6-22-2012

Estimating annual precipitation for the Colorado River Basin using oceanic-atmospheric oscillations

Ajay Kalra
University of Nevada, Las Vegas, Ajay.Kalra@dri.edu

Sajjad Ahmad
University of Nevada, Las Vegas, sajjad.ahmad@unlv.edu

Follow this and additional works at: https://digitalscholarship.unlv.edu/fac_articles

Part of the Atmospheric Sciences Commons, Climate Commons, Environmental Engineering Commons, Meteorology Commons, and the Water Resource Management Commons

Estimating annual precipitation for the Colorado River Basin using oceanic-atmospheric oscillations

Ajay Kalra[1,2] and Sajjad Ahmad[1]

[1] Estimating long-lead time precipitation under the stress of increased climatic variability is a challenging task in the field of hydrology. A modified Support Vector Machine (SVM) based framework is proposed to estimate annual precipitation using oceanic-atmospheric oscillations. Oceanic-atmospheric oscillations, consisting of Pacific Decadal Oscillation (PDO), North Atlantic Oscillation (NAO), Atlantic Multidecadal Oscillation (AMO), and El Niño–Southern Oscillation (ENSO) for a period of 1900–2008, are used to generate annual precipitation estimates with a 1 year lead time. The SVM model is applied to 17 climate divisions encompassing the Colorado River Basin in the western United States. The overall results revealed that the annual precipitation in the Colorado River Basin is significantly influenced by oceanic-atmospheric oscillations. The long-term precipitation predictions for the Upper Colorado River Basin can be successfully obtained using a combination of PDO, NAO, and AMO indices, whereas coupling AMO and ENSO results in improved precipitation predictions for the Lower Colorado River Basin. The results also show that the SVM model provides better precipitation estimates compared to the Artificial Neural Network and Multivariate Linear Regression models. The annual precipitation estimates obtained using the modified SVM modeling framework may assist water managers in statistically understanding the hydrologic response in relation to large scale climate patterns within the Colorado River Basin.

1. Introduction

1.1. Background

[2] Climatic fluctuations and increasing water demand in growing regions have captured the attention of scientific communities. This has led to the need for estimating regional and global precipitation of interannual and longer time scales [*Karl and Knight*, 1997; *Hidalgo and Dracup*, 2003]. Although precipitation is predominantly episodic, variability in precipitation results in such catastrophic events as floods and drought. The impact of these catastrophic events on water resources planning [*Qaiser et al.*, 2011; *Dawadi and Ahmad*, 2012]; water management [*Ahmad and Simonovic*, 2001, 2006], water infrastructure [*Forsee and Ahmad*, 2011; *Ahmad and Simonovic*, 2000, 2004], human settlements [*Mosquera-Machado and Ahmad*, 2007; *Simonovic and Ahmad*, 2005], and the environment [*Venkatesan et al.*, 2011a, 2011b; *Shrestha et al.*, 2011,

2012] have been studied by many researchers. Considering the impacts of floods and droughts on humans and environment and growing water demand in many parts of the world, efforts have increased to study and predict regional and global precipitation variation. Interannual, decadal, and multidecadal climatic signals, such as oceanic-atmospheric oscillations provide an exciting opportunity to estimate precipitation at longer lead-times. Oceanic-atmospheric oscillations, often termed as teleconnections, have been used by researchers across the world to study their relationship with precipitation over land surfaces at interannual and longer time scales; this relationship, in turn, controls such key components as streamflow, soil moisture, and evaporation of the hydrological cycle [*Diaz and Kiladis*, 1992; *Dracup and Kahya*, 1994; *Rajagopalan and Lall*, 1998; *Diaz et al.*, 2001; *Gutzler et al.*, 2002; *Viles and Goudie*, 2003; *Hidalgo and Dracup*, 2003; *Xu et al.*, 2004; *Hu and Feng*, 2001; *Stephen et al.*, 2010; *Puri et al.*, 2011]. The main objective of this study is to estimate annual precipitation using oceanic-atmospheric oscillations.

1.2. Oceanic Variability and Precipitation

[3] It is well established that the year-to-year variability of precipitation is primarily associated with the sea surface temperature (SST) anomalies in the tropical Pacific Ocean, known as the El Niño–Southern Oscillation (ENSO) [*Ropelewski and Halpert*, 1986; *Redmond and Koch*, 1991; *Piechota and Dracup*, 1996; *Kane*, 1999; *Barlow et al.*, 2002]. *Wang et al.* [2000] observed positive anomalies of precipitation in the

---

[1]Department of Civil and Environmental Engineering, University of Nevada, Las Vegas, Nevada, USA.

[2]Now at Division of Hydrologic Sciences, Desert Research Institute, Las Vegas, Nevada, USA.

Corresponding author: S. Ahmad, Department of Civil and Environmental Engineering, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, Las Vegas, NV 89154-4015, USA. (sajjad.ahmad@unlv.edu)

central Pacific and eastern Asia during extreme phases of ENSO cycles. *Lau and Wu* [2001] found that ENSO accounts for 30% of climate-related variability in Asian summer monsoon rainfall. *Hu and Feng* [2001] revealed that warmer phases of ENSO are related to increased summer precipitation within the central United States. A pseudo El Niño phenomenon, known as the El Niño Modoki, has been significantly linked to temperature and precipitation over many parts of the world, including Japan, Korea, New Zealand, and even the West Coast of the United States [*Ashok et al.*, 2007; *Weng et al.*, 2007].

[4] In the southwestern United States, particularly in the Colorado River Basin (CRB), various types of teleconnections are found to be related with the hydrologic variability [*Pulwarty and Melis*, 2001]. *Piechota and Dracup* [1996] used Palmer Drought Severity Index (PDSI) to relate the major dry and wet spells with historic ENSO events in the Lower Colorado River Basin (LCRB). *Kahya and Dracup* [1993] described 1941 and 1983 heavy rainfall events with ENSO phases in the southwestern U.S., including the CRB. *Cayan et al.* [1998] showed that there has been a change in the pattern and amount of precipitation within the CRB. *Merideth* [2000] observed wet periods during the start and end of 20th century, whereas dry periods during the mid-century in the Colorado Basin. Moreover, the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) routinely monitors the correlation between ENSO and southwest U.S. precipitation [*California Dep. of Water Resources*, 2005].

[5] While many studies have related ENSO with precipitation, there are other modes of atmospheric oscillations, for instance, Pacific Decadal Oscillation (PDO), Interdecadal Pacific Oscillation (IPO), North Atlantic Oscillation (NAO), Atlantic Multidecadal Oscillation (AMO), Arctic Oscillation (AO), Indian Ocean Dipole (IOD), and Pacific North American (PNA) indices. These also are important indicators of climate variability, and have been linked with precipitation both individually and in conjunction with ENSO [*Bjerknes*, 1966; *Hurrell*, 1995; *Giannini et al.*, 2001; *Gershunov and Barnett*, 1998; *Higgins et al.*, 2000; *Brito-Castillo et al.*, 2002; *McCabe et al.*, 2004; *Wang and Swail*, 2001; *Dickson et al.*, 2000; *Qian et al.*, 2000; *Ashok et al.*, 2001, 2003; *Wedgbrow et al.*, 2002; *Kiem and Franks*, 2004; *Pui et al.*, 2011]. Compared to ENSO and PDO, the AMO and NAO indexes exhibit considerable long-term variability [*Cancelliere et al.*, 2007; *McCabe et al.*, 2007].

1.3. Artificial Intelligence Models

[6] From the past studies, it is evident that oceanic-atmospheric oscillations do influence precipitation. In fact, there have been attempts to use oscillations as predictors to estimate precipitation. However, due to the complex interaction between precipitation and oceanic oscillation, it is difficult to construct a physically based mathematical model [*Lin et al.*, 2009]. An attractive alternative are the artificial intelligence (AI) models, also referred to as machine-learning or data-driven models. AI-type models are constructed by employing flexible and adaptive model structures in an empirical format and can be used to determine the relationship between inputs and outputs. By far, the most popular AI-type of model is the Artificial Neural

Network (ANN). Several researchers have used different types of ANN algorithms to forecast precipitation [*Raman and Sunilkumar*, 1995; *Kuligowski and Barros*, 1998; *Tokar and Johnson*, 1999; *Hsu et al.*, 1995, 1997; *Tokar and Markus*, 2000]. A detailed review of the ANN applications in hydrology is available from the work of the *ASCE Task Committee* [2000b].

[7] Although ANNs are powerful mathematical structures, they have certain drawbacks that include getting trapped in local minima and subjectivity in selection of model architecture [*Suykens*, 2001, *Lin et al.*, 2009]. Due to these drawbacks, there is a need for a more sophisticated AI-type data-driven model that is capable of efficiently representing the multifaceted interaction between oceanic-atmospheric oscillations and precipitation.

[8] Another type of data-driven model becoming popular in ANN-dominated fields is the Support Vector Machine (SVM). SVM is based on the structural risk minimization (SRM) principle, which helps it to efficiently relate the relative input to the desired output and perform better on unseen data [*Vapnik*, 1995, 1998]. In other words, SVM is a statistical tool that approaches the problem of training polynomial functions, radial basis functions, or neural network regression estimators in a way that is similar to neural networks; however, at the same time, it uses a new approach [*Liong and Sivapragasam*, 2002].

[9] In comparing SVMs and ANNs, there are two major differences. First, ANN is constructed based on the principle of empirical risk minimization (ERM), whereas SVMs are based on the SRM principle. The ERM principle only minimizes the total error (empirical risk), whereas the SRM principle minimizes empirical risk and addresses the model complexity against its success at fitting the training data set in order to avoid over fitting [*Haykin*, 2003]. This results in a better capability for generalization. Second, quadratic optimization problem is used to express the architecture and weight of SVM. The optimization problem can be solved using standard programming algorithms. Due to the set of weights used in SVM, useful information is revealed to the user in terms of model error and global optimum is guaranteed. Contrary to this, the weight and architecture of ANN is determine by trial and error procedure, which is iterative and time consuming. For these reason, SVM models have been applied successfully in hydrology to forecast streamflow [*Kalra et al.*, 2012; *Lin et al.*, 2010; *Kalra and Ahmad*, 2009; *Asefa et al.*, 2006; *Yu and Liong*, 2007; *Liong and Sivapragasam*, 2002; *Dibike et al.*, 2001], precipitation [*Lin et al.*, 2009; *Tripathi et al.*, 2006], and soil moisture [*Ahmad et al.*, 2010; *Gill et al.*, 2006]. The majority of these applications have shown the superiority of SVM modeling over the traditional ANN modeling approach.

[10] However, even though SVMs are considered superior to ANNs and other regression methods, they still are statistical data-driven models. A number of factors determine the accuracy of an SVM [*Boser et al.*, 1992; *Guyon et al.*, 1993; *Schölkopf et al.*, 1997; *Smola et al.*, 1998]: (1) the choice of kernel, which is responsible for the data transformation into the high-dimensional space in which SVM performs regression; (2) the determination of the hyper-parameters used in the model, which can result in overfitting or underfitting and can affect the accuracy of the

model's predictions. To overcome these issues, several algorithms and methods have been developed, and are available in the literature [*Smola and Schölkopf*, 2004].

1.4. Motivation for Current Research

[11] As evident from the documented literature, the qualitative relationship between oceanic-atmospheric oscillations and precipitation has been studied extensively [*Ropelewski and Halpert*, 1986; *Redmond and Koch*, 1991; *Hurrell*, 1995; *Piechota and Dracup*, 1996; *Cayan et al.*, 1998; *McCabe and Dettinger*, 1999; *Wang et al.*, 2000; *Giannini et al.*, 2001; *McCabe et al.*, 2004]. However, little attention has been paid to statistically understanding the coupled impact of oscillations on precipitation for the Colorado River Basin by using an AI-type data driven model. This current study uses the four oceanic-atmospheric oscillations—PDO, NAO, AMO, and ENSO—to estimate annual precipitation with a 1 year lead time for 17 climate divisions encompassing the Colorado River Basin. These are the only four oceanic indices for which reconstructed data have been developed by using tree ring information; data are available from the National Climate Data Network website http://www.ncdc.noaa.gov/paleo/recons.html [*Carrier et al.*, 2011]. These oscillations have been studied by numerous researchers, using different time scales to show their interaction with the western U.S. hydroclimatology, particularly the Colorado River Basin [*Piechota and Dracup*, 1996; *McCabe et al.*, 2004; *Webb et al.*, 2004; *Kim et al.*, 2006, 2008; *Ellis et al.*, 2010]. Hydroclimatic variability within the CRB, using PDO and ENSO as climate indices, has been reported extensively [*Kahya and Dracup*, 1993; *Piechota and Dracup*, 1996; *Mantua et al.*, 1997; *Cayan et al.*, 1999; *Pulwarty and Melis*, 2001]. PDO in combination with AO and PNA has been studied to improve the lead times for streamflow and precipitation [*Coulibaly et al.*, 2000; *Coulibaly*, 2006; *Kim et al.*, 2006, 2008]; however, the coupled impact of PDO and NAO for the Upper Basin has not been explored. Attempts have been made to study the relationship between AMO and the hydrologic conditions within the CRB [*Webb et al.*, 2004]. Contrary to this, the NAO primarily has been studied in relation to changes in mean sea level pressures (SLP) over the Arctic Ocean [*Walsh et al.*, 1996], trends in surface wave heights over the North Atlantic [*Kushnir et al.*, 1997], and predicting storm activity and shifts in storm tracks in the Atlantic Ocean [*Hurrell*, 1995]. Lesser attention has been given to the changes in precipitation in relation to NAO within the Colorado River Basin. Majority of the literature suggests that tropical Pacific drives the hydrologic variability over CRB. However, there have been attempts to study the relationship between Atlantic SST and Colorado River Basin [*Kim et al.*, 2006, 2008; *Ellis et al.*, 2010]. Although, it is well established that tropical Pacific modes of climate, i.e., PDO and ENSO are related to the hydrologic conditions in the western U.S.; there are other Atlantic climate patterns that are also believed to be statistically connected with the western U.S. hydroclimatology [*Pulwarty and Melis*, 2001]. The physical reason of the relationship between Atlantic SST's and western U.S. hydroclimatology is uncertain; it is still necessary to investigate both Pacific and Atlantic climate modes [*McCabe et al.*, 2007; *Thomas*, 2007; *Kim et al.*, 2008].

[12] The use of climate oscillation information has the potential to improve hydrologic forecasts within a basin [*Piechota and Dracup*, 1996; *Cayan et al.*, 1998; *Coulibaly et al.*, 2000; *Kim et al.*, 2008; *Kalra and Ahmad*, 2009; *Pui et al.*, 2011]. This research is an attempt to perform a statistical analysis, by incorporating oceanic-atmospheric oscillations in a SVM model, to estimate annual precipitation with a 1 year lead time for the Colorado River Basin. Once the annual precipitation with a 1 year lead time has been estimated, such stochastic techniques as K Nearest Neighbor (KNN) can be used to temporally disaggregate the precipitation into seasonal, or monthly rainfall, depending on the need of the end user [*Rajagopalan and Lall*, 1998, 1999; *Kalra and Ahmad*, 2011]. Furthermore, annual precipitation estimates are helpful in analyzing the sediment yield within the basin, which varies as a function of annual precipitation [*Wilson*, 1973]. Although there have been attempts to increase the precipitation lead time, whatever modest skills a climatologist may have at predicting a 3–9 month lead time arises from ENSO and its effect [*California Dep. of Water Resources*, 2005; *Kim et al.*, 2006, 2008].

[13] This current study aims to provide annual precipitation totals with a 1 year lead time by using oceanic-atmospheric indices in a data-driven model. It should be noted that when estimating annual cumulative precipitation, as done in the current study, the standard definition of "lead time," i.e., lags between the predictors and the predictand, may not be appropriate. The current study uses values for the previous year's oceanic oscillations, and estimates annual precipitation totals for the current year; this is referred to as a 1 year lead time. Thus, on 1 January of the current year, the cumulative total precipitation for the entire current year can be known using the proposed modeling approach.

[14] Generally, in any SVM type regression problem, the time series of the data is split into specific training and testing data sets. The model is trained on only one data set and tested on the other unseen data set. We propose a modified framework of SVM modeling approach, in which a data point is held out and the model is trained on the remaining data and then tested on the held out data. This process is repeated for the entire length of the data by moving the model forward in time. This SVM approach has been previously applied in evaluating contaminants in groundwater by *Khalil et al.* [2005] but has never been used in a climate related research. This is the first attempt to use this modified SVM predictive framework in a hydrologic forecasting study incorporating large-scale climate patterns.

[15] Along with extending the forecast lead time by using four oceanic-atmospheric oscillations, current research performs a rigorous sensitivity analysis to statistically determine the coupled impact and also the individual impact of each oscillation mode with respect to annual precipitation. Additionally, the robustness of the SVM approach is verified using a bootstrapped cross validation technique.

[16] Furthermore, the SVM precipitation estimates are compared with a feed-forward, back-propagation ANN model as well as a Multivariate Linear Regression (MLR) model. All models are evaluated using root-mean-square error (RMSE), mean error (ME), RMSE-observation standard deviation ratio (RSR), correlation coefficient (R), Nash Sutcliffe coefficient of efficiency (NSE), percent bias (Pbias), and linear error in probability space (LEPS) skill

score (SK). It should be noted that an exhaustive comparison between different methods for estimating precipitation, under different circumstances, is not the goal of this study. Instead, the results of the SVM model were compared to ANN and MLR models to explore how well each model was able to estimate precipitation within CRB by using oceanic-atmospheric oscillations.

[17] This paper is organized as follows. Section 2 presents the description of the study region. The precipitation data for the climate divisions is described in section 3. The description of the SVM modeling approach for estimating precipitation, and the performance measures used to evaluate the model performance are given in section 4. Section 5 describes the statistical properties of oscillations and annual precipitation. Section 6 provides a discussion of the SVM modeling results. A comparison of the SVM precipitation estimates is presented with estimates obtained from the ANN and MLR models. Section 7 summarizes and concludes the paper.

2. Study Region

[18] The Colorado River is the most regulated river in the United States, and is governed by the "Law of the River" [*Sax et al.*, 2000]. It encompasses seven states, and is a major source of water to the southwestern United States (Figure 1a). It provides industrial and municipal water to nearly 25 million people by means of existing reservoirs, Lake Powell, and Lake Mead [*Sax et al.*, 2000]. It provides agriculture water for nearly 3 million acres, and produces 11.5 billion kilowatt-hours of hydroelectric power. The Colorado River Basin is composed of the Upper Basin (Wyoming, Colorado, Utah, and New Mexico) and the Lower Basin (California, Nevada, and Arizona). The flow demand between the upper and the lower basin is established by the flow at Lee's Ferry (depicted by a triangle in Figure 1a), which acts as the hydrologic divide. The majority of the flow, nearly 90%, is generated in the Upper Basin from the spring-summer snowmelt [*Prairie and Callejo*, 2005]. Based on the flow contribution, the Upper Basin is subdivided further into eight subbasins (Figure 1b).

[19] The United States is divided into 344 climate divisions, based on the climatic boundaries (available at http://www.esrl.noaa.gov/psd/data/usclimate/map.html). The Colorado River Basin encompasses 29 climate divisions. Out of these 29 climate divisions, 17 divisions have greater than 30% of their area within the Colorado River Basin (Figure 1a). For the purpose of this study, the climate divisions have been sorted according to different states, and have been numbered from 1–17. Table 1 show the nomenclature used to identify each climate division within a particular state. Divisions 1–7 and 10–12 are within the Lower Basin, and Divisions 8–9 and 13–17 encompass the Upper Basin. An area-weighted method was employed in the Geographic Information System to compute the flow contribution from each of the Upper Basin climate divisions. The climate divisions were merged with the subbasins of the Upper Basin (Figure 1b) to compute the percentage contribution of each subbasin within that respective division. This resulted in a maximum flow generated by Climate Division 8 (~57%), followed by Climate Division 17 (~14%), and 16 (~11%), as shown in Figure 1c. The remaining divisions in the Upper Basin generate less than 6% flow,

individually. It should be noted that portions of the San Juan and Dirty Devil subbasins do not intersect with any of the Upper Basin climate divisions, and therefore were not included in the calculations. This accounts for approximately 5% of the flow. Therefore, the total flow percentage in Figure 1c adds up to less than 100%. In the LCRB, majority of the average annual streamflow occurs in the winter and spring seasons. The northwestern and central regions of LCRB generate approximately 70% flow, whereas approximately 35% flow is generated in the southern LCRB [*Thomas*, 2007].

3. Data

[20] The data sets used to estimate annual precipitation are the oceanic-atmospheric modes of the Pacific and the Atlantic Oceans (Figure 2) as well as the precipitation time series (Figure 3). A brief description of the precipitation data is provided in this paper. The oscillation data used in the current analysis is similar to the authors' previous work; a detailed description can be found in the work of *Kalra and Ahmad* [2009].

[21] The precipitation data used in this study is the average monthly time series data for 17 climate divisions, covering a period from 1901–2008. The monthly data set is added to obtain the annual precipitation time series for each of the climate division. This data is obtained from the National Climate Data Center (NCDC) http://www.esrl.noaa.gov/psd/cgi-bin/data/timeseries/timeseries1.pl. The NCDC prepares the data over each climate division by taking an average of temperature and precipitation from stations within a division, reported by the National Weather Service (NWS) Cooperative Observer Program (COOP). The data set is corrected for time bias by adjusting for the variations in average monthly mean temperatures, as described in the work of *Karl et al.* [1986]. The divisional averages are computed starting 1931 to present, whereas each station value was computed directly within a state from 1895–1930 [*Guttman and Quayle*, 1996]. Therefore the count and distribution of the stations within COOP have changed over time. This may be considered a limitation in the data set, but the data corresponds well both spatially and temporally to large-scale historic climate anomalies, such as drought [*Guttman and Quayle*, 1996].

[22] It should be noted that the CRB is composed of highly varied elevations and climate regimes, and it is difficult to integrate all precipitation contributions into a single time series that is representative of the entire basin precipitation. Therefore, the climate division data used in the analysis is helpful in representing the temporal and spatial variation of precipitation within CRB. The annual spread of the input data for each climate division is shown in vertical box plots in Figure 3. The horizontal line inside each box shows the median value. The box represents the 25th and 75th percentile (interquartile range) values, and the whiskers extend from 5th to 95th percentile values. The dot inside the box shows the historic mean of the input data.

[23] The box plots show that the annual precipitation within the CRB exhibits a higher degree of variability, as indicated by wider box plots for the majority of the climate divisions. Estimating this variability by using oceanic-atmospheric oscillations is a challenging task. It is noteworthy that the division with maximum flow in the upper basin, Climate Division 8 (Figure 1b), does not correspond to the largest
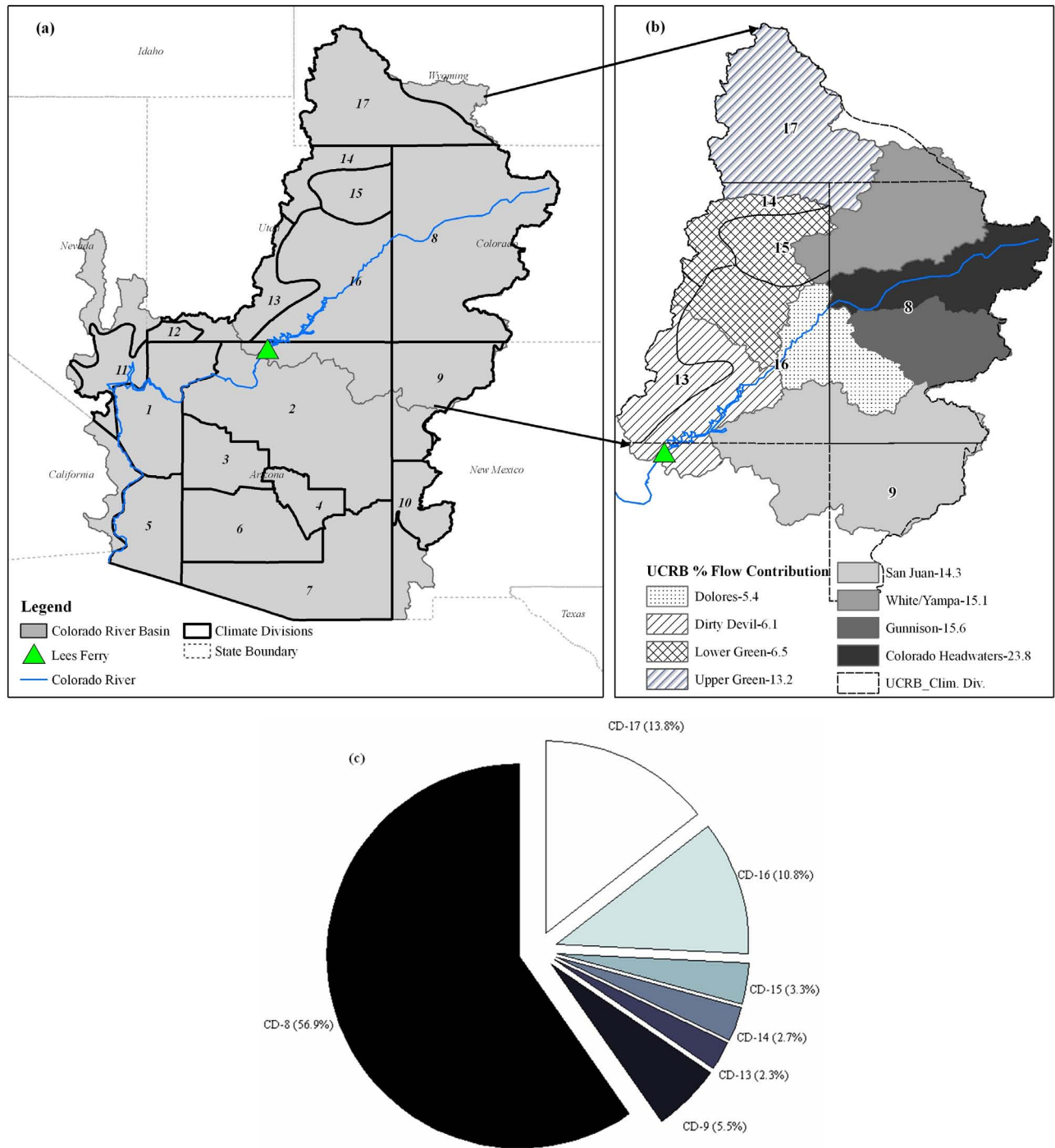
Figure 1. Map showing the location of (a) the Colorado River Basin and the 17 climate divisions, (b) percent flow contribution from UCRB to Colorado River, and (c) flow generated from each climate division in the Upper Basin. The location of Lees Ferry is indicated by a triangle.

precipitation, as shown in Figure 3. The division with a lower flow, Climate Division 14, experiences large precipitation amounts (Figure 3). The most probable cause for this difference is the effective precipitation, that is, precipitation that is realized as runoff. Precipitation in Climate Division 8 mostly is seen as upper elevation snowpack that gradually melts and contributes significantly to streamflow [*Hamlet et al.*, 2005;

Feng and Hu, 2007]. The precipitation in Climate Division 14 probably is due to lower elevation rainstorms, which are more susceptible to infiltration, evaporation, etc.; as a result, this water does not end up contributing significantly to the Colorado River Basin flow [*Colle*, 2004]. This emphasizes the fact that there is not a linear correlation between precipitation and streamflow within Colorado River Basin.

Table 1. List of Climate Divisions Used in the Study

Climate Division	Name	State	Region
1	NORTHWEST	AZ	Lower Basin
2	NORTHEAST	AZ	Lower Basin
3	NORTH CENTRAL	AZ	Lower Basin
4	EAST CENTRAL	AZ	Lower Basin
5	SOUTHWEST	AZ	Lower Basin
6	SOUTH CENTRAL	AZ	Lower Basin
7	SOUTHEAST	AZ	Lower Basin
8	CO DRAINAGE BASIN	CO	Upper Basin
9	NORTHWESTERN PLATEAU	NM	Upper Basin
10	SOUTHWESTERN MOUNTAINS	NM	Lower Basin
11	EXTREME SOUTHERN	NV	Lower Basin
12	DIXIE	UT	Lower Basin
13	SOUTH CENTRAL	UT	Upper Basin
14	NORTHERN MOUNTAINS	UT	Upper Basin
15	UINTA BASIN	UT	Upper Basin
16	SOUTHEAST	UT	Upper Basin
17	GREEN AND BEAR DRAINAGE	WY	Upper Basin

4. Method

[24] This section describes the modified SVM modeling framework abstracted from *Kalra* [2012] and also provides a description of the model evaluation performance measures.

4.1. SVM Modeling

[25] This section provides a brief explanation of the underlying principles of SVM abstracted from *Vapnik* [1995].

A more detailed discussion on the subject can be found in the work of *Vapnik* [1995, 1998]. The concept of Support Vector Regression (SVR) is to nonlinearly map the input data x into a higher dimensional feature space such that:

$$f = w\Phi(x_i) + b. \tag{1}$$

[26] In equation (1) $\Phi(x_i)$ is an input feature and both w and b are coefficients that are estimated by minimizing the regularized risk function,

$$R(f) = C\frac{1}{N}\sum_{i=1}^{N}(\xi_i + \xi_i^*) + \frac{1}{2}\|w\|^2 \tag{2}$$

$$\text{Subject to} \begin{cases} y_i - \sum_{j=1}^{K}\sum_{i=1}^{L}w_j x_{ji} - b \leq \varepsilon + \xi_i \\ \sum_{j=1}^{K}\sum_{i=1}^{L}w_j x_{ji} + b - y_i \leq \varepsilon + \xi_i^*. \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{3}$$

[27] In equations (2) and (3), C is the cost, K is the number of support vectors; and ε is called the Vapnik's insensitive loss function. The Vapnik's ε-insensitive loss functions act as a threshold in the sense that errors less than ε are not considered. Additionally, ξ_i and ξ_i^* are the slack variables.
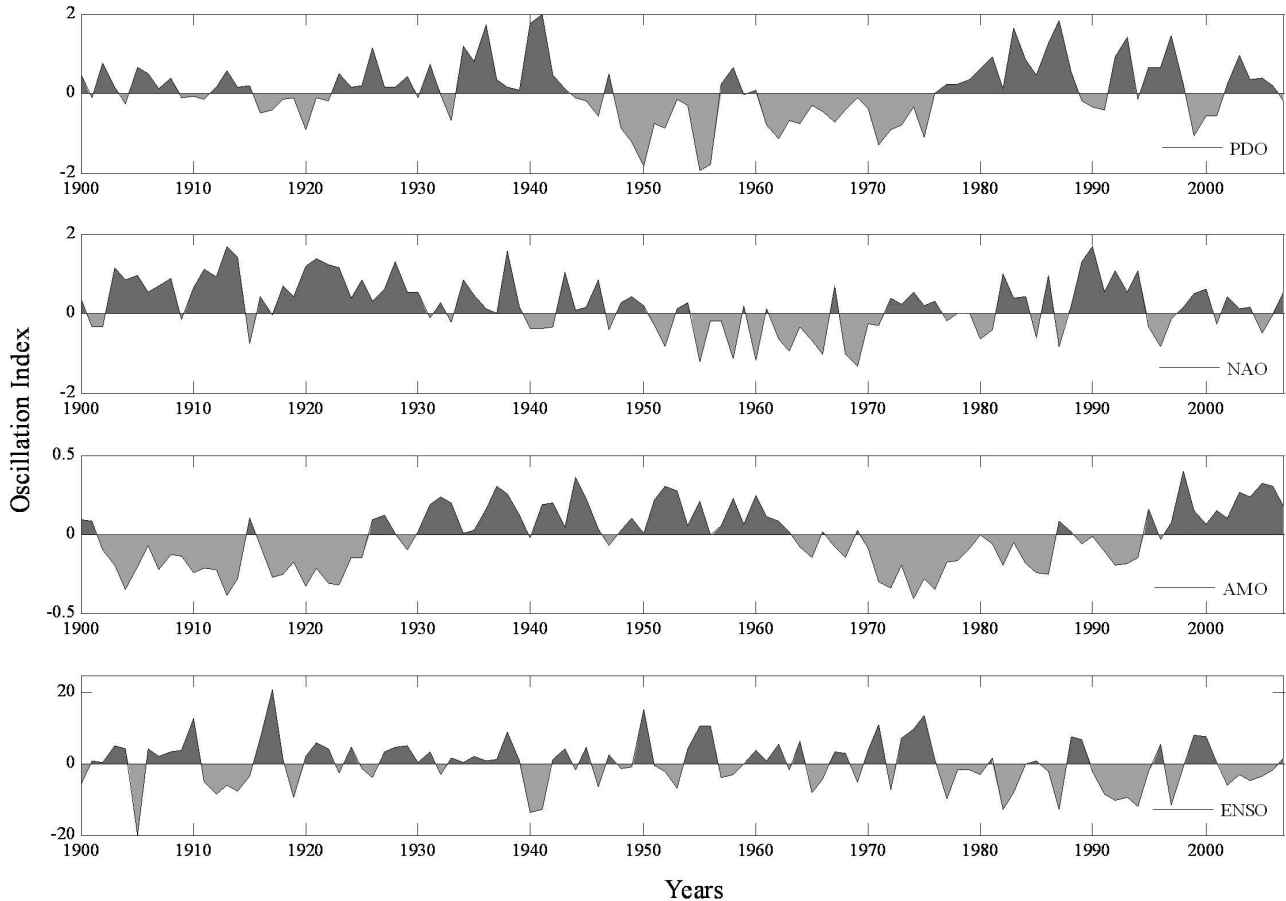

Figure 2. Time series plot of annual average oscillations used in the study.
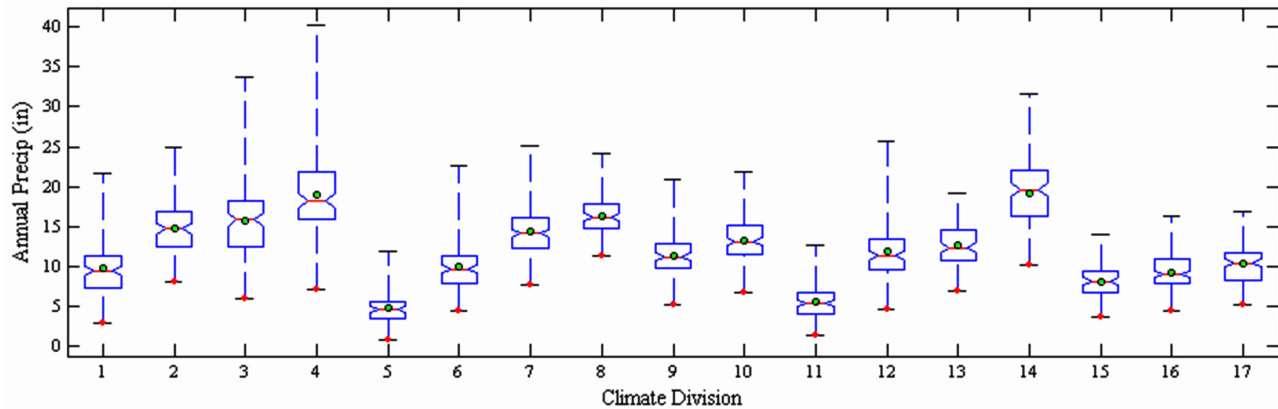
Figure 3. Box plots depicting annual precipitation data from 1901–2008 for 17 climate divisions encompassing the Colorado River Basin. The horizontal line inside the box shows the median value. The box represents the 25th and 75th percentile values (an interquartile range), and the whiskers extend from 5th to 95th percentile values. The circular dot inside the box represents the long–term mean of annual precipitation.

If errors are larger than ε, the slack variables are used to determine the degree to which samples will be penalized. Interested readers are referred to *Kalra and Ahmad* [2009] for the working mechanism as well as an example of the SVM modeling approach.

[28] As mentioned in the work of *Twarakavi et al.* [2006], a typical modeling framework of any AI model consists of the following four steps: (1) preparation of specific training and testing data suitable for model, (2) training the model using the specific trained data set, (3) testing the trained model using the testing data set, and (4) cross-validating the model using the entire data set. Step 1 is essential in every data-driven modeling application. Step 2 fits the model and step 3 is used to evaluate the model performance on unseen data. Step 4 is applied at the end to show the robustness of the model: it tests to see if different training and testing datasets do not yield different results that lead to different conclusions. To make the SVM modeling approach more robust and to improve the efficiency of the SVM approach as a better forecasting tool, the current research proposes a modified SVM modeling framework. The proposed modification tests the model for each sample in the data and is not limited to only single training and testing sample. Additionally, the robustness of the modified SVM prediction framework is verified using a bootstrapping cross validation technique.

4.1.1. Modified SVM Framework

[29] The modified SVM approach framework is a special case of the k-fold cross validation technique [*Geisser*, 1975; *Stone*, 1974] in which k = 1. Each data point in the time series is held out in turn; the model is trained on the remaining data set (N-k, where N is the number of observations in the data set) and tested on the held-out data point, i.e., k. This process is repeated for the entire length of data (k = N).

[30] Let i represent the current instance, which is also representative of the testing instance (i = 1). For each instance i, the training set will consist of [1, i) U (I, N]. This will train the model on all instances, except for instance i, and will test the model on instance i only. This

process is repeated for all instances and stops when i = N + 1. A brief description of the step-by-step algorithm is described below.

[31] Step 1. Let [Xi] represent the data matrix comprising of all observations used in the study of length N, where i is the featured instance and varies from 1:N.

[32] Step 2. Partition matrix [Xi] into two submatrices, [Ai] and [Bi], such that [Ai] is of length N-i and [Bi] is of length i.

[33] Step 3. Train the SVM model on [Ai] and test the model on [Bi].

[34] Step 4. Repeat steps 1–3 for all the featured instances.

[35] Step 5. Evaluate model performance for all instances (pooled) of [Bi].

[36] Step 6. Apply steps 1–5 for other climate divisions.

4.1.2. Bootstrap Cross Validation

[37] To establish the robustness of the SVM approach, it is necessary to determine that different training and testing samples does not yield different results. For this purpose a cross validation approach is needed to determine the performance of the predictive model. Several cross validation approaches, i.e., k-fold, leave out, bootstrapping etc. have been documented in available literature [*Khalil et al.*, 2005; *Asefa et al.*, 2006; *Twarakavi et al.*, 2006; *Chowdhury and Sharma*, 2009; *Lin et al.*, 2009; *Li et al.*, 2010]. All of the cross validation approaches have their advantages and disadvantages but help in assessing how the results of a statistical analysis will generalize on an independent data set. The current study used a bootstrap cross validation technique to test the stability of the SVM model results. Bootstrap or bagging is a statistical procedure that uses intensive resampling, with replacement to reduce the uncertainties in the data [*Efron*, 1979]. The bootstrap method is used to generate different realizations of the measured data that can be used to assess the mean and the variability of the estimates. In this study each SVM model is trained on *a* set of bootstrap samples and tested on the entire measured data set. The resampling procedure is repeated 100 times and results in 100 ensembles for each

measured value. The final estimated value is the mean of these 100 ensembles.

[38] Based on the modeling approach described, four SVM models are developed using annual oceanic-oscillations PDO, NAO, AMO, and ENSO for time step "t" in order to estimate precipitation at "t + 1" (t is in year) for 17 climate divisions encompassing the CRB. Each climate division is considered independent, and separate SVM models are developed for each division. The SVM models are developed by using the standard software package included in the Comprehensive R Achieve Network (available at http://www.r-project.org/). Model I is termed as the "base case," and uses all four oceanic modes to estimate annual precipitation. To estimate precipitation, Models II and III use a combination of 3 and 2 oscillation modes, respectively. Model IV uses a single oscillation mode to estimate precipitation. The major reason for developing Models II–IV is to evaluate the role of individual and coupled oceanic-oscillations in estimating precipitation within the basin.

[39] Based on equation (2), the performance of any SVR formulation depends on the selection of hyper-parameters: cost (C), insensitivity value (ε), and the radial basis kernel width (γ). Previous studies have used the following three procedures to estimate hyper-parameters in any SVR formulation: (1) user expertise, (2) grid based search, and (3) using the statistical properties of the training data in an analytical approach. In the current study, a grid-based search was adopted to compute the hyper-parameters. In grid based search every possible combination within a feasible hyper-parameters space is considered and prediction error is computed for each combination. The feasible parameter space for each hyper-parameter is constructed using the minimum and maximum possible values that are given a priori ($0.001 < C < 1000$, $0.001 < \gamma < 100$, and $0 < \varepsilon < 1$). An increment of 0.01 is selected that helps in ensuring the optimality and computational efficiency of the grid based approach. The hyper-parameters, which result in minimal mean square error, are selected as optimal values. It should be noted that the hyper-parameters for all years are not exactly the same (as training data changes) but correspond to a tight cluster. A number of previous studies have used similar approach [*Asefa et al.*, 2006; *Gill et al.*, 2006; *Kalra and Ahmad*, 2009; *Twarakavi et al.*, 2006, 2009; *Ahmad et al.*, 2010]. Furthermore, the performance of SVM depends on the choice of kernel as being a kernel's parameterization problem. In this study, a radial basis kernel is used in the SVM model; this has performed better when compared with other kernels, such as linear, polynomial, sigmoid or spline, as evident in the past studies [*Schölkopf et al.*, 1997; *Smola et al.*, 1998; *Dibike et al.*, 2001; *Yu and Liong*, 2007].

[40] Additionally, the performance of SVM model is compared with the two other hydrologic time series modeling approaches, i.e., ANN and MLR. A feed-forward backpropagation ANN model with one input layer, one hidden layer, and one output layer containing a single node was used in the current study. A tan sigmoid transfer function was used to transfer the input signal to the output through the hidden layer neurons. A linear transfer function was used at the output node. Furthermore, this ANN-type model has been used in other modeling studies involving different

hydroclimatic variables [*Raman and Sunilkumar*, 1995; *Kuligowski and Barros*, 1998; *Tokar and Johnson*, 1999; *Hsu et al.*, 1995; *Ahmad and Simonovic*, 2005; *Melesse et al.*, 2011]. A more detailed description on the theoretical aspects of ANN is available in the work of the *ASCE Task Committee* [2000a]. The other type of model developed is the parametric Multivariate Linear Regression model, which consists of oceanic-oscillations as the predictors and annual precipitation as the predictand. The modeling approach used to develop the corresponding ANN model and MLR model is similar to the SVM approach.

4.2.　Model Performance Evaluation

[41] The current study uses seven performance measures: RMSE, ME, RSR, R, NSE, Pbias, and LEPS SK. Lower RMSE and ME represent better model performance [*Singh et al.*, 2005]. R determines the linear association between the measured and predicted value. [*Legates and McCabe*, 1999]. RSR standardizes RMSE by using the standard deviation of observations [*Singh et al.*, 2004]; it is calculated as the ratio of RMSE and standard deviation of observed data. *Singh et al.* [2004] published the guidelines for RSR based on lower RMSE values. An RMSE value of less than half the standard deviation of observation is considered low. Therefore, *Moriasi et al.* [2007] used the recommended value (per *Singh et al.* [2004]) of less than 0.5 RSR to categorize model performance as "very good," and suggested a less stringent rating of 10% points and 20% points greater than 0.5 RSR to be "good" and "satisfactory." RSR as mentioned in the work of *Moriasi et al.* [2007] can be statistically expressed as

$$\text{RSR} = \frac{\text{RMSE}}{\text{STDEV}_{obs}} = \frac{\sqrt{\sum_{i=1}^{n} (Y_i^{obs} - Y_i^{est})^2}}{\sqrt{\sum_{i=1}^{n} (Y_i^{obs} - Y^{mean})^2}}. \quad (4)$$

[42] NSE is used to access predictive power of hydrological models [*Nash and Sutcliffe*, 1970]. NSE is computed as

$$\text{NSE} = 1 - \frac{\sqrt{\sum_{i=1}^{n} (Y_i^{obs} - Y_i^{est})^2}}{\sqrt{\sum_{i=1}^{n} (Y_i^{obs} - Y^{mean})^2}}. \quad (5)$$

[43] NSE ranges from $-\infty$ to 1, where NSE of 1 corresponds to a perfect match between the observed and predicted values. As reported in the work of *Moriasi et al.* [2007], Pbias measures the average tendency of the estimated data in comparison with the observed values and is computed as

$$\text{Pbias} = \frac{\sum_{i=1}^{n} (Y_i^{obs} - Y_i^{est}) * 100}{\sum_{i=1}^{n} (Y_i^{obs})}. \quad (6)$$

[44] The optimal value of Pbias is 0. Underestimation and overestimation of prediction bias is given by positive

and negative values [*Gupta et al.*, 1999]. In equations 4–6, Y^{obs} is the measured precipitation, Y^{est} is the estimated precipitation, and Y^{mean} is the long-term mean of the measured data.

[45] The model-estimated precipitation is compared against the "climatology," using the LEPS score. According to *Potts et al.* [1996] LEPS score measures the distance between the estimated value and the observed value in terms of their cumulative probability distributions. The LEPS score is defined as:

$$S = 3 * (1 - |P_f - P_o| + P_f^2 - P_f + P_o^2 - P_o) - 1. \quad (7)$$

[46] In equation (7), P_f and P_o are the estimated and observed cumulative probabilities. The LEPS score, which is computed for all the data points, shows the good and bad forecast years. A higher skill score is obtained when accurately forecasting values farther from the mean. Therefore, if the forecast is for a value near the mean, then it will have a lower skill (bad forecast) and a correct forecast farther from the mean will have a higher skill (good forecast). The average skill SK based on the LEPS score is defined as:

$$SK = \frac{\sum S}{\sum S_m} * 100, \quad (8)$$

where summation S is the sum of all years of LEPS score. If S is positive, then summation S_m is the sum of the best possible forecast, with $P_f = P_o$. If S is negative, then summation S_m is the sum of worst forecast computed by $P_f = 1$ or 0. SK ranges from –100 to 100, where a SK of 0 represents the climatological score or equivalently, random data. LEPS SK is considered "good" if it is greater than 10 [*Potts et al.*, 1996]. To evaluate the accuracy of the predictions, i.e., typical forecast error at each climate division, cumulative absolute forecasts errors are computed using the nonexceedence plots.

[47] Table 2, which is abstracted from *Moriasi et al.* [2007], shows the ratings for performance measures for RSR, NSE, R, and Pbias for estimating streamflow at a monthly time step. Based on Table 2, the model performance can be judged satisfactory if RSR ≤ 0.70 and NSE > 0.5 [*Moriasi et al.*, 2007]. *Moriasi et al.* [2007] recommended a value of R > ±0.5 for a model performance to be considered satisfactory; however, the current study used a much greater verification metric than the recommended measures and considered the model performance as satisfactory when R was >0.7.

Table 2. Recommended Performance Measures at Monthly Time Steps[a]

Performance Rating	RSR	R[b]	NSE	|Pbias|
Very good	0.0–0.50	0.85–1.0	0.75–1.0	<10
Good	0.51–0.60	0.81–0.85	0.65–0.75	≤10 and >15
Satisfactory	0.61–0.70	0.71–0.80	0.51–0.65	≤15 and >25
Unsatisfactory	>0.70	≤0.70	≤0.50	>25

[a]Performance measures for RSR, NSE, and Pbias are taken directly from *Moriasi et al.* [2007].
[b]The performance measure for R is greater than the recommended measure.

[48] It should be noted that the variable evaluated in the current analysis is precipitation at the annual time step. *Moriasi et al.* [2007] suggested performance measures for evaluating model performance for estimating streamflow at a monthly time step. Precipitation exhibits greater variability than streamflow, so it becomes questionable to use monthly streamflow performance measures for evaluating annual precipitation estimates. Having said this, the selection of model evaluation guidelines depends on the scope and magnitude of the research problem. Stricter guidelines are needed for projects involving significant consequences, such as congressional testimony or the development of new laws [*Moriasi et al.*, 2007]. Moderate performance ratings would suffice the purpose for explanatory research or technology assessment where no litigation is involved (U.S. EPA, 2002). Furthermore, *U.S. Environmental Protection Agency* [2002] has indicated that the measures used to accept, reject, or qualify model results should be documented before evaluating the model. Also, the annual precipitation used in the analysis has reduced variability compared to monthly precipitation values. Therefore, the stricter performance measures mentioned above are adequate to evaluate the SVM, ANN, and MLR models.

5. Statistical Properties of Annual Precipitation and Its Relation With Oscillation Modes

[49] First, the linear correlation coefficients were computed between the oscillation modes and lag 1 precipitation to evaluate the persistence over time of the different oscillations and precipitation for the CRB. The correlation coefficients generally are computed to examine the potential predictors [*Grantz et al.*, 2005; *Singhrattna et al.*, 2005]. Table 3 shows linear correlation coefficient of lag 1 between the oscillation modes and annual precipitation for 17 climate divisions encompassing the CRB. Significant correlation for each of the combinations is highlighted in bold, and the lowest and the highest correlation values for a particular set are also presented.

Table 3. Correlation Coefficient Between Oscillation Modes and Annual Precipitation for 17 Climate Divisions at 90% Significance Level[a]

Climate Division	Region	PDO	NAO	AMO	ENSO
1	Lower Basin	0.15	**0.19**	−0.12	−0.06[b]
2	Lower Basin	**0.24**	**0.18**	**−0.33**	−0.16
3	Lower Basin	0.12	**0.18**	**−0.33**	−0.10
4	Lower Basin	0.08	**0.19**	**−0.32**	−0.11
5	Lower Basin	**0.23**	0.16	**−0.24**	−0.13
6	Lower Basin	0.13	**0.17**	**−0.30**	**−0.18**
7	Lower Basin	**0.18**	**0.18**	**−0.34**[b]	−0.14
8	Upper Basin	**0.18**	0.10	**−0.23**	−0.12
9	Upper Basin	**0.28**	0.11	**−0.24**	−0.13
10	Lower Basin	**0.32**[b]	0.08	**−0.19**	−0.12
11	Lower Basin	**0.22**	**0.17**	**−0.21**	**−0.22**
12	Lower Basin	**0.29**	0.05	−0.14	**−0.29**[b]
13	Upper Basin	**0.27**	0.04	−0.06[b]	**−0.24**
14	Upper Basin	0.07[b]	−0.01[b]	−0.07	**−0.18**
15	Upper Basin	**0.20**	0.02	−0.15	**−0.16**
16	Upper Basin	0.16	0.05	**−0.19**	−0.15
17	Upper Basin	0.15	**0.22**[b]	**−0.18**	**−0.19**
90% Significant		*11*	*8*	*12*	*7*

[a]The significant correlations are shown in bold.
[b]The lowest and highest correlation values for each subset.

[50] When correlating the PDO index with the 17 climate divisions for a 1 year lead time, the correlation values resulted in 11 climate divisions exceeding 90% significance, with Climate Division 10 having the highest correlation and Climate Division 14 having the lowest correlation. Five of these divisions are in the Upper Basin, and the remaining six are in the Lower Basin. In the case of NAO, 8 climate divisions exceeded the 90% significance, with Division 17 having the highest correlation and Division 14 having the lowest correlation. Seven of these divisions are in the Lower Basin, and the remaining division is in the Upper Basin. When relating AMO and ENSO, Climate Divisions 12 and 7, respectively, exceeded the 90% significance. PDO showed statistically significant correlation with the Upper Basin precipitation, whereas AMO showed a comparable stronger correlation with the Lower Basin precipitation. On the other hand, NAO and ENSO did not show statistically significant correlation with the CRB precipitation (Table 3). Additionally, among the predictors, as expected, only the combinations of PDO-ENSO (±0.50) and NAO-AMO (±0.41) had significant correlations at 90% significance level.

[51] However, to form the basis for a skillful forecast, coefficient values could not be considered alone. Correlations analysis was a first step to verify the potential predictors for each climate division that showed significant relationships with annual precipitation in the CRB. Furthermore, each climate index individually and in combination

was analyzed by means of the SVM model; the results are discussed in the ensuing section.

6. Results and Discussion

[52] First, 1 year lead time precipitation estimates for the 17 climate divisions encompassing the CRB, using the SVM Model I (base case), are discussed. Second, the coupled and individual effects of oscillations on annual precipitation (Models II–IV) are analyzed. Lastly, the annual precipitation estimates obtained using the SVM model is compared with the ANN and MLR model estimates.

6.1. Model I

[53] In Model I, all four oscillations modes are used to estimate annual precipitation with a lead time of 1 year. Figure 4 shows the scatter plots between the measured and estimated annual precipitation for the 17 climate divisions for the period of record (1901–2008). A good match is obtained between the measured and estimated annual precipitation for the majority of the Upper Basin climate divisions compared to the Lower Basin. This is evident by the sample points following the 45° bisector line for majority of the Upper Basin Climate Divisions, indicating a good model fit. For Climate Divisions 1, 5, 6, 11, and 12 in the Lower Basin, the model does fairly well at the low values; however, a few of the high values are scattered away from the bisector, indicating that the model was not able to capture them satisfactorily. The correlation values are greater
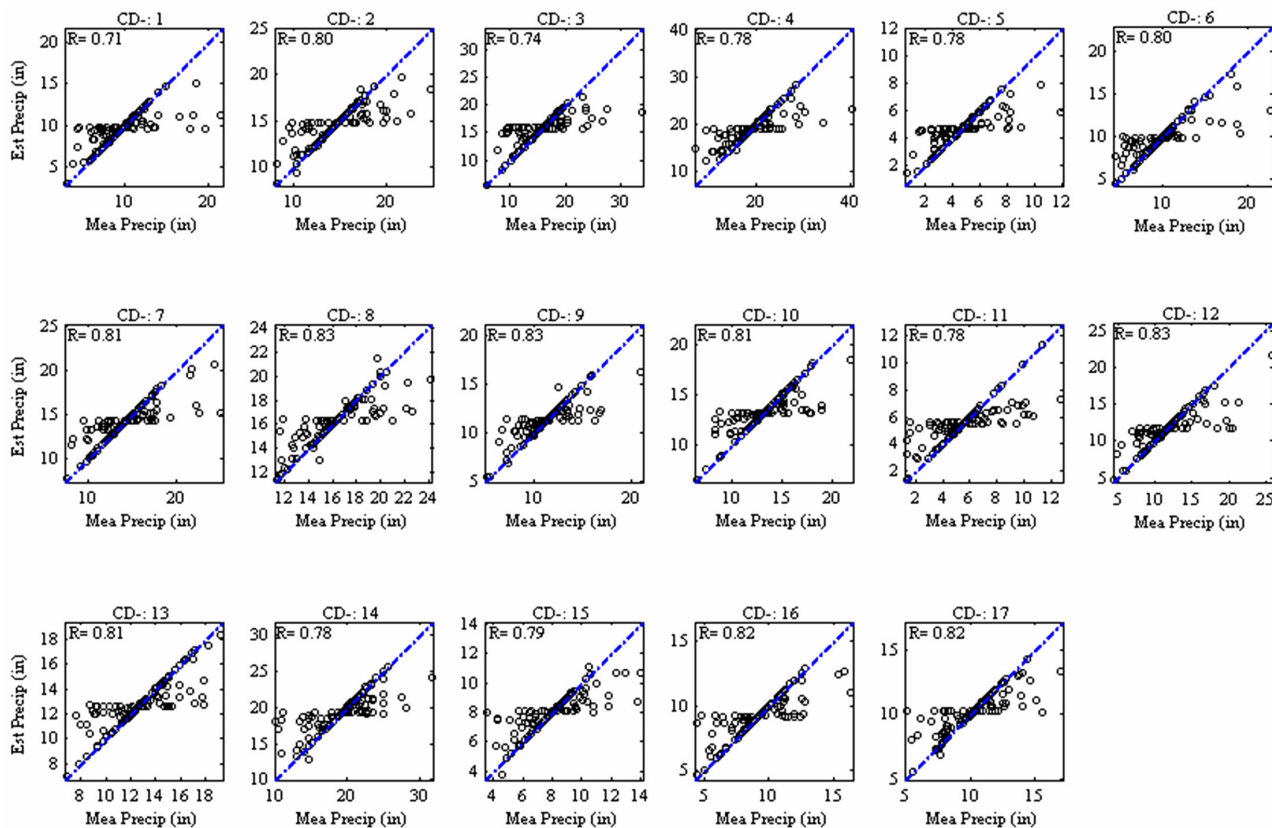

Figure 4.   Scatterplot between measured and SVM estimated precipitation for 17 climate divisions for Model I. Dashed diagonal line is the 45° bisector.

than 0.70 for all the climate divisions. Although, satisfactory correlation values are achieved, it is noticed that for some climate divisions, estimated values are not different for different years; however the observed values are different, evident from the values saturated along a horizontal line. This is mostly witnessed in the Lower Basin climate divisions.

[54] Based on Table 2, Figure 5 shows the spatial map of the 17 climate divisions, depicting the three performance measures of RSR, R, and NSE. Based on RSR, the model shows good precipitation estimates for the majority of the climate divisions (except climate divisions 13 and 14) in the Upper Basin, and satisfactory estimates (except climate division 1) for the climate divisions in the Lower Basin.

[55] Based on Figure 1c, climate divisions generating more than 90% of the Upper Basin flow had RSR error statistics in the range of 0.51–0.60 (good estimates per Table 2). All the divisions in the Lower Basin, except Division 1, had RSR in the satisfactory range (0.61–0.70). The RSR measure indicated that the model performed unsatisfactorily in estimating precipitation for Climate Division 1 (RSR > 0.7). The correlation statistics R agreed with the results of RSR, indicating that a good correlation (0.81 < R < 85) was achieved between the measured and estimated precipitation for the Upper Basin climate divisions and a satisfactory value (R > 0.7) for the climate divisions in the Lower Basin. In case of NSE error statistics, the climate divisions that generate approximately 60% of the Upper Basin flow had good precipitation estimates (NSE > 0.65). The remaining Upper Basin divisions had satisfactory estimates (0.51–0.65), and the Lower Basin's divisions were in the satisfactory range, except for Division 1. Overall, the SVM model was able to provide satisfactory estimates for all the climate divisions, except for Division 1. It should be noted that even in the case of Climate Division 1, results are only slightly below or above the satisfactory levels, with RSR, R, and NSE values of 0.71, 0.71 and 0.49, respectively.

Similar results were noticed for RMSE and ME error statistics, as evident in Table 4. A Pbias of approximately ±2% was achieved for all the climate divisions. A lower value of Pbias signifies that there is less bias in the estimated values and majority of the estimated values are similar to the measured values. Although, there are some values that are overestimated and underestimated, still the values are within a tolerance level having lesser bias and may be considered good. The LEPS SK value ranged between 53% and 64% (Table 4). A LEPS SK score greater than 10% indicates a "good" model. Therefore, the precipitation estimates obtained using the SVM model is much better than the climatology.

[56] Based on Model I result, the SVM model performs satisfactorily in capturing the variability in annual precipitation for a 1 year lead time. The scatter plots (Figure 4) and the spatial maps of the performance measures (Figure 5) show that the model produced good precipitation estimates for the Upper Basin divisions generating two-thirds of the flow; satisfactory estimates were obtained for the Lower Basin divisions, except for Climate Division 1.

[57] Overall, annual precipitation estimates for the CRB are in the range of satisfactory to good for Model I at "t + 1" when using all the four oscillation indices. Additionally, the estimates serve as a better predictor than the "climatology." Although, Model I results indicated that all four oscillations have some influence on the hydroclimatology of the CRB, this needs to be further investigated to evaluate the coupled and individual effects of each oscillation mode. Moreover, the coupled and individual impacts of the oscillations in relation to precipitation may vary within the Upper and Lower Basins.

6.2. Coupled and Individual Response of Oscillation in Relation to Annual Precipitation

[58] To analyze the coupled and individual response of oscillation in relation to annual precipitation, separate
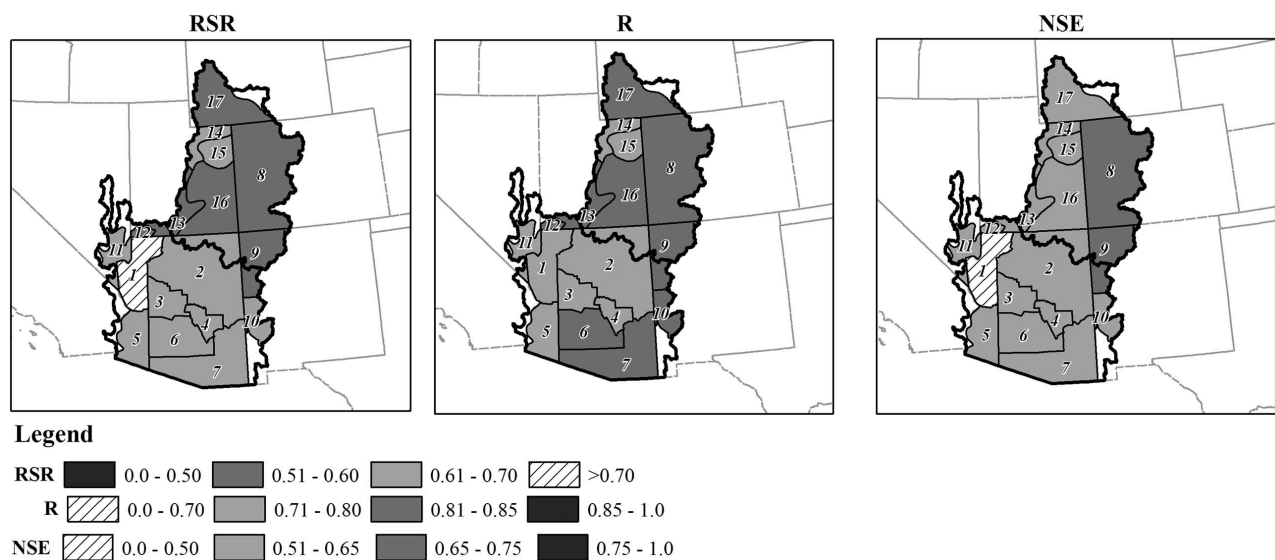

Figure 5. Spatial maps showing the range of performance measures for 17 climate divisions for Model I: (a) RSR, (b) R, and (c) NSE.

Table 4. Performance Measures for SVM Model I Output[a]

Climate Division	RMSE	ME	RSR	R	NSE	Pbias	SK
1	2.38	−0.05	0.71	0.71	0.49	1.07	55.1
2	2.03	−0.01	0.63	0.80	0.60	1.01	58.1
3	3.09	−0.03	0.69	0.74	0.52	0.86	57.0
4	3.50	−0.04	0.66	0.78	0.56	0.53	61.3
5	1.29	−0.10	0.66	0.78	0.56	1.37	54.3
6	2.04	−0.04	0.62	0.80	0.61	0.71	62.7
7	1.98	−0.01	0.61	0.81	0.62	1.30	60.3
8	1.53	−0.01	0.57	0.83	0.67	0.50	62.6
9	1.47	−0.01	0.57	0.83	0.67	0.70	59.5
10	1.64	−0.02	0.60	0.81	0.63	−0.37	57.9
11	1.53	−0.10	0.66	0.78	0.57	2.18	53.0
12	2.14	−0.02	0.59	0.83	0.65	1.61	63.8
13	1.57	−0.03	0.59	0.81	0.65	−0.60	61.6
14	2.57	−0.03	0.64	0.78	0.58	−0.16	55.9
15	1.29	−0.03	0.63	0.79	0.59	0.70	57.3
16	1.38	−0.03	0.60	0.82	0.64	0.21	63.3
17	1.37	−0.02	0.59	0.82	0.65	0.26	62.4

[a]The RMSE and ME values are in inches. The Pbias and SK are in percentage.

SVM models (Models II–IV) were created for each climate division. Precipitation estimates obtained using Model I are used as a baseline to compare precipitation estimates obtained from Models II–IV. This will help to better understand the coupled and individual responses of oscillation modes in relation to precipitation within the two regions of the CRB. Similar to Figure 5 spatial maps depicting the three performance measures RSR, R, and NSE were created for Models II–IV.

[59] In Model II, one oscillation was dropped every time, and the remaining three oscillations were used to predict annual precipitation. This resulted in four models for each climate division; the results are shown in Figure 6. Dropping ENSO and using a combination of PDO, NAO, and AMO as input resulted in an improvement in RSR (Figure 6a), R (Figure 6b), and NSE (Figure 6c) for five climate divisions in the Upper Basin and eight climate divisions in the Lower Basin, compared to Model I results. The divisions showing improvement in the Upper Basin generated approximately 79% of the flow in the Colorado River, based on Figure 1c. The climate divisions showing improvement in Lower Basin encompassed the northern and central portion, generating approximately 70% of the flow [*Thomas*, 2007] in the Lower Basin. Deterioration in RSR, R, and NSE error statistics was noted only for Climate Division 13 in the Upper Basin, generating approximately 2% of the flow. Climate Division 14 showed no change in error statistics compared to Model I results. Based on the flow contribution, the precipitation estimates were in the range of "very good" for the Upper Basin and good for the Lower Basin.

[60] Dropping AMO and using the combination of PDO, NAO, and ENSO as input resulted in improvement in RSR, NSE and R for five climate divisions in the Upper Basin. Whereas, for the Lower Basin, RSR and NSE improved in two climate divisions and R in one climate division (Climate Division 10). The divisions showing improvement in the Upper Basin generated approximately 27% of the flow in the Colorado River. No significant improvements were noted for the Lower Basin climate divisions. The results for the Upper Basin were in the range of "good" whereas satisfactory estimates were obtained for the Lower Basin.

[61] Dropping NAO and using combination of PDO, AMO, and ENSO as input, the three error statistics showed deterioration for majority of the climate division in the Upper Basin (five); however, no significant improvements were noticed for the climate divisions of the Lower Basin. Dropping PDO and using combination of NAO, AMO, and ENSO as inputs, significant improvement in error statistics was noticed for Lower Basin climate divisions. Nine out of ten divisions showed improvement in all the three performance measures (Figure 6). Precipitation estimates for the Lower Basin were "good", compared to "satisfactory" Model I estimates.

[62] To quantify the improvement in performance measures, Table 5 is presented that highlights the percent change in NSE error statistics compared to Model I results. Table 5 clearly indicated that overall the best estimates for the Upper Basin were obtained using a combination of PDO, NAO, and AMO, with six climate divisions showing improvement compared to Model I results. By coupling NAO, AMO, and ENSO precipitation estimates improved for the Lower Basin for a 1 year lead time, with nine climate divisions showing improvement (Table 5). The worst precipitation estimates for the upper basin were obtained when NAO is dropped from the model whereas; dropping AMO resulted in worst estimates for the lower basin (Table 5). Based on Table 2, "very good" precipitation estimates were obtained for the Upper Basin, and "good estimates" were obtained for the Lower Basin climate divisions. Very good precipitations estimates within the UCRB may be helpful for the water managers because, on average, 90% of the streamflow is generated in the Upper Basin above Lees Ferry. These results also indicated that statistically ENSO has a weak association with Upper Basin precipitation and NAO has a strong association (Table 5). Moreover, removing PDO from the model may provide better precipitation estimates for the Lower Basin; whereas deterioration in precipitation estimates is witnessed when AMO is dropped from the model (Table 5).

[63] The spatial map (Figure 6) shows the error statistics between the measured and estimated precipitation for Model II, based on the performance ratings established in Table 2. Figure 7 demonstrates the capability of the proposed modeling framework in relation to capturing high and low annual precipitation values by using the best Model II estimates, with PDO, NAO, and AMO indices, for the Upper Basin. The divisions selected for visualizing the results were Climate Divisions 8, 9, and 16. These climate divisions, which contribute approximately 73% flow to the Colorado River, indicated good-to-very good annual precipitations estimates (Figure 6). The scatterplot shows a close match between the measured and estimated annual precipitations (Figures 7a–7c). The majority of the points are saturated around the 45° bisector, indicating good model fit. Higher R value was achieved. A positive Pbias of less than a half percent was achieved for Climate Division 8, and a negative Pbias of less than a half percent was achieved for Climate Divisions 9 and 16, indicating very good estimates (Table 2). A smaller value of ME is achieved indicating a good match between the measured and estimated precipitation values.
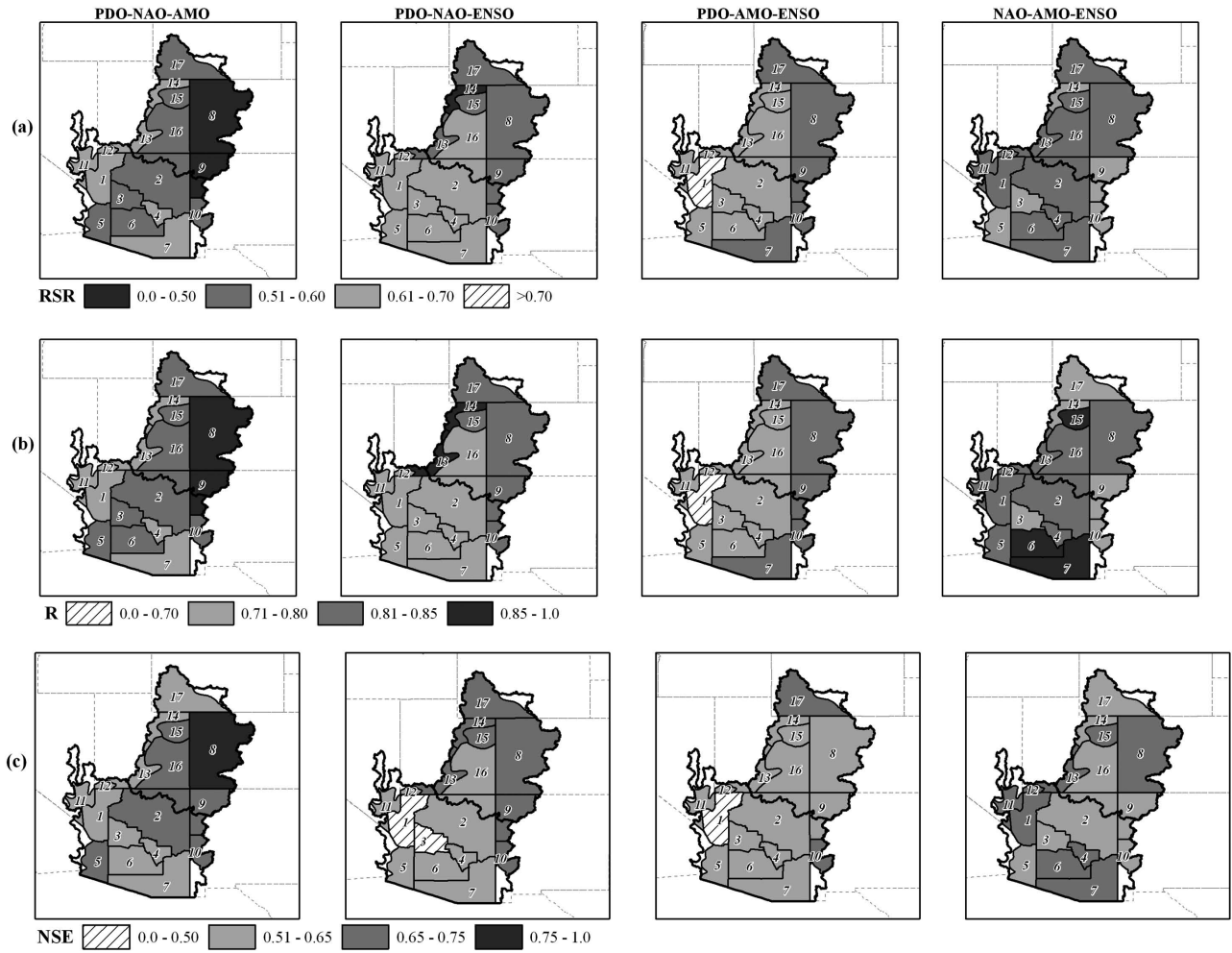
Figure 6. Spatial maps showing the range of performance measures for 17 climate divisions for Model II: (a) RSR, (b) R, and (c) NSE.

Table 5. Percentage Change in NSE Error Statistic for Model II and Model III Compared to Model I Results[a]

Climate Division	Model II Oscillations				Model III Oscillations					
	123	124	134	234	12	13	14	23	24	34
1	15.3	4.1	−15.6	35.3	11.1	−9.3	−29.2	27.3	21.7	20.7
2	9.2	−5.7	−0.3	6.0	4.7	12.7	−20.8	12.9	−2.1	30.4
3	25.2	−4.1	7.4	14.0	20.4	11.4	−11.1	19.2	20.3	36.5
4	10.6	−0.3	−4.8	20.1	12.3	−2.0	−5.3	−5.6	22.3	13.9
5	21.1	−6.5	5.7	11.1	7.7	16.1	−2.7	−2.4	14.7	11.8
6	5.3	−7.4	−1.7	13.0	−3.6	−10.8	−8.8	−4.9	9.4	21.9
7	−2.1	−7.1	2.8	14.2	0.9	−1.9	−6.3	−4.6	−14.7	5.4
8	15.2	0.5	−2.7	0.8	−0.5	2.3	−23.6	−0.6	−4.7	−1.7
9	11.3	3.0	−2.3	−8.5	12.7	−6.9	−36.9	7.5	−12.8	−0.4
10	8.2	10.9	6.0	−13.7	19.2	9.5	−16.1	2.7	−14.7	16.8
11	9.4	−1.7	2.2	17.7	5.1	27.6	1.7	6.0	12.7	23.4
12	−3.8	−14.4	−15.5	6.4	−5.7	−10.6	−17.2	−7.2	2.8	8.4
13	−2.9	13.9	−7.5	6.0	−1.2	12.9	−17.1	4.0	1.1	0.7
14	6.8	28.4	4.9	3.6	17.2	−7.8	21.7	9.8	22.4	14.4
15	20.0	17.1	−7.5	21.7	21.0	10.5	−9.4	10.4	7.0	−6.6
16	10.9	−6.3	−9.1	−0.4	−6.7	−2.6	−34.7	0.6	−0.8	1.7
17	0.7	10.9	2.7	−1.4	5.3	−2.0	1.2	11.4	22.2	−6.8

[a]The negative sign implies decrease in NSE value. 1, 2, 3, and 4 refer to PDO, NAO, AMO, and ENSO.
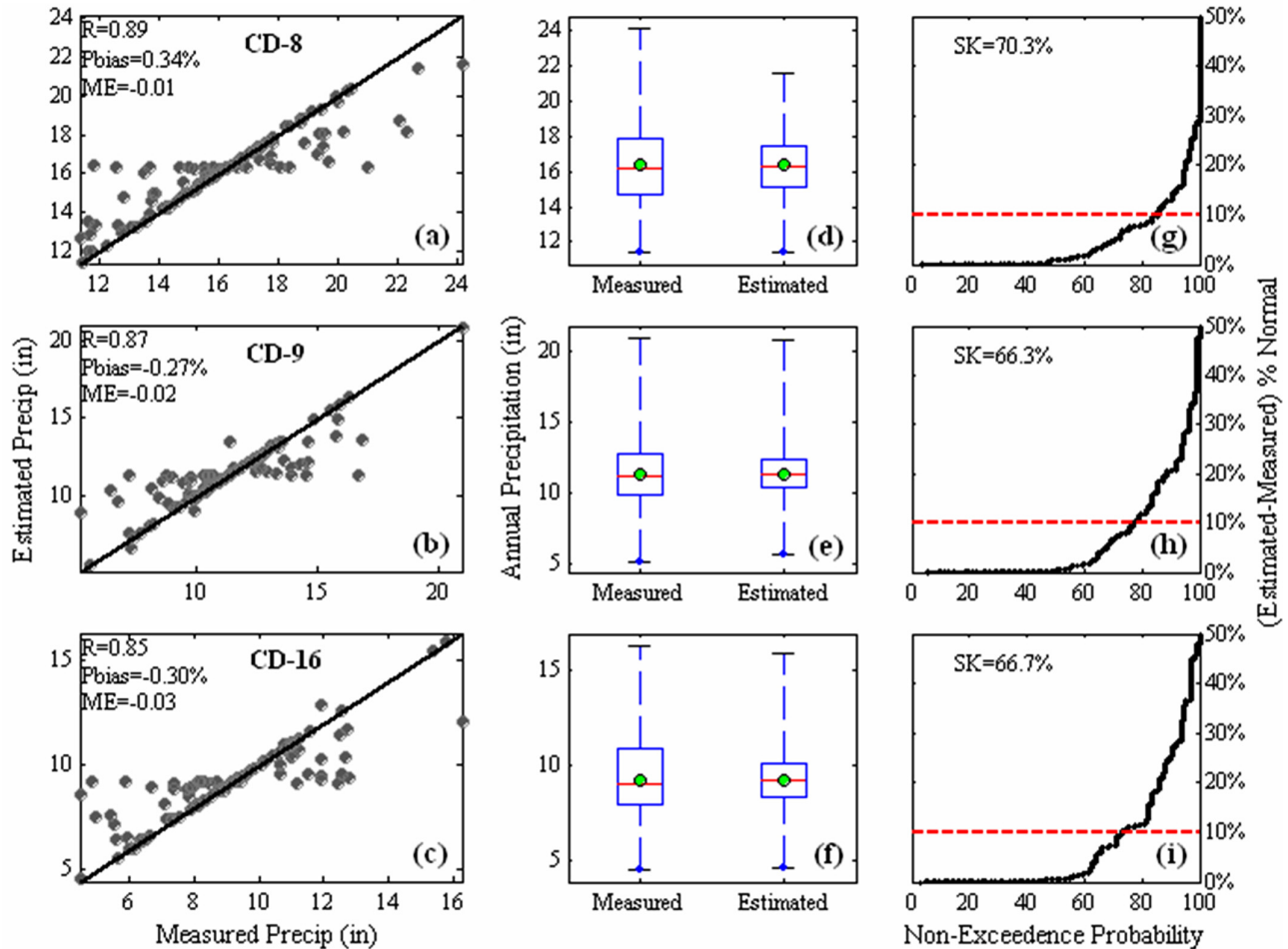
Figure 7. (a–c) scatterplot, (d–f) box plot, and (g–i) nonexceedance probability plots between the measured and estimated precipitation for selected climate divisions in the UCRB using PDO, NAO and AMO oscillation indices for Model II. The horizontal dotted line in the nonexceedance probability plots shows an error value of 10%.

[64] The variability (Figures 7d–7f) and probabilistic error (Figures 7g–7i) of the estimated precipitation values, compared to the measured values, also were analyzed. Figures 7d–7f shows the box plot between measured and estimated annual precipitation for three climate divisions: 8, 9, and 16. The span of the box represents interquartile range (25th–75th percentile) with horizontal line inside the box indicating the median (50th percentile) value. The whiskers extend from the box to 5th and 95th percentile on the lower and upper side, respectively. It was noticed that the model was able to capture the extreme (high and low values) for Climate Divisions 9 (Figure 7e) and 16 (Figure 7f), but missed the high values for Climate Division 8 (Figure 7d). The model performed reasonably well in representing the variability exhibited by annual precipitation. This was evident by the similarity in the interquartile range of the measured and estimated precipitation values. Overall, the model performed satisfactorily in capturing the low values compared to the high values.

[65] Probabilistic cumulative error between the measured and estimated annual precipitation was computed for the selected divisions in the Upper Basin (Figures 7g–7i). It

was noticed that for approximately 60% of the predictions, a negligible error—close to 1%—is achieved. Approximately 80% predictions have a 10% error. Based on past research that addressed difficulties in estimating precipitation [Bell, 1987; Olsson, 1998; Guenni and Bardossy, 2002], an error of less than 10% for approximately 80% of the estimates shows the robustness of the SVM approach in capturing the variability in annual precipitation in relation to large-scale climate patterns. Furthermore, it was seen that a SK value of greater than 10% was achieved for the selected divisions, indicating "skillful" forecasts. Overall, the model performed better in capturing the low values compared to the high values.

[66] In Model III, oscillations were dropped in pairs, and the remaining two oscillation modes were used to predict annual precipitation. The three performance measures are shown on a spatial map in Figure 8. Based on RSR (Figure 8a), R (Figure 8b), and NSE (Figure 8c) error statistics, three climate divisions in the Upper Basin—14, 15, and 17—showed improvements compared to the base case when using a combination of NAO and ENSO indices. These divisions accounted for approximately 21% of the flow in
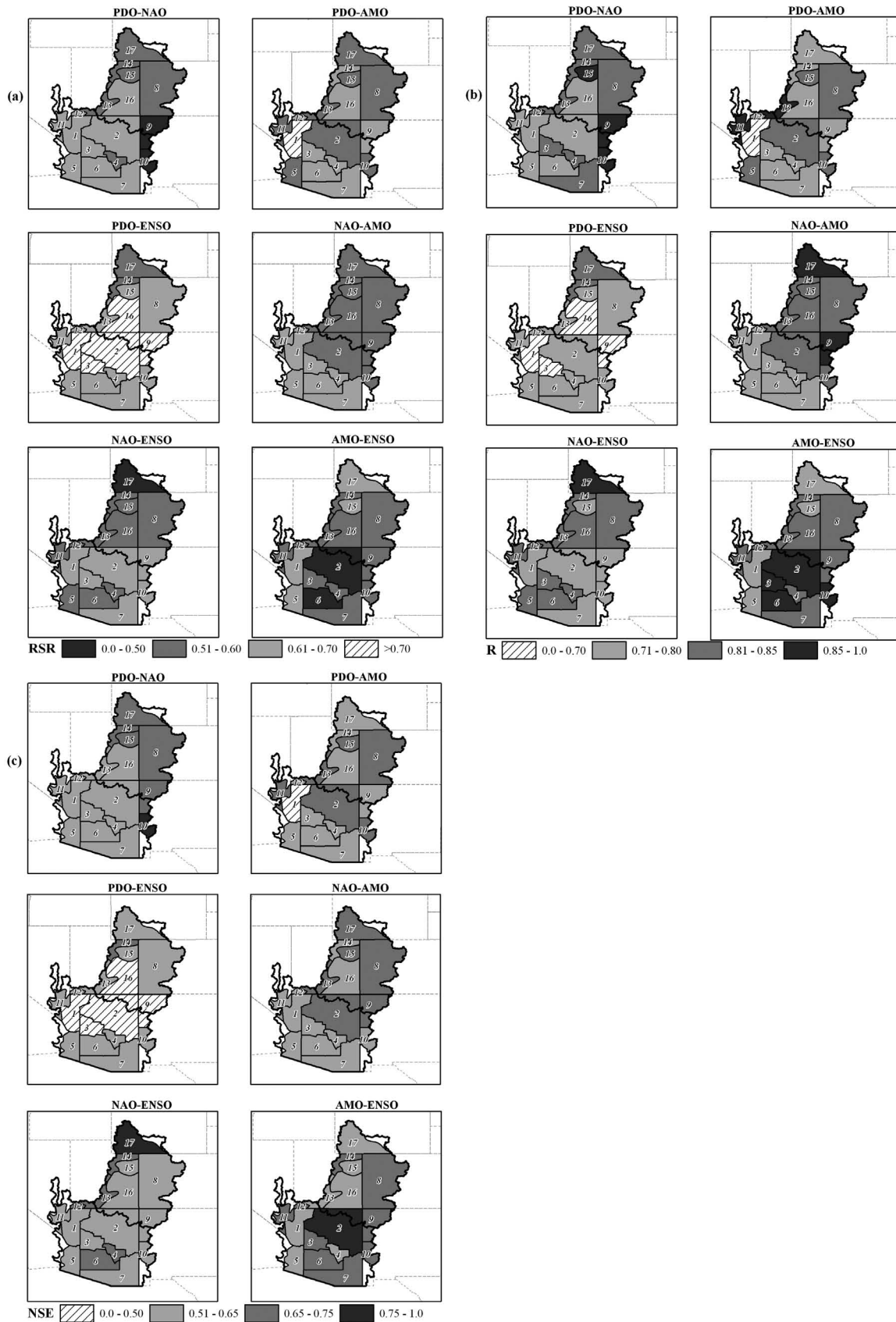
Figure 8. Spatial maps showing the range of performance measures for 17 climate divisions for Model III: (a) RSR, (b) R, and (c) NSE.

the Colorado River. The estimates were in the range of "good" to "very good."

[67] Additionally, coupling NAO and AMO also showed improvement for five climate divisions in the Upper Basin, compared to Model I results. These divisions generate approximately 27% of the flow in the Upper Basin. Overall, coupling NAO and AMO resulted in "good" precipitation estimates for the majority of the Upper Basin climate divisions. For the majority of the Upper Basin climate divisions, all other combinations resulted in deterioration of error statistics compared to Model I results. In indicating that NAO has a stronger presence in the Upper Basin, Model III results agreed with Model II results. This was evident by NAO being one of the inputs in the best Model III estimates for the Upper Basin, based on the three error statistics. The best predictions were obtained using a combination of NAO and AMO, whereas combination of PDO and ENSO resulted in worst predictions for majority of the upper basin climate divisions (Table 5). Overall, none of the combinations resulted in better predictions for the majority of the climate divisions, compared to Model II best results using PDO-NAO-AMO for the Upper Basin.

[68] All the three error statistics were in agreement in indicating that best precipitation estimates for the Lower

Basin climate divisions are obtained by using a combination of AMO and ENSO, compared to Model I results. All three performance measures showed improvement in the climate divisions covering the north central portion of LCRB, which region generates 70% of the flow in the lower basin. The estimates for the majority of the climate divisions were in the range of "very good," compared to "satisfactory" estimates for Model I and "good" estimates for the best combination of Model II. This implies that a combination of AMO and ENSO had statistically a stronger influence on precipitation in the LCRB compared to any other combination of indices. Additionally, it was noticed that a combination of PDO and ENSO resulted in unsatisfactory estimates for the majority of the Lower Basin divisions, as compared to the Upper Basin (Table 5).

[69] The visual inspection of results of the best Model III precipitation estimates using AMO and ENSO for the Lower Basin are shown in Figure 9. The divisions selected to show the results are Climate Divisions 2, 3, and 6. The scatterplot shows a match between measured and estimated precipitation for the selected climate divisions (Figures 9a–9c). The majority of the points for the selected divisions follow the 45° bisector, indicating good model fit. The sample points are aligned vertically along the diagonal, suggesting different
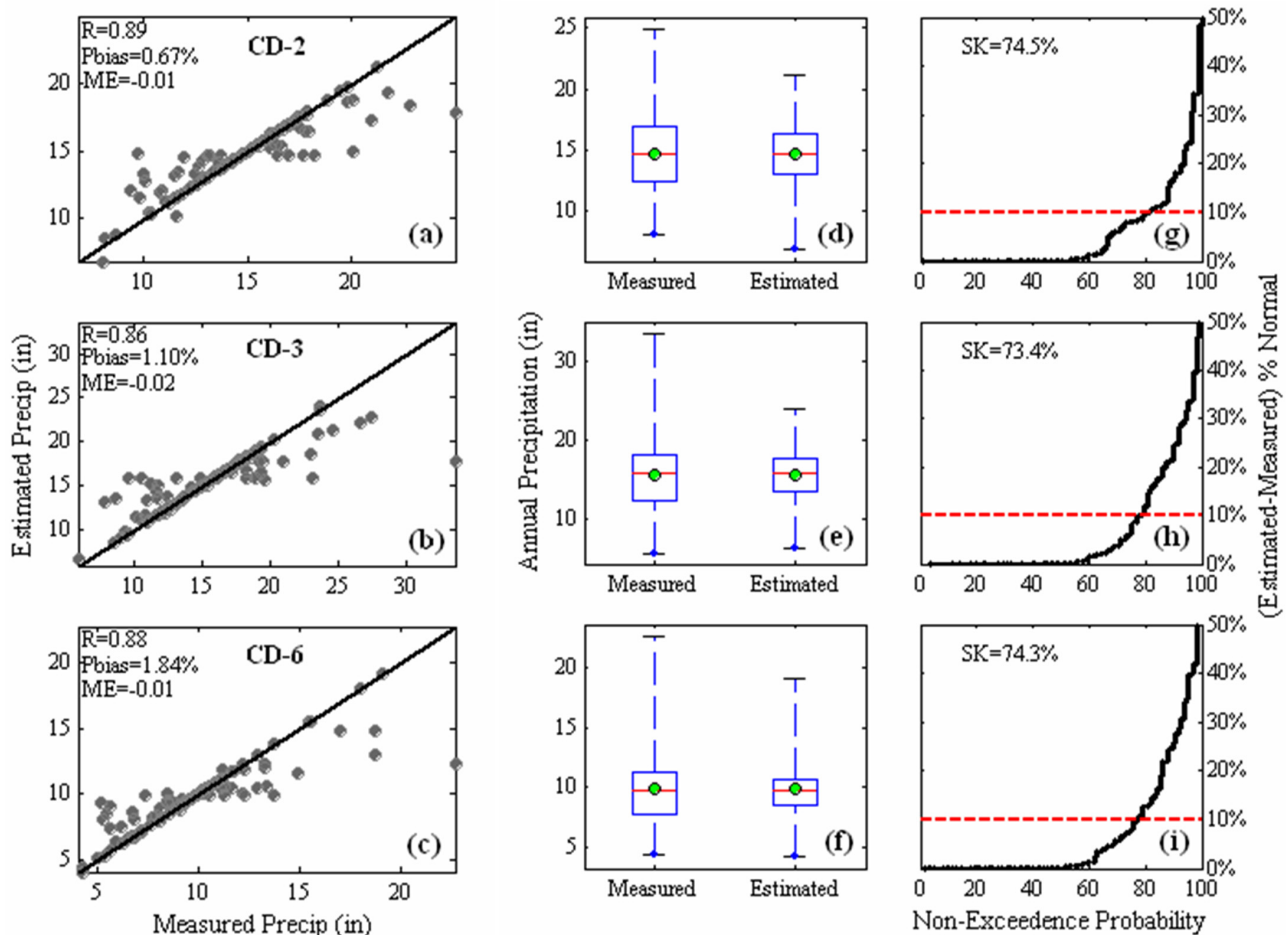

Figure 9. (a–c) scatterplot, (d–f) box plot, and (g–i) nonexceedance probability plots between the measured and estimated precipitation for selected climate divisions in the LCRB using AMO and ENSO oscillation indices for Model III.

estimated precipitation values for different years, which have different measured precipitation values. Higher correlation is achieved between measured and estimated precipitation. A positive Pbias ranging from 0.67% to 1.84% and smaller ME is achieved for the selected divisions indicating a good match between the measured and estimated precipitation.

[70] The model is able to perform well for low values, as the 5th percentile whiskers of estimated values match the measured values better compared to such high values as the 95th percentile (Figures 9d–9f). The model has negligible error for approximately 60% of the predictions (Figures 9g–9i); approximately 80% of the predictions have an error close to 10%. This is similar to the Upper Basin results and shows the robustness of the modified SVM modeling approach in capturing the variability in precipitation in relation to oceanic-atmospheric oscillations.

[71] In Model IV, each oscillation was used individually to estimate precipitation for each climate division. Figure 10 shows the spatial map representing the three performance measures, i.e., RSR (Figure 10a), R (Figure 10b), and NSE (Figure 10c). A satisfactory correlation (Figure 10b) between measured and estimated precipitation was achieved for majority of the Upper Basin climate divisions and also for a

few Lower Basin divisions using AMO as the sole input in the model. It is noticed that none of the oscillations, when used individually, resulted in improved precipitation estimates for CRB, as compared to Model I results. This is in agreement with findings from previous studies, where researchers showed that a qualitative understanding of the relationship between oscillations and precipitation within the CRB could be enhanced by evaluating the coupled response of oscillation indices rather than by using an individual oscillation mode [*Kim et al.*, 2006, 2008; *Canon et al.*, 2007; *Hidalgo and Dracup*, 2003; *McCabe et al.*, 2007].

[72] To evaluate and show the robustness of the proposed SVM approach, a bootstrap cross validation technique was used. Input data was resampled 100 times, and for each sample the SVM model was fitted and then tested on the entire data set. The final estimated value was the mean of these 100 ensembles. Figure 11 shows the time series plot between the measured, estimated and bootstrapped precipitation for the selected Upper Basin (Figure 11a) and Lower Basin (Figure 11b) climate divisions for the best model combinations. The time series plots show that estimated and boot strapped precipitation values were close to the measured values for the selected climate divisions. The

Figure 10. Spatial maps showing the range of performance measures for 17 climate divisions for Model IV: (a) RSR, (b) R, and (c) NSE.
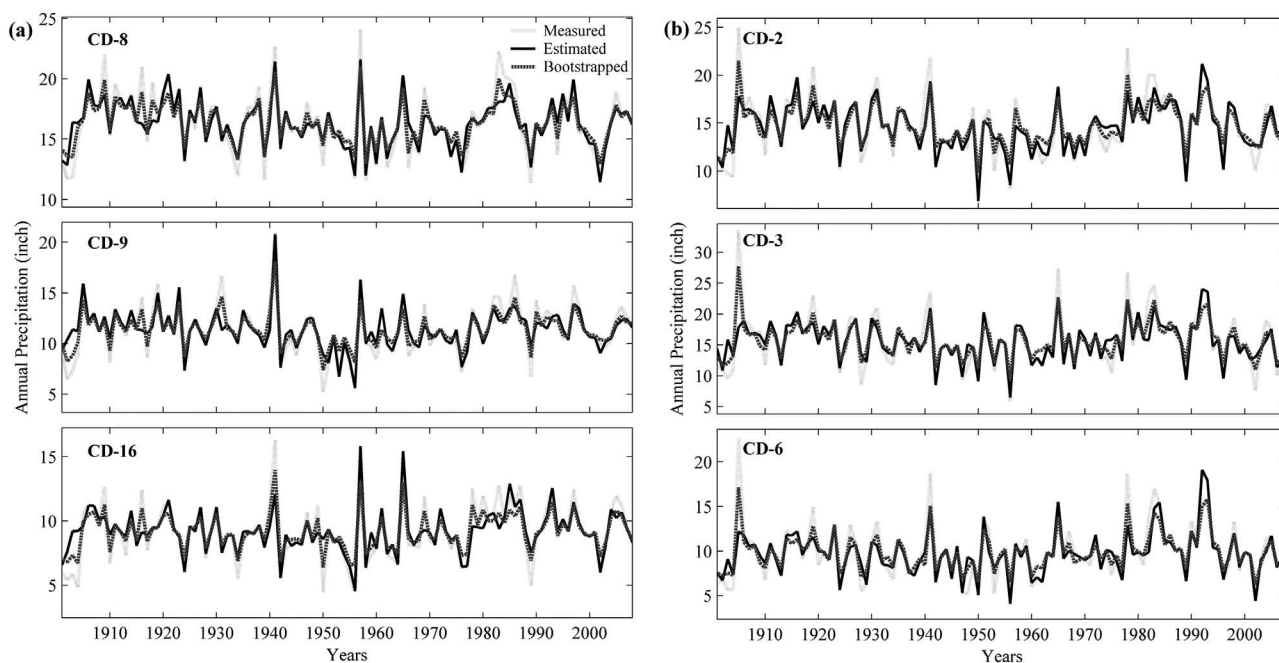
Figure 11. Precipitation time series plot for selected climate divisions in the (a) UCRB (using PDO, NAO, AMO indices) and (b) LCRB (using AMO and ENSO indices) showing the measured, estimated by SVM using actual data, and bootstrapped data.

estimated and bootstrapped precipitation followed a trend similar to the measured precipitation. It is also evident that the model missed the extremes in the early part of the 20th century (around 1910) and later for few years between 1970 and 1990. Additionally, it was seen that the model performed adequately in capturing high and low precipitation values during the severe sustained drought within the CRB, occurring from 2000–2008. Both the estimated and bootstrapped values were in agreement in indicating that the model does fairly well in capturing the low values compared to the high values.

[73] The cross validation results highlighted the robustness of proposed SVM framework. The results showed that the model is stable and findings are not altered using different training and testing data sets. The estimated and bootstrapped time series have a close match and exhibit similar pattern.

[74] Overall, based on Model II–IV results and bootstrapping cross validation analysis, it can be inferred that a combination of PDO, NAO, and AMO (Model II) has statistically the strongest association with the annual precipitation for a 1 year lead time for the majority of the UCRB. Interdecadal hydroclimatic variations in the UCRB that are related to possible PDO influences have been investigated by *Hidalgo and Dracup* [2003]. Their study indicated that the shifts in the mean of precipitation and streamflow within UCRB are related with decadal PDO changes. Similar to our findings, *Hidalgo and Dracup* [2003] concluded that ENSO associations are not always consistent, and may not be linked with the hydrologic fluctuations in UCRB. On the other hand, documented literature has shown the linkages between NAO and precipitation over the Europe and the Mediterranean basin [*Hurrell*, 1995; *Qian et al.*, 2000]; however, little attention has been given to NAO's

association with precipitation in the western United States, particularly to the CRB [*Kim et al.*, 2008]. Our results indicate that NAO, coupled with other indices, can improve the precipitation estimates in the UCRB. Furthermore, *Webb et al.* [2004] indicated that a combination of AMO and PDO may help to explain more thoroughly the long-term fluctuations in streamflow within the Colorado River Basin. AMO usually reflects the conditions in the Atlantic Ocean that may affect the climate in North America [*Enfield et al.*, 2001]. The findings of the current study are in partial agreement with the other studies that showed that PDO and AMO in combination with other indices can serve as useful predictors, to some extent, for extending lead times of different hydroclimatic variables—in this case, precipitation within the Upper Basin [*Hidalgo and Dracup*, 2003; *Webb et al.*, 2004; *Kim et al.*, 2006, 2008; *McCabe et al.*, 2007].

[75] In the case of the Lower Basin, best estimates are obtained using the combination of AMO and ENSO (Model III). This finding is in agreement with the available literature, indicating that ENSO effects are more pronounced in the Lower Basin than in the Upper Basin [*Redmond and Koch*, 1991; *Webb and Betancourt*, 1992; *Kahya and Dracup*, 1993; *Piechota and Dracup*, 1996]. ENSO has been linked to the occurrence of floods in the Lower Basin [*Webb and Betancourt*, 1992]. Additionally, *Thomas* [2007] identified that AMO—both individually and in combination with PDO and ENSO—can explain the streamflow variability in the Lower Basin. Furthermore, *Ellis et al.* [2010] indicated that much of drought variance in the Lower Basin can be explained using AMO compared to other indices. The dominance of AMO over PDO in influencing CRB droughts have become more in phase during the later half of 20th century. The three teleconnections, AMO, PDO, and ENSO, can be used to explain much of

the hydrologic variability within the Lower Basin compared to the Upper Basin [*Ellis et al.*, 2010]. This is in partial agreement with the current findings. The physical significance of the combined effect of AMO and ENSO on the hydrologic conditions in the Lower Basin is yet to be explored.

[76] It is evident from the literature that various features of climate are interrelated in a complicated fashion [*Webb et al.*, 2004; *McCabe et al.*, 2007]. Their interrelationship depends on the variability both in space and time. There may be additional, yet unidentified, factors that may contribute to the interrelationship of climate patterns [*California Dep. of Water Resources*, 2005; *Webb et al.*, 2004; *McCabe et al.*, 2007]. Though many studies have demonstrated that ENSO and PDO are mainly teleconnneted with monthly, seasonal, and annual precipitation variability in the U.S., it still is necessary to investigate the influence of other climate indices [*Ropelewski and Halpert*, 1986; *Ropelewski and Jones*, 1987; *Redmond and Koch*, 1991; *Kahya and Dracup*, 1993; *Webb et al.*, 2004].

6.3. Comparison of SVM With ANN and MLR Models

[77] The SVM model results also were compared with ANN and MLR model results. The scatter plots between measured and estimated precipitation for ANN and MLR models, using all four oscillations indices, are shown in Figure 12. Table 6 shows the comparison of different performance measures between the measured and estimated precipitation for ANN and MLR. The scatter plots for ANN

(Figure 12a) and MLR (Figure 12b) models show that for the majority of the climate divisions, the model estimates the mean of the measured data. Low correlation values were obtained for all climate divisions, indicating the inability of the model to capture the extremes. In general, the model estimates are parallel to the *x* axis instead of following the bisector line, showing poor prediction capability. The LEPS score SK was less than 10%, for majority of the climate divisions indicating an unsatisfactory forecast. This implies that model estimates are worse than the "climatology".

[78] The performance measures obtained from ANN and MLR are in the unsatisfactory range (Table 6). Although the SVM model outperforms both the ANN and MLR models, all three models perform comparatively better for the climate divisions within the Upper Basin compared to the Lower Basin divisions. Similar to Model I, Model II–IV (results not shown) was also created, and the results showed a better performance of SVM over both the ANN and MLR models.

[79] Based on the results, it was noticed that SVM outperformed both MLR and ANN models. There has been sufficient evidence from other studies in different fields of hydrology that show the superiority of SVM over the regular ANN and MLR modeling approaches [*Lin et al.*, 2009, 2010; *Ahmad et al.*, 2010; *Kalra and Ahmad*, 2009; *Gill et al.*, 2006; *Asefa et al.*, 2006; *Dibike et al.*, 2001]. This is because the SVM model has a better ability to generalize, relating the input to the desired output. In addition, the
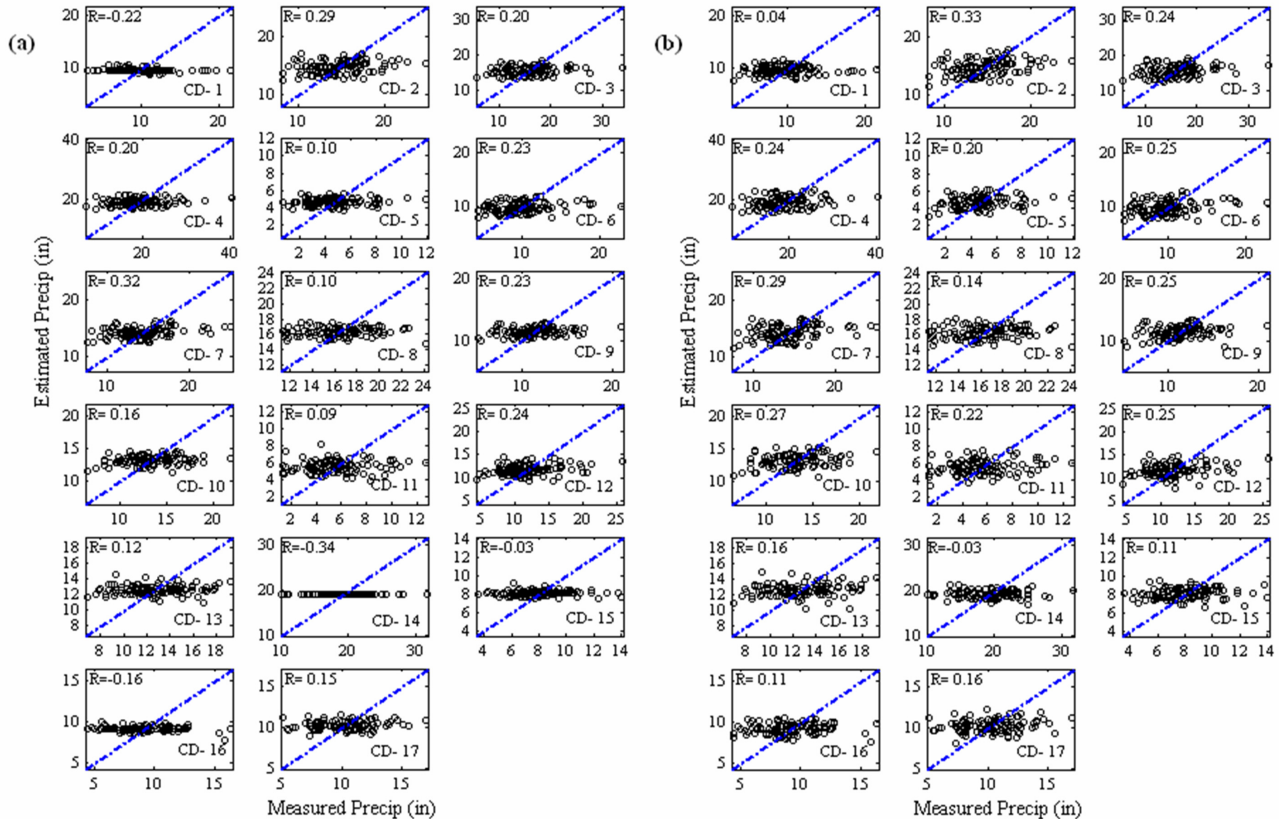

Figure 12. Scatterplot between measured and estimated precipitation for (a) ANN and (b) MLR model for 17 climate divisions for Model I. Dashed diagonal line is the 45° bisector.

Table 6. Comparison of Performance Measures for ANN and MLR Outputs for Model I[a]

Climate Division	ANN							MLR						
	RMSE	ME	RSR	R	NSE	Pbias	SK	RMSE	ME	RSR	R	NSE	Pbias	SK
1	3.40	−0.13	1.02	−0.22	−0.05	−0.18	2.4	3.39	−0.12	1.02	0.04	−0.04	0.04	1.3
2	3.09	−0.04	0.96	0.29	0.08	0.09	10.7	3.05	−0.04	0.94	0.33	0.10	−0.02	13.7
3	4.39	−0.08	0.98	0.20	0.03	0.08	7.3	4.36	−0.08	0.97	0.24	0.04	−0.01	10.6
4	5.20	−0.08	0.97	0.20	0.04	0.22	4.4	5.19	−0.08	0.97	0.24	0.04	−0.01	9.0
5	1.95	−0.23	1.00	0.10	−0.01	0.41	0.9	1.93	−0.20	0.99	0.20	0.02	0.12	4.9
6	3.21	−0.10	0.97	0.23	0.05	0.28	7.1	3.21	−0.10	0.97	0.25	0.05	0.03	9.1
7	3.06	−0.04	0.94	0.32	0.10	0.22	9.1	3.10	−0.04	0.96	0.29	0.07	−0.02	11.4
8	2.67	−0.03	1.00	0.10	0.00	−0.05	1.9	2.68	−0.03	1.00	0.14	−0.01	0.05	4.0
9	2.49	−0.05	0.97	0.23	0.05	−0.15	6.5	2.49	−0.05	0.97	0.25	0.05	0.00	9.8
10	2.69	−0.05	0.99	0.16	0.01	−0.12	2.8	2.62	−0.04	0.96	0.27	0.06	0.00	7.0
11	2.37	−0.25	1.01	0.09	−0.04	−0.12	1.6	2.29	−0.23	0.98	0.22	0.03	0.03	4.7
12	3.52	−0.09	0.97	0.24	0.05	0.13	4.9	3.53	−0.09	0.97	0.25	0.05	0.00	5.7
13	2.66	−0.04	1.00	0.12	0.00	0.19	3.4	2.66	−0.04	1.00	0.16	0.00	0.05	4.8
14	3.97	−0.05	1.00	−0.34	0.00	0.00	0.1	4.08	−0.05	1.02	−0.03	−0.05	0.04	0.0
15	2.06	−0.07	1.01	−0.03	−0.03	0.33	0.6	2.05	−0.07	1.00	0.11	−0.02	0.10	3.4
16	2.37	−0.07	1.02	−0.16	−0.05	0.11	0.0	2.33	−0.07	1.01	0.11	−0.02	0.04	5.0
17	2.28	−0.06	0.99	0.15	0.02	−0.28	3.5	2.29	−0.05	0.99	0.16	0.00	0.07	6.2

[a]The RMSE and ME values are in inches.

optimization algorithm used in SVM is more robust than the one used in regular ANN models. In case of MLR, the models are based on the assumption of normality, and can be used efficiently to relate simple processes. In case of hydrological processes where the data does not follow the usual normal distribution, MLR models fail to capture the variability. The feed-forward back-propagation algorithm used for ANN in the current analysis is simple and widely used. There are other ANN architectures and activation functions [see *Dibike et al.*, 1999] that may be able to better capture the relationship between precipitation and ocean-atmospheric indices. An exhaustive comparison of methods was not the focus of our work.

7. Summary and Conclusion

[80] In this study we explored the association between individual and coupled oceanic-atmospheric indices and precipitation in the Colorado River Basin. We used an AI-type model to capture the relationship between oceanic-atmospheric indices and precipitation, and used this model to extend the lead time for precipitation estimation up to 1 year. For this purpose, a modified Support Vector Machine predictive framework incorporating oceanic-atmospheric oscillations was constructed for the 17 climate divisions encompassing the Colorado River Basin. The oceanic-atmospheric oscillations used in this study were PDO, NAO, AMO, and ENSO.

[81] Annual precipitation within CRB is variable both at the temporal and spatial scales. It is difficult to construct a single precipitation time series that is representative of the entire basin. For this reason, monthly time series data for the climate divisions were used that extend more than a century in record and spatially cover the entire basin. Currently, CPC issues 3 month forecast for lead times of 0.5 to 12.5 months with modest skill for 3–9 month lead time based on ENSO and its indices. Forecasts are termed "skillful" if they show improvement over the long-term averages of the precipitation record used in the analysis. In general, they have no skill for summer precipitation during ENSO years and no skill for winter precipitation during non-ENSO

years [*California Dep. of Water Resources*, 2005; *Regonda et al.*, 2006]. During ENSO years, the precipitation forecast is higher in the southern part of the basin (LCRB) and has no skill in the headwaters that generate majority of the runoff in the Colorado River [*Redmond and Koch*, 1991]. Therefore, obtaining accurate estimates of precipitation within the CRB is a formidable challenge. However, the advancement along several scientific fronts has opened doors for statistical forecast possibilities. In an attempt to address this challenge, we evaluated the link between individual and coupled oceanic-atmospheric indices and temporal variability in precipitation, and developed a data driven model to estimate annual precipitation with a lead time of 1 year. It should be noted that the precipitation estimates obtained in the current study are not compared with the CPC estimates as both have different temporal resolution. The current work contributes to the existing literature on the use of statistical approaches for estimating precipitation.

[82] Multiple SVM models incorporating individual and coupled oceanic-atmospheric oscillations were developed. Results indicate that coupled PDO, NAO, and AMO have a statistically stronger association with precipitation in the Upper Basin. This combination resulted in very good precipitation estimates for climate divisions that contribute approximately 67% (two-thirds) of the flow in Colorado River. Coupled AMO and ENSO have a statistically stronger association with precipitation in majority of the Lower Basin. The estimates for the majority of the Lower Basin divisions were in the range of "very good;" these divisions account for generating approximately 70% of flow in the Lower Basin. The findings of the SVM approach were confirmed using a bootstrapped cross validation approach. The cross validation analysis indicated that the results are stable and do not change with different training and testing data sets. Overall, the results showed that the SVM approach does better in capturing the low precipitation values compared to the higher values. The inability of SVM model in estimating some high precipitation values may be attributed to fitting of outliers in the training phase. This drawback associated with SVM under the condition of non-Gaussian

outliers has been indicated in prior studies [*Asefa et al.; 2006; Gill et al.*, 2006; *Twarakavi et al.*, 2006]. Therefore, removing the outliers may result in improving the model performance. Also, there are other SVM machines such as the simple Least Square Support Vector Machine (LS-SVM) or robust LS-SVM that have shown to improve model performance in case of noisy data and outliers [*Suykens*, 2001; *Twarakavi et al.*, 2006]. The results also indicated that the SVM approach performed better in capturing the interaction of oscillation indices and precipitation when compared with ANN and MLR.

[83] The major contributions of this research are as follows. First, there is no single oscillation that can be used to explain the climate variability within CRB. It is evident that various oscillations are interrelated and can be used in combination to improve annual precipitation forecasting with a 1 year lead time within CRB [*Webb et al.*, 2004; *Kim et al.*, 2008]. Second, NAO coupled with other indices can improve the precipitation estimates in the UCRB. This requires further attention and should be investigated. Third, the modified SVM predictive framework showed statistically significant results; conclusions drawn from this analysis are cross validated and are not specific to any particular period. The approach uses the entire data and may be used in situation when limited data is available. Fourth, the analysis performed is comprehensive in nature and spatially covers the entire CRB. There have been attempts to estimate precipitation within CRB, but no other study has attempted to utilize a data driven model coupled with major Pacific and Atlantic Ocean climate patterns to estimate annual precipitation for the entire Colorado River Basin.

[84] There are also some limitations associated with the current work. The study used climate division precipitation data and does not differentiate between precipitation as rainfall and snowfall. Second, the current research used the four most common climate indices; there are other indices, such as IPO, PNA, and predefined SST regions [*Kalra et al.*, 2012] that potentially can be used to improve the results. Although, the proposed modeling framework does not suffer from the so-called "curse-of-dimensionality," if more predictors and a longer data set are involved, the trade-off in computation time and accuracy will need to be considered.

[85] The current study was able to successfully estimate precipitation using SVM approach, there is still some unexplained variability, which cannot be addressed using the statistical approach. These variations were seen with a particular predictor combination resulting in different hydrological response among adjacent climate divisions. For example, a combination of NAO and ENSO had stronger associations with Climate Divisions 14 and 17 precipitation but not for Climate Divisions 8 and 15. The cause for this variability may be somewhat explained through Table 3, which highlighted that climate indices are correlated differently with the CRB precipitation. The indices individually show different hydrologic response to climate divisions in higher elevations compared to lower elevation. Also, this response varies among regions having snowfall as their primary precipitation compared to rainfall. Besides climate other topographic features can also alter this response. Furthermore, when coupled response of indices is evaluated in

relations to any hydrologic variable, understanding the dynamics becomes even more challenging. Therefore, it should be noted that the results presented in this study are statistical in nature, and the physical mechanisms that drive these relationships are not fully understood at this time.

[86] The results from the current research help in statistically understanding the association between different oceanic-atmospheric indices and precipitation in the Upper and Lower CRB. Using the modified SVM predictive framework, cumulative precipitation totals for the current year can be made available as early as 1 January of that year. The annual precipitation values can be disaggregated into seasonal or monthly resolution depending on the need of end user [*Lall and Sharma*, 1996; *Lall et al.*, 1996; *Sharma et al.*, 1997; *Rajagopalan and Lall*, 1999; *Ahmad and Prashar*, 2010; *Kalra and Ahmad*, 2011] for other applications. The predictive framework incorporates global climate information therefore can be potentially used in other catchments and river basins. However, the predictive potential of model will vary depending on the strength of connections between input climate indices and precipitation in the region. Besides precipitation, the proposed SVM framework may be applied to estimate other hydrological variables such as streamflow, groundwater levels, and soil moisture.

[87] Water managers have considered past climate variability over different time scales for water resources planning and management [*Vedwan et al.*, 2008]. With the projections of future climate being uncertain, multiple plausible options need to be considered. Simple, robust, and parsimonious statistical techniques can serve as good predictor of hydroclimatology. Overall, the SVM model used in this study provides very good precipitation estimates that have the potential to improve water management within the basin.

References

Ahmad, S., and D. Prashar (2010), Evaluating municipal water conservation policies using a dynamic simulation model, *Water Resour. Manage.*, 24(13), 3371–3395.

Ahmad, S., and S. P. Simonovic (2000), System dynamics modeling of reservoir operations for flood management, *ASCE J. Comput. Civ. Eng.*, 14(3), 190–198.

Ahmad, S., and S. P. Simonovic (2001), Integration of heuristic knowledge with analytical tools for selection of flood control measures, *Can. J. Civ. Eng.*, 28(2), 208–221.

Ahmad, S., and S. P. Simonovic (2004), Spatial system dynamics: A new approach for simulation of water resources systems, *ASCE J. Comput. Civ. Eng.*, 18(4), 331–340.

Ahmad, S., and S. P. Simonovic (2005), An artificial neural network model for generating hydrograph from hydro-meteorological parameters, *J. Hydrol.*, 315(1–4), 236–251.

Ahmad, S., and S. P. Simonovic (2006), An intelligent decision support system for management of floods, *Water Resour. Manage.*, 20(3), 391–410.

Ahmad, S., A. Kalra, and H. Stephen (2010), Estimating soil moisture using remote sensing data: A machine learning approach, *Adv. Water Resour.*, 33(1), 69–80.

ASCE Task Committee (2000a), Artificial neural networks in hydrology I: Preliminary concepts, *J. Hydrol. Eng.*, 5(2), 115–123.

ASCE Task Committee (2000b), Artificial neural networks in hydrology II: Hydrologic applications, *J. Hydrol. Eng.*, 5(2), 124–137.

Asefa, T., M. Kemblowski, M. McKee, and A. Khalil (2006), Multi-time scale stream flow predictions: The support vector machines approach, *J. Hydrol.*, *318*, 7–16.

Ashok, K., Z. Guan, and T. Yamagata (2001), Impact of the Indian Ocean dipole on the relationship between the Indian monsoon rainfall and ENSO, *Geophys. Res. Lett.*, *28*(23), 4499–4502, doi:10.1029/2001GL013294.

Ashok, K., Z. Guan, and T. Yamagata (2003), Influence of the Indian Ocean Dipole on the Australian winter rainfall, *Geophys. Res. Lett.*, *30*(15), 1821, doi:10.1029/2003GL017926.

Ashok, K., S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata (2007), El Niño Modoki and its possible teleconnection, *J. Geophys. Res.*, *112*, C11007, doi:10.1029/2006JC003798.

Barlow, M., H. Cullen, and B. Lyon (2002), Drought in central and southwest Asia: La Nina, the warm pool, and Indian Ocean precipitation, *J. Clim.*, *15*(7), 697–700.

Bell, T. L. (1987), A space-time stochastic model of rainfall for satellite remote-sensing studies, *Geophys. Res. Lett.*, *92*(D8), 9631–9643.

Bjerknes, J. (1966), A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature, *Tellus*, *18*, 820–829.

Boser, B., I. Guyon, and V. Vapnik (1992), A training algorithm for optimal margin classifiers, in *Annual Conference on Computational Learning Theory*, pp. 144–152, ACM Press, Pittsburgh, Pa.

Brito-Castillo, L., A. Leyva-Contreras, A. V. Douglas, and D. Lluch-Belda (2002), Pacific decadal oscillation and the filled capacity of dams on the rivers of the Gulf of California continental watershed, *Atmosfera*, *15*, 121–138.

California Dep. of Water Resources (2005), *Colorado River Basin Climate: Paleo, Present, Future*, Spec. Publ. of Assoc. Calif. Water Agencies and Colorado River Water Users Assoc. Conf., Calif. Dep. Water Res., Sacramento, Calif., 80 pp.

Cancelliere, A., D. Mauro, B. Bonaccorso, and G. Rossi (2007), Investigating the potential of NAO index to forecast droughts in Sicily, in *Hydrology Days*, edited by J. A. Ramirez, pp. 75–86, Colorado State Univ., Fort Collins, Colo.

Canon, J., J. Gonzalez, and J. Valdes (2007), Precipitation in the Colorado River Basin and its low frequency associations with PDO and ENSO signals, *J. Hydrol.*, *333*, 252–264.

Carrier, C., A. Kalra, and S. Ahmad (2011), Using proxy reconstructions for streamflow forecasting, in *World Environmental and Water Resources Congress: Bearing Knowledge for Sustainability*, edited by R. E. Beighley II and M. W. Killgore, pp. 3124–3133, Palm Springs, Calif.

Cayan, D. R., M. D. Dettinger, H. F. Diaz, and N. E. Graham (1998), Decadal climate variability of precipitation over western North America, *J. Clim.*, *11*(12), 3148–3166.

Cayan, D. R., K. T. Redmond, and L. G. Riddle (1999), ENSO and hydrologic extremes in the western United States, *J. Clim.*, *12*, 2881–2893.

Chowdhury, S., and A. Sharma (2009), Multisite seasonal forecast of arid river flows using a dynamic model combination approach, *Water Resour. Res.*, *45*, W10428, doi:10.1029/2008WR007510.

Colle, B. A. (2004), Sensitivity of Orographic precipitation to changing ambient conditions and terrain geometries: An idealized modeling perspective, *J. Atmos. Sci.*, *61*, 588–606.

Coulibaly, P. (2006), Spatial and temporal variability of Canadian seasonal precipitation (1900–2000), *Adv. Water Resour.*, *29*(12), 1846–1865.

Coulibaly, P., F. Anctil, P. Rasmussen, and B. Bobee (2000), A recurrent neural networks approach using indices of low-frequency climatic variability to forecast regional annual runoff, *Hydrol. Processes*, *14*, 2755–2777.

Dawadi, S., and S. Ahmad (2012), Changing Climatic Conditions in the Colorado River Basin: Implications for Water Resources Management, *J. Hydrol.*, *430–431*, 127–141, doi:10.1016/j.jhydrol.2012.02.010.

Diaz, H. F., and G. N. Kiladis (1992), Atmospheric teleconnection associated with the extreme phase of the Southern Oscillation, in *El Nino: Historical and Paleoclimatic Aspects of the Southern Oscillation*, edited by H. F. D. A. V. Markgraf, pp. 7–28, Cambridge Univ. Press, New York.

Diaz, H. F., M. P. Hoerling, and J. K. Eischeid (2001), ENSO variability, teleconnections and climate change, *Int. J. Climatol.*, *21*, 1845–1862.

Dibike, Y. B., D. Solomatine, and M. B. Abbott (1999), On the encapsulation of numerical hydraulic models in artificial neural network, *J. Hydraul. Res.*, *37*(2), 147–161.

Dibike, Y. B., S. Velickov, D. Solomatine, and M. B. Abbott (2001), Model induction with support vector machines: Introduction and application, *J. Comput. Civ. Eng.*, *15*(3), 208–216.

Dickson, R. R., T. J. Osborn, J. W. Hurrell, J. Meincke, J. Blindheim, B. Adlandsvik, T. Vinje, G. Alekseev, and W. Maslowski (2000), The Arctic Ocean response to the North Atlantic oscillation, *J. Clim.*, *13*, 2671–2696.

Dracup, J. A., and E. Kahya (1994), The relationship between US streamflow and La Nina, *Water Resour. Res.*, *30*(7), 2133–2141.

Efron, B. (1979), Bootstrap methods: Another look at the jacknife, *Ann. Stat.*, *7*, 1–26.

Ellis, A. W., G. B. Goodrich, and G. M. Garfin (2010), A hydroclimatic index for examining patterns of drought in the Colorado River Basin, *Int. J. Climatol.*, *30*, 236–255.

Enfield, D. B., A. M. Mestas-Nunez, and P. J. Trimble (2001), The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S., *Geophys. Res. Lett.*, *28*, 2077–2080.

Feng, S., and Q. Hu (2007), Changes in winter snowfall/precipitation ratio in the contiguous United States, *J. Geophys. Res.*, *112*, D15109, doi:10.1029/2007JD008397.

Forsee, W. J., and S. Ahmad (2011), Evaluating urban stormwater infrastructure design in response to projected climate change, *J. Hydrol. Eng.*, *16*(11), 865–873, doi:10.1061/ASCE-HE.1943-5584.0000383.

Geisser, S. (1975), The predictive sample reuse method with applications, *J. Am. Stat. Assoc.*, *70*, 320–328.

Gershunov, A., and T. P. Barnett (1998), Interdecadal modulation of ENSO telecommunications, *Bull. Am. Meteorol. Soc.*, *79*, 2715–2726.

Giannini, A., M. A. Cane, and Y. Kushnir (2001), Interdecadal changes in the ENSO teleconnection to the Caribbean region and the North Atlantic oscillation, *J. Clim.*, *14*, 2867–2879.

Gill, M. K., T. Asefa, M. Kemblowski, and M. McKee (2006), Soil moisture prediction using support vector machines, *J. Am. Water Resour. Assoc.*, *42*(4), 1033–1046.

Grantz, K., B. Rajagopalan, M. P. Clark, and E. A. Zagona (2005), A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts, *Water Resour. Res.*, *41*, W10410, doi:10.1029/2004WR003467.

Guenni, L., and A. Bardossy (2002), A two steps disaggregation method for highly seasonal rainfall, *Stoch. Environ. Res. Risk Assess.*, *16*, 188–206.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1999), Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration, *J. Hydrol. Eng.*, *4*(2), 135–143.

Guttman, N. B., and R. G. Quayle (1996), A historical perspective of U.S. climate divisions, *Bull. Am. Meteorol. Soc.*, *77*, 293–303.

Gutzler, D. S., D. M. Kann, and C. Thornbrugh (2002), Modulation of ENSO-based long-lead outlooks of Southwestern U.S. winter precipitation by the pacific decadal oscillation, *Bull. Am. Meteorol. Soc.*, *17*, 1163–1172.

Guyon, I., B. Boser, and V. Vapnik (1993), Automatic capacity tuning of very large VC-dimension classifiers, in *Advances in Neural Information Processing Systems*, edited by S. J. Hanson, J. D. Cowan, and C. L. Giles, pp. 147–155, Morgan Kaufmann Publ., San Francisco, Calif.

Hamlet, A. F., P. W. Mote, M. P. Clark, and D. P. Lettenmaier (2005), Effects of temperature and precipitation variability on snowpack trends in the western United States, *Am. Meteorol. Soc.*, *18*, 4545–4561.

Haykin, S. (2003), *Neural Networks: A comprehensive foundation*, 4th edition, Pearson Educ., Singapore.

Hidalgo, H. G., and J. A. Dracup (2003), ENSO and PDO effects on hydroclimatic variations of the Upper Colorado River Basin, *J. Hydrometeorol.*, *4*, 5–23.

Higgins, R. W., W. A. Leetmaa, Y. Xue, and A. Barnston (2000), Dominant factors influencing the seasonal predictability of U.S. precipitation and surface air temperature, *J. Clim.*, *13*, 3994–4017.

Hsu, K.-L., H. V. Gupta, and S. Sorooshian (1995), Artificial neural network modeling of the rainfall-runoff process, *Water Resour. Res.*, *31*(10), 2517–2530.

Hsu, K. L., X. Gao, S. Sorooshian, and H. V. Gupta (1997), Precipitation estimation from remotely sensed information using artificial neural networks, *J. Appl. Meteorol.*, *36*, 1176–1190.

Hu, Q., and S. Feng (2001), Variations of teleconnection of ENSO and interannual variation in summer rainfall in the central United States, *J. Clim.*, *14*, 2469–2480.

Hurrell, J. W. (1995), Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation, *Science*, *269*(5224), 676–679.

Kahya, E., and J. A. Dracup (1993), U.S. streamflow patterns in relation to the El Nino/southern oscillation, *Water Resour. Res.*, *29*(8), 2491–2503, doi:10.1029/93WR00744.

Kalra, A. (2012), Association of oceanic-atmospheric oscillations and hydroclimatic variables in the Colorado River Basin, Ph.D. dissertation, 220 pp., Univ. of Nev., Las Vegas, Nev.

Kalra, A., and S. Ahmad (2009), Using oceanic-atmospheric oscillations for long lead time streamflow forecasting, *Water Resour. Res.*, *45*, W03413, doi:10.1029/2008WR006855.

Kalra, A., and S. Ahmad (2011), Evaluating changes and estimating seasonal precipitation for Colorado River Basin using stochastic nonparametric disaggregation technique, *Water Resour. Res.*, *47*, W05555, doi:10.1029/2010WR009118.

Kalra, A., W. P. Miller, K. W. Lamb, S. Ahmad, and T. Piechota (2012), Using large-scale climatic patterns for improving long lead time streamflow forecasts for Gunnison and San Juan River Basins, *Hydrol. Processes*, doi:10.1002/hyp.9236, in press.

Kane, R. P. (1999), El Nino timings and rainfall extremes in India, Southeast Asia and China, *Int. J. Climatol.*, *19*, 653–672.

Karl, T. R., and R. W. Knight (1997), Secular trends of precipitation amount, frequency, and intensity in the United States, *Bull. Am. Meteorol. Soc.*, *79*(2), 231–241.

Karl, T. R., C. N. Williams Jr., and P. J. Young (1986), A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperatures for United States, *J. Clim. Appl. Meteorol.*, *25*, 145–160.

Khalil, A., M. N. Almasri, M. McKee, and J. J. Kaluarachchi (2005), Applicability of statistical learning algorithms in groundwater quality modeling, *Water Resour. Res.*, *41*, W05010, doi:10.1029/2004WR003608.

Kiem, A. S., and S. W. Franks (2004), Multi-decadal variability of drought risk, eastern Australia, *Hydrol. Processes*, *18*(11), 2039–2050.

Kim, T.-W., J. B. Valdes, B. Nijssen, and D. Roncayolo (2006), Quantification of linkages between large-scale climate patterns and precipitation in the Colorado River Basin, *J. Hydrol.*, *321*, 173–186.

Kim, T.-W., C. Yao, and J.-H. Ahn (2008), Influence of climate variation on seasonal precipitation in the Colorado River Basin, *Stoch. Environ. Res. Risk Assess.*, *22*, 411–420.

Kuligowski, R. J., and A. P. Barros (1998), Experiments in short-term precipitation forecasting using artificial neural networks, *Mon. Weather Rev.*, *126*, 470–482.

Kushnir, Y., V. J. Cardone, J. G. Greenwood, and M. A. Cane (1997), The recent increase in the North Atlantic wave heights, *J. Clim.*, *10*, 2107–2113.

Lall, U., and A. Sharma (1996), A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resour. Res.*, *32*(3), 679–693.

Lall, U., B. Rajagopalan, and D. G. Tarboton (1996), A nonparametric wet/dry spell for resampling daily precipitation, *Water Resour. Res.*, *32*(9), 2803–2823.

Lau, K. M., and H. T. Wu (2001), Principal modes of rainfall-SST variability of the Asian summer monsoon: A reassessment of the Monsoon-ENSO relationship, *J. Clim.*, *14*, 2480–2895.

Legates, D. R., and G. J. McCabe (1999), Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.*, *35*(1), 233–241.

Li, P. H., H. H. Kwon, L. Sun, U. Lall, and J. J. Kao (2010), A modified support vector machine based prediction model on streamflow at the Shih-men Reservoir, Taiwan, *Int. J. Climatol.*, *30*, 1256–1268.

Lin, G.-F., G.-R. Chen, M.-C. Wu, and Y.-C. Chou (2009), Effective forecasting of hourly typhoon rainfall using support vector machines, *Water Resour. Res.*, *45*, W08440, doi:10.1029/2009WR007911.

Lin, G.-F., G.-R. Chen, and P.-Y. Huang (2010), Effective typhoon characteristics and their effects on hourly reservoir inflow forecasting, *Adv. Water Resour.*, *33*, 887–898.

Liong, S.-Y., and C. Sivapragasam (2002), Flood stage forecasting with support vector machines, *J. Am. Water Resour. Assoc.*, *38*(1), 173–186.

Mantua, N. J., Y. Z. Hare, J. M. Wallace, and R. C. Francis (1997), A pacific interdecadal climate oscillation with impacts on salmon production, *Bull. Am. Meteorol. Soc.*, *78*, 1069–1079.

McCabe, G. J., and M. D. Dettinger (1999), Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States, *Int. J. Climatol.*, *19*, 1399–1410.

McCabe, G. J., M. A. Palecki, and J. L. Betancourt (2004), Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States, *Proc. Natl. Acad. Sci.*, *101*(12), 4136–4141.

McCabe, G. J., J. L. Betancourt, and H. G. Hidalgo (2007), Associations of decadal to multidecadal sea-surface temperature variability with Upper Colorado river flow, *J. Am. Water Resour. Assoc.*, *43*(1), 183–192.

Melesse, A. M., S. Ahmad, M. E. McClain, X. Wang, and Y. H. Lim (2011), Suspended sediment load prediction of river systems: An artificial neural network approach, *Agric. Water Manage.*, *98*(5), 855–866.

Merideth, R. (2000), *A primer on climatic variability and change in the southwest*, 28 pp., Udall Cent. for Stud. in Public Policy and the Inst. of the Study of Plant Earth, Univ. of Ariz., Tucson, Ariz.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Trans. Am. Soc. Agric. Biol. Eng.*, *50*(3), 885–900.

Mosquera-Machado, S., and S. Ahmad (2007), Flood hazard assessment of Atrato river in Colombia, *Water Resour. Manage.*, *21*(3), 591–609.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models: Part 1. A discussion of principles, *J. Hydrol.*, *10*(3), 282–290.

Olsson, J. (1998), Evaluation of a scaling cascade model for temporal rainfall disaggregation, *Hydrol. Earth Syst. Sci.*, *2*(1), 19–30.

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

Potts, J. M., C. K. Folland, I. T. Jolliffe, and D. Sexton (1996), Revised "LEPS" scores for assessing climate model simulations and long-range forecasts, *J. Clim.*, *9*, 34–53.

Prairie, J., and R. Callejo (2005), *Natural Flow and Salt Computation Methods, Calendar Years 1971–1995*, U.S. Dep. Interior, Bur. of Reclam., Boulder City, Nev., 112 pp.

Pui, A., A. Lal, and A. Sharma (2011), How does the Interdecadal Pacific Oscillation affect design floods in Australia?, *Water Resour. Res.*, *47*, W05554, doi:10.1029/2010WR009420.

Pulwarty, R. S., and T. S. Melis (2001), Climate extremes and adaptive management on the Colorado River: Lessons from the 1997–1998 ENSO event, *J. Environ. Manage.*, *63*, 207–324.

Puri, S., H. Stephen, and S. Ahmad (2011), Relating TRMM precipitation radar land surface backscatter response to soil moisture in the southern United States, *J. Hydrol.*, *402*(1–2), 115–125.

Qaiser, K., S. Ahmad, W. Johnson, and J. Batista (2011), Evaluating the impact of water conservation on fate of outdoor water use: A study in an arid region, *J. Environ. Manage.*, *92*(8), 2061–2068.

Qian, B., J. Corte-real, and H. Xu (2000), Is the North Atlantic Oscillation the most important atmospheric pattern for precipitation in Europe, *Geophys. Res. Lett.*, *105*, 11,901–11,910.

Rajagopalan, B., and U. Lall (1998), Interannual variability in western US precipitation, *J. Hydrol.*, *210*, 51–67.

Rajagopalan, B., and U. Lall (1999), A k-nearest neighbor simulator for daily precipitation and other weather variables, *Water Resour. Res.*, *35*(10), 3089–3101.

Raman, H., and N. Sunilkumar (1995), Multivariate modelling of water resources time series using artificial neural networks, *J. Hydrol. Sci.*, *40*(2), 145–163.

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

Regonda, S. K., B. Rajagopalan, M. Clark, and E. Zagona (2006), A multi-model ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin, *Water Resour. Res.*, *42*, W09404, doi:10.1029/2005WR004653.

Ropelewski, C. F., and M. S. Halpert (1986), North American precipitation and temperature patterns associated with El-Nino-Southern Oscillation (ENSO), *Mon. Weather Rev.*, *114*, 2165–2352.

Ropelewski, C. F., and P. D. Jones (1987), An extension of the Tahiti-Darwin Southern Oscillation Index, *Am. Meteorol. Soc.*, *115*(9), 2161–2165.

Sax, J. L., B. H. Thompson, J. D. Leshy, and R. H. Abrams (2000), *Legal Control of Water Resources: Cases and Materials*, West Group, St. Paul, Minn., 956 pp.

Schölkopf, B., K.-K. Sung, C. J. C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik (1997), Comparing support vector machines with gaussian kernels to radial basis function classifiers, *IEEE Trans. Signal Process.*, *45*(11), 2758–2765.

Sharma, A., D. G. Tarboton, and U. Lall (1997), Streamflow simulation: A nonparametric approach, *Water Resour. Res.*, *33*(2), 291–308.

Shrestha, E., S. Ahmad, W. Johnson, P. Shrestha, and J. Batista (2011), Carbon footprint of water conveyance versus desalination as alternatives to expand water supply, *Desalination*, *27*(1–3), 120–127.

Shrestha, E., S. Ahmad, W. Johnson, and J. Batista (2012), The carbon footprint of water management policy options, *Energy Policy*, *42*, 201–212, doi:10.1016/j.enpol.2011.11.074.

Simonovic, S. P., and S. Ahmad (2005), Computer-based model for flood evacuation emergency planning, *Nat. Hazards*, *34*(1), 25–51.

Singh, J., H. V. Knapp, and M. Demissie (2004), Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT, *Rep. ISWS CR 2004-08*, Ill. Water Surv., Champaign, Ill.

Singh, J., H. V. Knapp, J. G. Arnold, and M. Demissie (2005), Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT, *J. Am. Water Resour. Assoc.*, *41*(2), 361–375.

Singhrattna, N., B. Rajagopalan, M. P. Clark, and K. K. Kumar (2005), Seasonal forecasting of Thailand summer monsoon rainfall, *Int. J. Climatol.*, *25*, 649–664.

Smola, A. J., and B. Schölkopf (2004), A tutorial on support vector regression, *Stat. Comput.*, *14*, 199–222.

Smola, A. J., B. Schölkopf, and K. R. Muller (1998), The connection between regularization operators and support vector kernels, *Neural Networks*, *11*, 637–649.

Stephen, H., S. Ahmad, T. C. Piechota, and C. Tang (2010), Relating surface backscatter response from TRMM precipitation radar to soil moisture: Results over a semi-arid region, *Hydrol. Earth Syst. Sci.*, *14*(2), 193–204.

Stone, M. (1974), Cross-validatory choice and assessment of statistical predictions, *J. R. Stat. Soc.*, *36*, 111–147.

Suykens, J. A. K. (2001), Nonlinear modeling and support vector machines, in *Proceedings of IEEE Instrumentation and Measurement Technology Conference*, vol. 1, pp. 287–294, Budapest, Hungary, doi:10.1109/IMTC. 2001.928828.

Thomas, B. E. (2007), Climatic fluctuations and forecasting of streamflow in the Lower Colorado River Basin, *J. Am. Water Resour. Assoc.*, *46*(6), 1550–1569.

Tokar, A. S., and P. A. Johnson (1999), Rainfall-runoff modeling using artificial neural networks, *J. Hydrol. Eng.*, *4*(3), 232–239.

Tokar, A. S., and M. Markus (2000), Precipitation-Runoff modeling using artificial neural networks and conceptual models, *J. Hydrol. Eng.*, *5*(2), 156–161.

Tripathi, S., V. V. Srinivas, and R. S. Nanjundiah (2006), Downscaling of precipitation for climate change scenarios: A support vector machine approach, *J. Hydrol.*, *330*, 621–640.

Twarakavi, N. K. C., D. Mishra, and S. Bandopadhyay (2006), Prediction of arsenic in bedrock derived stream sediments at a gold mine site under conditions of sparse data, *Nat. Resour. Res.*, *15*(1), 15–26.

Twarakavi, N. K. C., J. Simunek, and M. G. Schaap (2009), Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines, *Soil Sci. Soc. Am. J.*, *73*(5), 1443–1452.

U.S. Environmental Protection Agency (2002), Guidance for quality assurance project plans for modeling, *EPA QA/G-5M, Rep. EPA/240/R-02/007*, Off. of Environ. Inf., U.S. Environ. Protect. Agency, Washington, D.C.

Vapnik, V. (1995), *The Nature of Statistical Learning Theory*, Springer, New York.

Vapnik, V. (1998), *Statistical Learning Theory*, Wiley, New York.

Vedwan, N., S. Ahmad, F. Miralles-Wilhelm, K. Broad, D. Letson, and G. Podesta (2008), Institutional evolution in Lake Okeechobee management in Florida: Characteristics, impacts, and limitations, *Water Resour. Manage.*, *22*(6), 699–718.

Venkatesan, A. K., S. Ahmad, W. Johnson, and J. Batista (2011a), Salinity reduction and energy conservation in direct and indirect portable water reuse, *Desalination*, *272*(1–3), 120–127.

Venkatesan, A. K., S. Ahmad, W. Johnson, and J. Batista (2011b), System dynamics model to forecast salinity load to the Colorado River due to urbanization within the Las Vegas valley, *Sci. Total Environ.*, *409*(13), 2616–2625.

Viles, H. A., and A. S. Goudie (2003), Interannual, decadal and multidecadal scale climate variability and geomorphology, *Earth Sci. Rev.*, *61*, 105–131.

Walsh, J. E., W. L. Chapman, and T. Shy (1996), Recent decrease of sea level pressure in the Central Arctic, *J. Clim.*, *9*, 480–486.

Wang, B., R. G. Wu, and X. H. Fu (2000), Pacific-East Asia teleconnections: How does ENSO affect East Asian climate?, *J. Clim.*, *13*, 1517–1536.

Wang, X. L., and V. R. Swail (2001), Changes of extreme wave heights in Northern Hemisphere oceans and related atmospheric circulation regimes, *J. Clim.*, *14*, 2204–2221.

Webb, R. H., and J. L. Betancourt (1992), Climate variability and flood frequency of the Santa Cruz River, Pima County, Arizona, *U.S. Geol. Surv. Water Suppl. Pap., 2379*, pp. 40.

Webb, R. H., G. J. McCabe, R. Hereford, and C. Wilkowske (2004), Climatic fluctuations, drought, and flow in the Colorado River Basin, *U.S. Geol. Surv. Fact Sheet*, *3062-04*, 4 pp.

Wedgbrow, C. S., R. L. Wilby, H. R. Fox, and G. O'Hare (2002), Prospects for seasonal forecasting of summer drought and low river flow anomalies in England and Wales, *Int. J. Climatol.*, *22*(2), 219–236.

Weng, H., K. Ashok, S. K. Behera, S. A. Rao, and T. Yamagata (2007), Impacts of recent El Niño Modoki on dry/wet conditions in the Pacific rim during boreal summer, *Clim. Dyn.*, *29*(2–3), 113–129, doi:10.1007/s00382-00007-00234-00380.

Wilson, L. (1973), Variations in mean annual sediment yield as a function of mean annual precipitation, *Am. J. Sci.*, *273*, 335–349.

Xu, Z. X., K. Takeuchi, and H. Ishidaira (2004), Correlation between El Nino-Southern Oscillation (ENSO) and precipitation in South-east Asia and the Pacific region, *Hydrol. Process.*, *18*, 107–123.

Yu, X., and S.-Y. Liong (2007), Forecasting of hydrologic time series with ridge regression in feature space, *J. Hydrol.*, *332*, 290–302.