indices to physical variable the RMSE value decreases, which shows with more variables the target variable can be better predicted. For Nash-Sutcliffe error, the closer the model efficiency is to 1, the more accurate the model is. The model with all variables showed the highest correlation coefficient, the lowest RMSE and the highest NSE.

4.Result

4.1 Prediction Analysis

Based on Principal Component Regression model that applies multiple climate variables, there are two basic components that have more influence in the prediction, which are accumulated precipitation and mean temperature for the three months before the spring [Rosenberg et al., 2011]. Feather River Basin's major source of streamflow is rainfall and snow [Kim, 2005]. For spring season streamflow prediction, streamflow is mainly from snow-melt water from winter snow. In Northern California, precipitation in winter months can be in solid or liquid forms depending on surface temperature [Asztalos et al., 2006]. Compared with monthly forecasting, seasonal forecasting reduced uncertainty as seasonal streamflow is much more consistent. For example, in a colder year, snow melt water will be released later to the watershed than a warmer year. The seasonal storage could be similar on average, however monthly storage will have a huge difference due to this [Unterfrauner, 2009]. After selecting all possible variables to simulate streamflow, by calculating the eigenvalue, the contribution from each variable can be found. With the reordered variables, Fig. 11 shows the top six variables after transfer, which accounts for 95% variance explained.

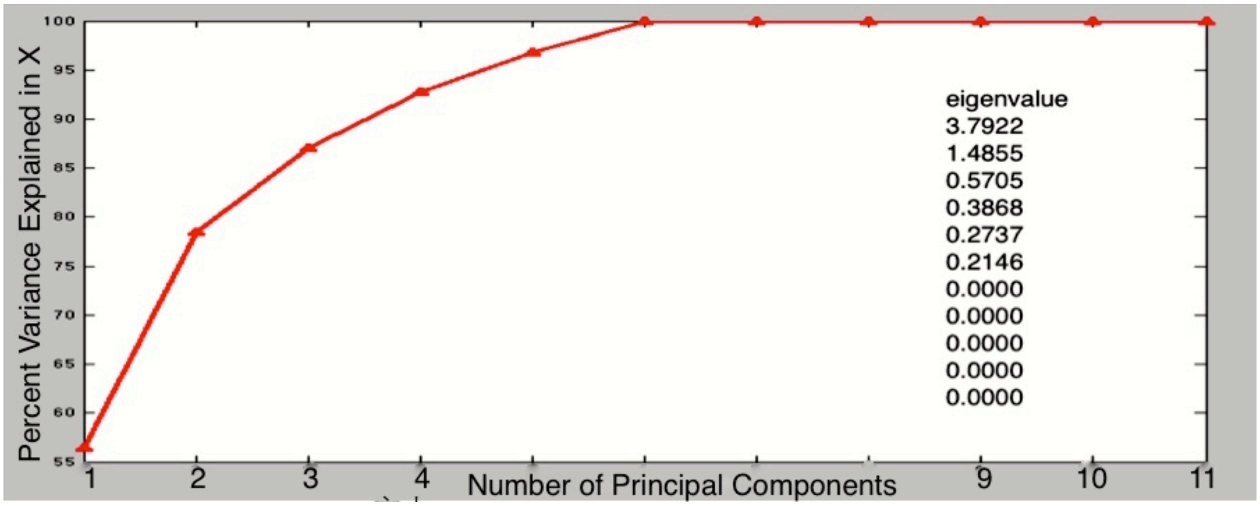


Figure 11 Number of principal components used in regression model and the percent variance explained in the hold variables

As accumulated precipitation in winter season has a high correlation with spring streamflow, it can be assumed that accumulated precipitation is the major component in the regression model. However, winter season temperature has an important impact on streamflow [Kelly and Roy, 1991]. The basic two variables before simulation can be set as precipitation and temperature to test monthly data and climate based indices contributed in regression, and the result is in Fig. 12. The curve in Fig. 9 illustrates the number of principal components used in the regression model, and the eigenvalue of each component is the contribution of reordered variables. With only one variable, the percent variance that explains X is around 57%. By adding on variables, the independent variable X can be well explained.

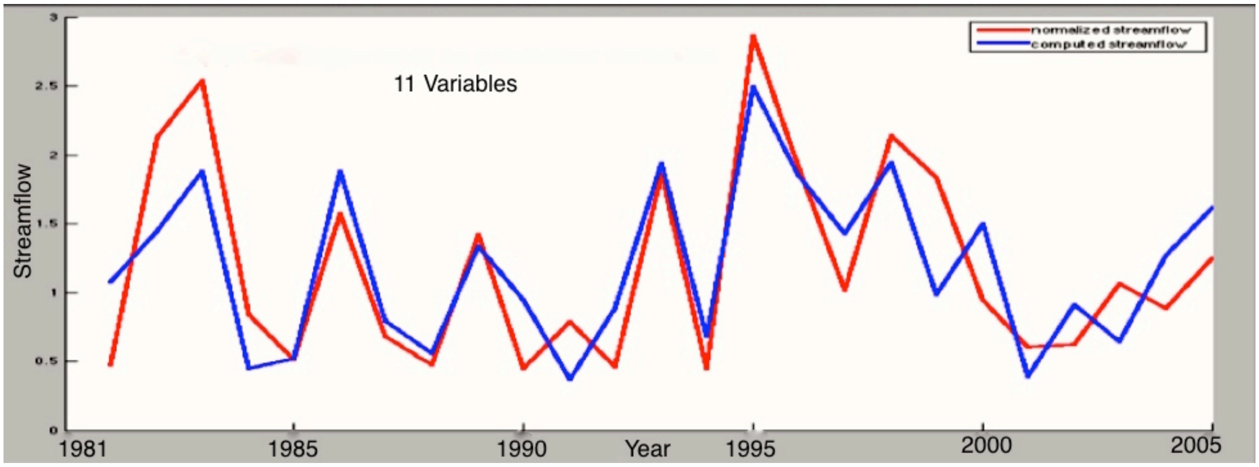


Figure 12 Standardized observed streamflow and simulated streamflow with 11 variables using the Principal Component for 25 years (1982-2005)

PCR is used to reorder the correlated variables by calculating the contribution score of each variable, and the eigenvalues show the number of components that are necessary to represent the observed data [Helland, 1990]. For example, by using the first three components, it can account for almost 95% of simulation. In this case, according to eigenvalues, the observed streamflow can be represented by six main components instead of using all eleven components. In PCR, it only shows the reordered eigenvalue. Given this information, variables need to be manually reduced. The basic components for regression are accumulated precipitation for the winter season, which is December, January, and February, mean temperature for the same three months and monthly precipitation and temperature in December. The result is shown in Fig. 10. In December, precipitation is the major source of storage, while temperature tells us if precipitation will produce runoff before the spring season or remain in storage as snow for spring water usage. Higher temperature during December means most of the precipitation turns into runoff instead of keeping the shape of snow for spring usage.

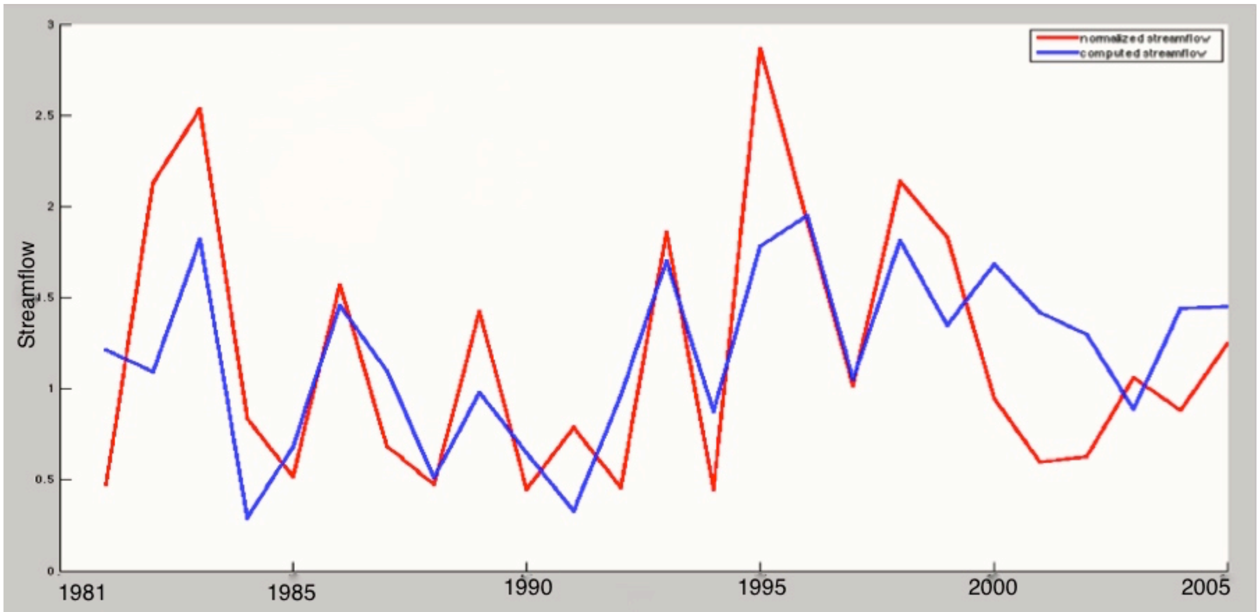
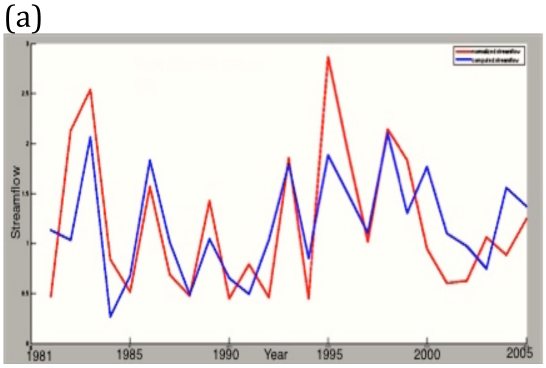


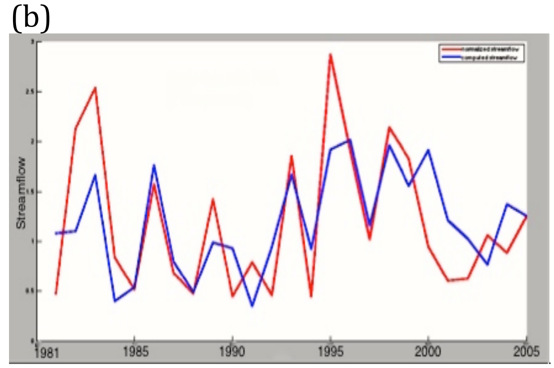
Figure 13 Standardized observer streamflow and simulated streamflow with 4 components using PCR by using accumulated precipitation of Dec, Jan, and Feb, and monthly precipitation of Dec, and mean temperature, and monthly temperature for Dec for 25 years (1981-2005)

By comparing different combinations of basic variables and climate indices, the results are slightly different, according to the table of correlation coefficient between simulated streamflow and observed streamflow. *Table 1* shows the different contributions of each climate index. For different regions, climate indices may have different influence on prediction. According to different combinations of variables, variables after regression can generally follow the pattern of the observed streamflow dataset (after normalization).

(a)

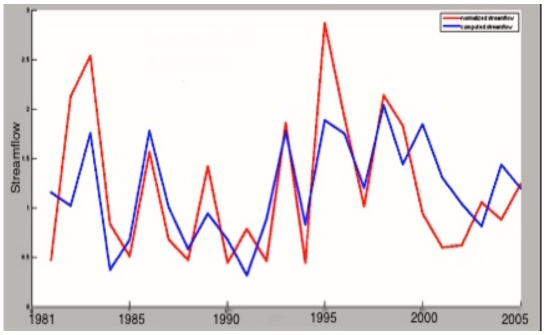


(b)

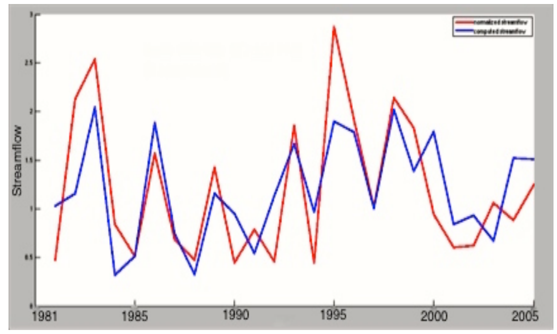


(c)

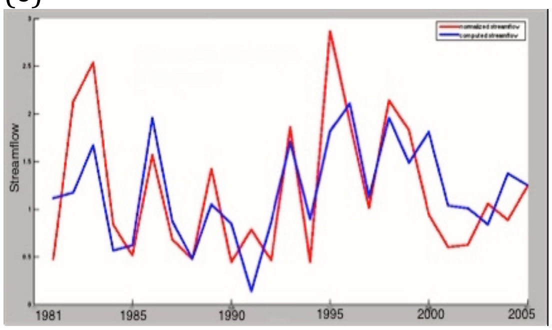
(d)

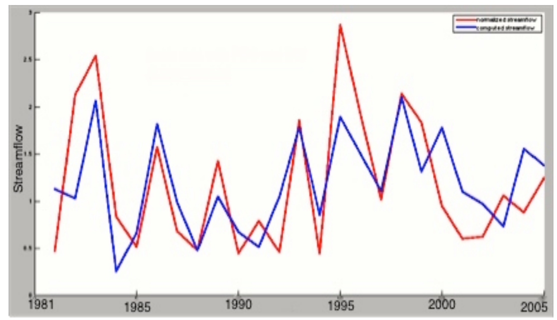


(e)



(f)





(g)

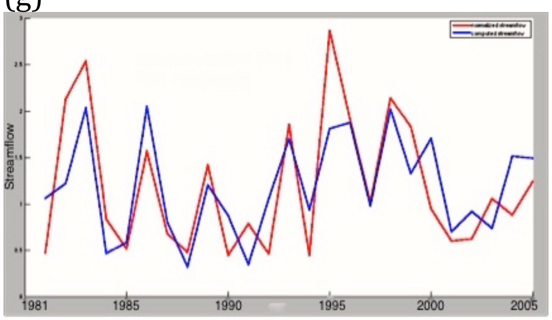


Figure 14 a) basic variables with SOI b) basic variables with PNA c) basic variables with PDO d) Basic variables with SOI& PNA e) Basic variables with PNA&PDO f) basic variables with SOI&PDO g) basic variables with SOI&PNA&PDO (cfs)

By comparing all seven plots in Fig. 15, the slight difference between combinations of variables demonstrates different information. However, the pattern of simulated results are not the same. For example, Fig. 15(d) and Fig. 15(e), as basic variables works with SOI and PNA, and the peak is not the same as basic variables with PNA and PDO. To have a better understanding of the difference between the results, the correlation coefficient expresses how close the simulation is to observation.

With the basic concept of PCR for streamflow prediction, the first thing to consider is which of the components should be used for prediction. The identified components are the accumulated precipitation and the average temperature for three months before the prediction season.

By comparing numbers shown in Table 1, the correlation between simulated streamflow and observed streamflow shows that the simulated streamflow changes in the same pattern as the observed data changes. As we have the NSE between 0 and 1, the more variables used, the better predictor the model will be. However, the model shows that with six variables it can cover almost all information for prediction.

Based on the result above, the best components are basic precipitation and temperature with SOI and PNA. While PDO does contribute, PDO does not provide a better indicator than the other two variables. Fig. 15 shows the evaluation of the PCR model with six components.

4.2 Validation

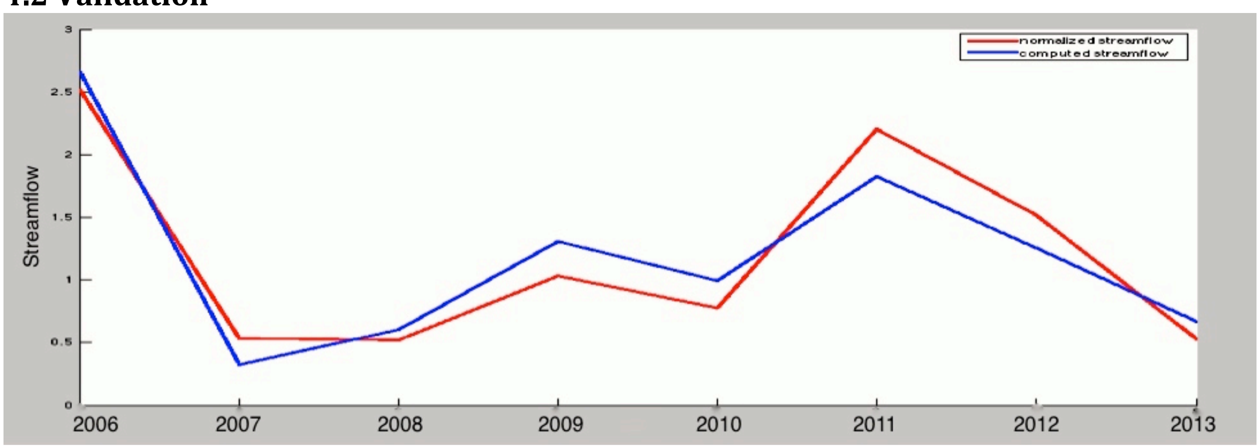


Figure 15 Model Validation

To validate the model when using eleven variables, the result shows the correlation coefficient between simulated data and observation is 0.94 during the specific time period of 2006-2013. However, it is not necessarily the model that should be used for streamflow forecasting. DWR and NOAA [Rosenberg et al., 2011] have been using HEPEX and hybrid model for streamflow forecasting using simulated gridded dataset, and the predictor that has been often utilized has been snow water equivalent (SWE). In this study, as temperature has a direct coherence with snowmelt water accumulation, it is a major variable in regression model. In the case study, it is demonstrated that the proposed approach can work well for forecasting streamflow over spring season for the test basin (the North Fork of the Feather River).

5. Summary and Conclusions

The result shown in Fig. 14 demonstrated information from climate variables in winter season are correlated to spring season streamflow forecasting. Physical variables (precipitation and temperature) correlated to streamflow changes in spring season, with precipitation types (snow or rain) and average temperature affecting when water is released to streamflow. By adding climate indices to streamflow prediction regression model, the result is better than just predicting with physical variables. With more variables in the regression model, the result is better. By using principal components with the largest variances, PCR effectively reduces the dimensionality of streamflow estimation [Ozkale, 2009]. Following the regression progress, the principal components are selected after they are reordered and scored. With the eigenvalue, each variance of principal components is explained.

Other basins would require different forecasting information, due to climate conditions in the individual geographic region. Information regarding climate indices would contribute to the forecasting progress, and increase the accuracy of forecasts. As variables are selected around the Feather River Basin, the model and selected variables can be utilized regionally. To test if the approach can be utilized throughout California or for all U.S. basins, more station information is needed. Additional testing of gauge conditions can contribute to the regression model, and can be enhanced and potentially utilized in a wider basin region.

In conclusion, it was found that adding on climate indices to physical variables like precipitation and temperature indicated a higher correlation coefficient between the simulated and observed data.

6.References

Alan F. Hamlet, Dennis P. Lettenmaie (1999), Columbia River streamflow Forecasting Based on ENSO and PDO Climate Signals, Journal of Water Resources Planning and Management/ November/ December

Asztalos, J., Kirnbauer, R., Escher-Vetter, H., & Braun, L. (2006, October). A distributed energy balance snowmelt model as a component of a flood forecasting system for the Inn river. In Proceedings of the Alpine Snow Workshop, Munich.

Barnston, A. G., and R.E., Livezey, 1987; Classification, seasonality, and persistence of low-frequency atmospheric circulatio patterns. Mon, Wea, Recv., 115, 1083-1126

Brekke, L. D. (2009). Climate change and water resources management: A federal perspective. DIANE Publishing.

Climate Prediction Center - ENSO FAQ. (n.d.). Retrieved March 17, 2015, from http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensofaq.shtml#HOWLONG

Climatology. (n.d.). Retrieved March 17, 2015, from http://en.wikipedia.org/wiki/Climatology

(n.d.). Retrieved March 17, 2015, from http://worldwidescience.org/topicpages/e/ensemble-

based probabilistic forecasting.html

Chowdhury, M. R. (2003). The El Nino-Southern Oscillation (ENSO) and seasonal flooding–Bangladesh. Theoretical and applied climatology, 76(1-2), 105-124.

David C. Garen, Danny Marks (2005), Spatially Distributed Energy balance snowmelt modelling in a mountainous river basin: estimation of meteorological inputs and verification of model results, Journal of Hydrology doi:10.1016/j.jhydrol.2005.03.026

El Niño Southern Oscillation. (n.d.). Retrieved March 15, 2015, from http://en.wikipedia.org/w/index.php?title=El_Niño_Southern_Oscillation&redirect=no

Eric A. Rosenberg (2011), Statistical applications of physically based hydrologic models to seasonal streamflow forecasts, Water Resources Research, VOL.47, WOOH14, doi: 10.1029/2010WR010101, 2011

Franz, K. J., H. C. Hartmann, S. Sorooshian, and R. Bales (2003), Verification of National Weather Service ensemble streamflow predictions for water supply forecasting in the Colorado River basin, J. Hydrometeorol., 4, 1105–1118.

Feather River. (n.d.). Retrieved March 17, 2015, from http://en.wikipedia.org/wiki/Feather_River

From the Water Resources Research Editor-in-Chief. (n.d.). Retrieved March 17, 2015, from http://www.academia.edu/2825843/From_the_Water_Resources_Research_Editor-in-Chief

Glantz, M. H. (1982), Consequences and Responsibilities in Drought Forecasting: The case of Yakima, 1977, Water Resource. Rex. 18(1), 3-13, doi:10.1029/WR018i001p00003

Global Patterns - Pacific/North American (PNA) | State Climate Office of North Carolina. (n.d.). Retrieved March 15, 2015, from http://www.nc-climate.ncsu.edu/climate/patterns/PNA.html

Haggag, M., Yamashita, T., Kim, K. O., & Lee, H. S. (2010). Simulation of the North Indian Ocean Tropical Cyclones Using the Regional Environment Simulator: Application to Cyclone Nargis in 2008. In Indian Ocean Tropical Cyclones and Climate Change (pp. 73-82). Springer Netherlands.

Hart, D., and F. Gehrke (1990), Status of the California Cooperative SnowSurveys Program, *58th Western Snow Conference, April 17–19, Sacramento, California.*

Helland, I. S. (1990). Partial least squares regression and statistical models.Scandinavian Journal of Statistics, 97-114.

He, Y., Bárdossy, A., & Zehe, E. (n.d.). A review of regionalisation for continuous streamflow simulation. Hydrology and Earth System Sciences, 3539-3553.

Huntington, T. G. (2006). Evidence for intensification of the global water cycle: review and synthesis. Journal of Hydrology, 319(1), 83-95.

Ilin, A., & Raiko, T. (2010). Practical approaches to principal component analysis in the presence of missing values. Journal of Machine Learning Research 11.

Integrated Climate Data Center - ICDC. (n.d.). Retrieved March 19, 2015, from http://icdc.zmaw.de/climate_indices.html?&L=1#

Jeong, D. I., & Kim, Y. O. (2005). Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. Hydrological processes, 19(19), 3819-3835.

Jon M, John H., Patrick M. (1992), Which Principal Components to Utilize for Principal Component Regression, 1992 by John Wiley& Sons,Ltd.

John C., Schaake (2007), HEPEX the Hydrological Ensemble Prediction Experiment, American Meteorological Society

J.T.Gene Hwang, Dan Nettleton (2002), Principal Components Regression with Data- Chosen Components and Related Methods

Kelly T., Roy W. (1991), Surface Climate and Streamflow Variability in the Western United States and Their Relationship to Large-Scale Circulation Indices, Water Resources Research VOL. 27, No.9, Pages 2381-2399, September 1999

Kennedy, A. M., Garen, D. C., & Koch, R. W. (2009). The association between climate teleconnection indices and Upper Klamath seasonal streamflow: Trans-Niño Index. Hydrological processes, 23(7), 973-984.

Kim, J. (2005). A projection of the effects of the climate change induced by increased CO2 on extreme hydrologic events in the western US. Climatic Change, 68(1-2), 153-168.

Können, G. P., Jones, P. D., Kaltofen, M. H., & Allan, R. J. (1998). Pre-1866 extensions of the Southern Oscillation Index using early Indonesian and Tahitian meteorological readings. Journal of Climate, 11(9), 2325-2339.

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

Li, H., L. Luo, E. F. Wood, and J. Schaake (2009), The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, J. Geophy. Res., 114, D04114.

Lindgren, W., & Knowlton, F. H. (1911). The tertiary gravels of the Sierra Nevada of California (No. 73). Govt. print. off..

Iris T., Daniel R. (2004), Changes toward Earlier Streamflow Timing across Western North America, 2005 American Meteorological Society

Milija, Zupanski(1992),Regional four-dimensional variational data assimilation in a Quasi-Operational Forecasting Environment, UCAR Visiting Research Program, National Meteorological Center, Washington,D.C, Monthly weather review

Mantua, N. J., & Hare, S. R. (2002). The Pacific decadal oscillation. Journal of Oceanography, 58(1), 35-44.

Nijssen, B., O'Donnell, G. M., Hamlet, A. F., & Lettenmaier, D. P. (2001). Hydrologic sensitivity of global rivers to climate change. Climatic change, 50(1-2), 143-175.

North_American_teleconnection_pattern http://en.wikipedia.org/wiki/Pacific%E2%80%93North_American_teleconnection_pattern

North Fork Feather River. (n.d.). Retrieved March 17, 2015, from http://en.wikipedia.org/wiki/North_Fork_Feather_River

Pacific decadal oscillation. (n.d.). Retrieved March 15, 2015, from http://en.wikipedia.org/wiki/Pacific_decadal_oscillation

Pacific–North American teleconnection pattern. (n.d.). Retrieved March 15 http://en.wikipedia.org/wiki/Pacific–North_American_teleconnection_pattern

Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H., Buizza, R., & Roo, A. (2008, January 1). New dimension in early flood warning across the globe using grand-ensemble weather *predictions. Geophysical Research Letters.*

Putcger, E. J., M. L., Blackmon, G. T. Bates and S. Munoz, 1988: The effect of norht Pacific sea surface temperature anomalies on the January climate of a general circulation model, J. Atoms. Sci., 45, 173-188

Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., ... & Kirtman, B. (2007). Climate models and their evaluation. In Climate Change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR) (pp. 589-662). Cambridge University Press.

Raiko, T., Ilin, A., & Karhunen, J. (n.d.). Principal Component Analysis for Large Scale Problems with Lots of Missing Values.

Reynolds, R., Dettinger, M., Cayan, D., Stephens, D., Highland, L., & Wilson, R. (1997). Effects of El Niño on streamflow, lake level, and landslide potential. US Geological Survey Open file report.

Roger C., Noah P. (2006), Mountain Hydrology of the western United States, Water Resources Research, VOL. 42, W08432, doi: 10.1029/2005WR004387, 2006

Rodionov, S., & Assel, R. (n.d.). A new look at the Pacific/North American Index. Geophysical Research Letters, 1519-1519.

Sacramento River Watershed Program. (n.d.). Retrieved March 17, 2015, from http://www.sacriver.org/

Satish K., Blaji R. (2003), Seasonal Cycle Shifts in Hydroclimatology over the Western United States, Journal of Climate, VOL. 18

Seckler, D. W. (Ed.). (1971). California water: A study in resource management. Univ of California Press.

Scott C, Rober A (1999), ENSO indices Based on Patterns of Satellite-Derived Precipitation, 2000 American Meteorological Society

Southern Oscillation Index (SOI). (n.d.). Retrieved March 15, 2015, from http://www.ncdc.noaa.gov/teleconnections/enso/indicators/soi/

Stenseth, N. C., Ottersen, G., Hurrell, J. W., Mysterud, A., Lima, M., Chan, K. S., ... & Ådlandsvik, B.

(2003). Studying climate effects on ecology through the use of climate indices: the North Atlantic Oscillation, El Nino Southern Oscillation and beyond. Proceedings of the Royal Society of London B: Biological Sciences, 270(1529), 2087-2096.

Sutter, J. M., Kalivas, J. H., & Lang, P. M. (1992). Which principal components to utilize for principal component regression. Journal of chemometrics, 6(4), 217-225.

Syed, F. S., Giorgi, F., Pal, J. S., & King, M. P. (2006). Effect of remote forcings on the winter precipitation of central southwest Asia part 1: observations.Theoretical and applied climatology, 86(1-4), 147-160.

Thomas C. Pagano, Davi C. (2009), Daily Updating of operational Statistical Seasonal Water Supply Forecasts for the Western U.S., Jpunal of the American Water Resources Association, VOL.45, No.3

Timothy B. (2010), Understanding El Nino for your Fire Management Unit, Desert Research Institute, University of Idaho

Unterfrauner, F. (2009). Investigation of water storage and flow processes in snow, firn and ice (Doctoral dissertation, uniwien).

Wallace, J.M, and D.S. Guzler, Teleconnections in the geopotential height field during the northern hemisphere winter, Mon. Weather Rev. 109,784-812, 1981

Wei, W. (2010). Use of hydroclimatic forecasts for improved water management in central Texas. Michigan Technological University.

WilliamJ. Sydeman et al. (2013) Increasing variance in North Pacific climate related to unprecedented ecosystem variability off California, Global Change Biology (2013)19, 1662-1675,doi:10.1111/bcb.12165*

Wilson, J. P., Mitasova, H., & Wright, D. J. (2000). Water resource applications of geographic information systems. Urisa Journal, 12(2), 61-79.

Wood, A. W., and D. P. Lettenmaier (2006), A test bed for new seasonalhydrologic forecasting approaches in the western United States, Bull.Am. Meteorol. Soc., 87(12), 1699–1712.

Yarnal, B., and H. F. Diaz, Relationships between extremes of the Southern Oscillations and *the winter climate of the Anglo- American Pacific coast, J. Clim.,6,197-219,1986.*

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble-based weather forecasts. Bull. Amer. Meteor. Soc., 83, 73–83.

(n.d.). Retrieved March 15, 2015, from http://prism.oregonstate.edu/documents/PRISM_datasets.pdf

(n.d.). Retrieved March 17, 2015, from http://www.learner.org/courses/envsci/unit/pdfs/textbook.pdf

(n.d.). Retrieved March 18, 2015, from http://www.feather-river-crm.org/pdf/MOU/IRWMP_063005.pdf

Özkale, M. R. (2009). Principal components regression estimator and a test for the restrictions. Statistics, 43(6), 541-551.