Climatic Change (2010) 101:575–597
DOI 10.1007/s10584-009-9705-0

Potential extents for ENSO-driven hydrologic drought forecasts in the United States

Jae H. Ryu · Mark D. Svoboda · John D. Lenters · Tsegaye Tadesse · Cody L. Knutson

Received: 25 August 2008 / Accepted: 26 August 2009 / Published online: 7 November 2009

Abstract The relationship between the El Niño-Southern Oscillation (ENSO) and hydrologic variability in the United States is investigated using Empirical Orthogonal Function (EOF)/Principal Component Analysis (PCA). The multivariate ENSO index (MEI) is utilized to identify strong coherences associated with multiple months (1-, 2-, 4-, 6-, 12-, 24-, 48-month) of the Log-Standardized Hydrologic Drought Index (LSHDI) in the conterminous states for the period 1950–2005. Based on 56 years of monthly streamflow data for 102 forecast climate divisions, this research explores the spatial and temporal variation of hydrologic responses corresponding to ENSO events. Preliminary results show that a potential predictor of the dominant streamflow modes in the northern Great Plains is identified from streamflows in western Arizona. Also, positive relationships between hydrologic drought and El Niño were found in the Pacific Northwest (Washington, Oregon, and northern California), whereas negative relationships were detected in southern California and the northern Great Plains. These findings will provide useful insights to help improve streamflow forecast potential and capabilities, and minimize the impacts of hydrologic events (e.g. floods and droughts) associated with ENSO events.

J. H. Ryu (✉) · M. D. Svoboda · J. D. Lenters · T. Tadesse · C. L. Knutson
School of Natural Resources, University of Nebraska, Lincoln, NE 68588-0988, USA
e-mail: jryu2@unl.edu

M. D. Svoboda
e-mail: msvoboda2@unl.edu

T. Tadesse
e-mail: ttadesse2@unl.edu

C. L. Knutson
e-mail: cknutson1@unl.edu

J. D. Lenters
Department of Geosciences, University of Nebraska, Lincoln, NE, USA
e-mail: jlenters2@unl.edu

1 Introduction

Hydrologic events, such as floods and droughts, cause significant economic, social, and environmental impacts every year in the United States. Droughts, in particular, produce a complex sequence of intertwined effects that ripple throughout society, affecting both the short- and long-term viability of many activities. As a result, drought is one of the costliest natural disasters. In fact, the U.S. Department of Commerce's National Climatic Data Center has recorded 13 drought years in the United States from 1980 to 2007 that have exceeded $1.0 billion in damages/costs. The total cost for the droughts and associated heat waves is nearly $157 billion. Although a rough estimate, this represents an annual average of at least $5.6 billion dollars in direct drought losses. Better understanding of hydro-climatic relationships and the development of enhanced forecasts has potential for fostering more proactive planning and hazard response activities that will help reduce these costs, as well as the stress on society and the environment.

Improved knowledge of climate dynamics and teleconnections explained by ocean-atmosphere phenomena such as the El Niño-Southern Oscillation (ENSO) (Battisti and Sarachik 1995), the Pacific Decadal Oscillation (PDO; Mantua et al. 1997), and the Arctic Oscillation (AO; Thompson and Wallace 2000) has raised significant issues within scientific communities, including its relation to climate (e.g., precipitation and temperature), hydrologic variability (e.g., streamflow), and regional water resource planning (e.g., sustainability). In particular, ENSO, the most prominent climate signal at seasonal to interannual scales, has been studied extensively and utilized as an indicator to identify precipitation and temperature patterns from regional scales (Andrews et al. 2004) to global scales (Ropelewski and Halpert 1987), as well as the nature and magnitude of streamflow variability in the United States (Kahya and Dracup 1993; Piechota and Dracup 1996; Redmond and Koch 1991).

The evidence of teleconnections between ENSO and hydrologic variability, which could perhaps lead to improvements in mid-range (up to 3–6 months) streamflow forecast capabilities, would provide useful insights for water resource managers in making their systems less vulnerable to drought. For instance, such forecasts would allow managers to provide extra flood control volumes when high flows are expected, to store more water at the beginning of a drawdown season when less than average flows are expected, and to encourage wise water use during unseasonably dry summer months.

The application of ENSO to drought studies by Ropelewski and Halpert (1987) and others (Barlow et al. 2001; Hidalgo and Dracup 2003; Karl and Koscielny 1982) has initiated new opportunities to include both increased levels of complexity associated with climate dynamics and other details of hydrologic variations. However, few studies have focused solely on hydrologic drought in the United States (Piechota and Dracup 1996; Ropelewski and Halpert 1986). Therefore, in this research, the authors identify relationships between ENSO and major hydrologic droughts in the conterminous Unites States based not only on their severity, but also on their spatial and temporal extent over the water years from 1950 to 2005. Potential opportunities for ENSO-driven hydrologic drought prediction in the United States are also discussed.

Two major spatial analysis techniques, harmonic analysis and empirical orthogonal function (EOF)/Principal Components Analysis (PCA), have been utilized to plot areal extents as well as temporal patterns of drought. Ropelewski and Halpert (1987) used a harmonic analysis to compute the monthly composite value of temperature and precipitation to identify regional coherence with respect to the evolution of ENSO episodes. This approach was expanded by Piechota and Dracup (1996) to identify several coherent regional responses to ENSO related to hydrologic drought. As another approach, EOF analysis has been widely used in science communities to investigate both spatial and temporal variations explained by the multivariate fields. For drought studies, Karl and Koscielny (1982) investigated spatial coherence of identifiable patterns of Palmer Drought Severity Indices (PDSI) using EOF and spectral analysis, and found that the duration of droughts is significantly higher in the interior portions of the United States than in areas closer to the coasts. Since then, Lins (1985) has extended this approach to investigate streamflow variability in the United States from 1931 to 1978.

None of these previous studies, however, has utilized the ENSO signal as an indicator to identify potential extents of ENSO-driven hydrologic droughts based on the most recent data up to year 2005. Therefore, this research explores the spatial and temporal relationships between ENSO and streamflow at seven different timescales (e.g. 1-, 2-, 4-, 6-, 12-, 24-, and 48-month timescales) for the forty-eight contiguous states using an EOF/PC analysis, and we discuss how to best use the ENSO signal to anticipate future droughts in the United States. The results from this research will help bridge the gap between operational water management and climate dynamics, and also contribute to the enhancement of drought prediction.

In this paper, a brief description of the data is first presented in order to discuss the sources and quality of data. The research methodology is then described, including extension of the streamflow data record, description of the hydrologic drought index, and statistical analysis of the spatial and temporal variability. Next, the potential opportunities for developing and utilizing hydrologic droughts forecasts associated with the Multivariate ENSO Index (MEI) are demonstrated. Finally, the paper concludes with a discussion of the research results and recommendations for future work.

2 Data

Records of unregulated streamflow (i.e., natural streamflow, as opposed to streams with upstream diversions or dam operations) are extremely important for understanding the effects of hydrologic drought on regional water resources. Currently, the United States Geological Survey (USGS) maintains 1,659 unregulated gauge stations, known as the Hydro-Climatic Data Network (HCDN), across the United States, to provide naturalized streamflow data for research activities related to climate change and other subjects (http://pubs.usgs.gov/wri/wri934076/1st_page.html). To analyze the effects of ENSO on streamflow, this research utilized a set of 102 streamflow gauging stations (a subset of the 1,659 naturalized streamflow gauge stations) distributed throughout the conterminous United States for which daily data are available for water years 1950–2005.

Selection criteria for streamflow gauge stations were chosen based on their relevance for developing hydrologic drought forecasts. At present, the National Oceanic and Atmospheric Administration's (NOAA's) Climate Prediction Center (CPC) is providing seasonal forecast information for 102 climate divisions using a 2-degree spatial grid domain. Monthly average values of streamflow have been assembled for 102 climate divisions across the country (Fig. 1). It was felt that the project stations should be well distributed and suitable for representing the regional climatology using a hydrologic runoff model, and that they should be real-time stations that can be used for improving streamflow forecasts associated with climate forecast models when they become available. Additionally, it was decided that at least one station in each climate division must have a minimum of 56 years of data and be located close to the centroids of the 102 climate division polygons in order to minimize systematic errors associated with spatial interpolation and to represent the flow responses at stream gauging station link directly to climate in the same geographic location. Note that streamflow at a particular gauging location integrates responses coming from the larger contributing area so that such a flow could be partially decoupled from climate variables within the same climate division. These facts are also considered during selection processes.

To identify stations meeting this criterion, the project researchers first selected those USGS HCDN gauge stations that were well maintained and had a serially complete dataset.

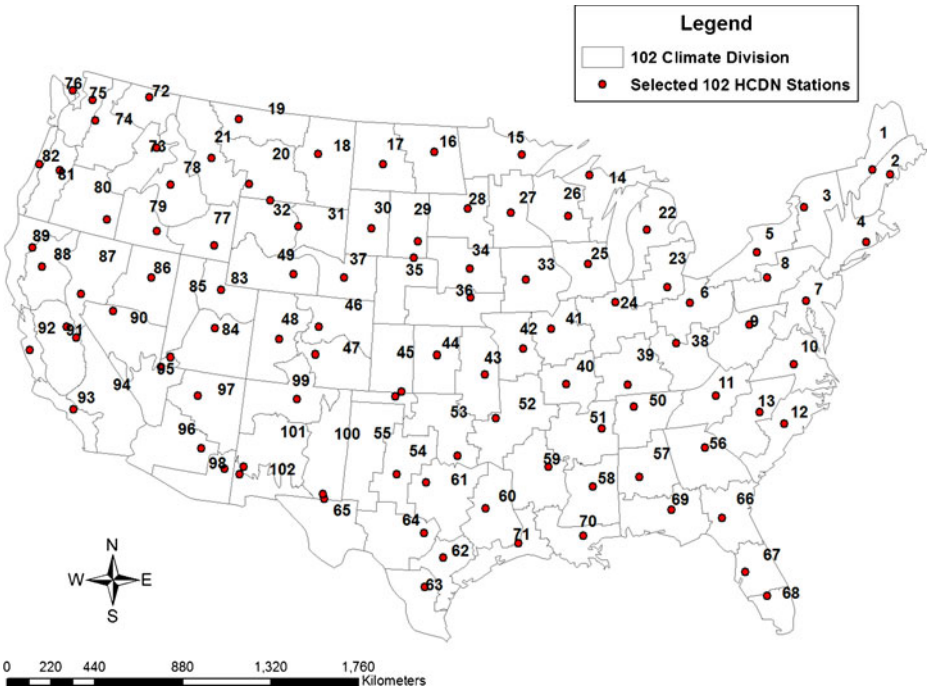

Fig. 1 The 102-station HCDN network and corresponding climate divisions

Next, the best candidates were chosen from the selected stations by identifying those closest to the centroids of the individual climate division polygons. The Multivariate ENSO Index (MEI) (Wolter 1987; Wolter and Timlin 1998) was then utilized to identify the relationship between climate variability and hydrologic drought. The MEI is dynamically composed of diverse ocean interactions including sea-level pressure, wind, surface temperature, and the total amount of cloudiness. The MEI value is extended from one month to the first week of the following month based on real-time observations in the ocean, and is updated monthly. Positive values of the MEI represent the warm phase of ENSO (El Niño), while negative values of the MEI indicate the cold phase (La Niña; http://www.cdc.noaa.gov/ENSO).

3 Methodology

3.1 Streamflow extension

Most selected HCDN gauging stations have a complete data record for the water years from 1950 to 2005. However, four gauging stations—6677500 (Climate Division 37), 9408400 (Climate Division 85), 1024900 (Climate Division 90), and 9431500 (Climate Division 102)—were extended to complete their datasets for analysis periods from 1950 to 2005 because no other valid sites are available in this area.

The Maintenance of Variance Extension Type 4 (MOVE4) method was used in this study and is described briefly in the following paragraphs. More details can be found in Appendix and in water resources literature by Hirsch (1982) and Vogel and Stedinger (1985). The advantage of this method is that it provides an efficient estimation of the mean and variance of flows at a short-record gauge using statistical information at a nearby long-record gauge. Two observed streamflows are denoted by

$$\begin{matrix} x_1, \ldots, x_{n1}, x_{n1+1}, \ldots, x_{n1+n2} \\ y_1, \ldots, y_{n1} \end{matrix}$$

where, n_1 is the length of the short streamflow record (y), and $n_1 + n_2$ is the length of the long streamflow record (x). n_1 is also the length of concurrent observations in the longer records, x. The MOVE4 method, which is used for filling values, is an extended version of linear regression expressed by

$$y_i = a + bx_i \tag{1}$$

where the parameters a and b can be estimated to minimize the difference between observed and estimated values of y (See Appendix for details).

For instance, streamflow data for Horse Creek near Lyman, Nebraska (USGS 6677500) were unavailable from September 30, 1998, through September 30, 2005. Thus, historic observed streamflow for this station had to be estimated using the Cheyenne River station at Edgemont, South Dakota (USGS 6395000), as a base station. This was possible since there was a good correlation between the two stations. To evaluate whether or not the gap-filling procedure is adequate, a series of statistics have been measured, such as mean, standard deviation, and the correlation coefficient of monthly flows and monthly volume in between the base station and

short-record station. All statistical measures are satisfactory (in fact, between 0.90 and 0.98) so that extended records can be considered for further analysis.

3.2 Log-Standardized Hydrologic Drought Index (LSHDI)

In general, drought can be characterized from a variety of perspectives, thus several classifications of drought have emerged, including meteorological, agricultural, hydrologic, and water resource management drought (Wilhite and Glantz 1985). For this study, hydrologic drought has been investigated because HCDN gauge stations are characterized by streamflow quantity. To better understand the occurrence of hydrologic drought, a complete set of streamflow data were analyzed to identify the relationship between the MEI and a Log-Standardized Hydrologic Drought Index (LSHDI). The project researchers used a total of seven different stages of monthly moving averages from 1, 2, 4, 6, 12, 24, and up to 48 months without weighting. The main reason for incorporating a moving average approach is to avoid quick changes for low-frequency processes (e.g., droughts), which can be dominated by extreme events such as low flows or high flows. For instance, if low flow events occur between high flow events during the wet season, the critical low flows would be undetected through a peak flow analysis. A 2-month total moving average is defined as

$$\begin{aligned} Y_t &= x_t, \quad \text{when} \quad t = 1 \\ Y_t &= \frac{(x_{t-1} + x_t)}{2}, \quad \text{when} \quad t > 1 \end{aligned} \tag{2}$$

where

t 1, 2, ..., T is the monthly time step,
x_t total monthly streamflow at time t,
Y_t two-month moving average streamflow at time t.

The procedure utilized in calculating the LSHDI is to de-seasonalize the logarithm of the monthly moving average streamflows by subtracting the long-term mean of the monthly logarithm and then dividing by the standard deviation of the logarithm for the given month to create a zero-mean process (Log-Standardized Hydrologic Drought Index), which is suitable in matching the MEI units. The advantages of using a logarithmic transformation have been well documented in the water resources literature (Box and Cox 1964; Stedinger 1980). The correlation coefficients between the MEI and LSHDI are then calculated to identify the spatial signature of ENSO effects on hydrologic drought in the United States. The LSHDI is calculated as:

$$LSHDI_{t,i} = \frac{\left(Y_t - \bar{Y}_i\right)}{\sigma_i} \tag{3}$$

where

t 1, 2, ..., T is the monthly time step,
i 1, 2, ..., 12 is the month index,
Y_t logarithm of the monthly, moving average streamflows at time t,
$\bar{Y}_i$ long-term mean of the logarithm of the moving average streamflow for month i,

σ_i standard deviation of the logarithm of the moving average streamflow for month i,
$LSHDI_{t,i}$ Log-Standardized Hydrologic Drought Index at time t and month i.

Like other drought indices such as the Palmer Drought Severity Index (PDSI) and Standardized Precipitation Index (SPI), the LSHDI is standardized, which allows direct comparisons of diverse streamflow variability across the country. The LSHDI is a measure of the streamflow abnormality over a period of low flow, average flow, and peak flow. Thus, the lower the index, the more severe the hydrologic drought associated with low flow, while the higher the index, the more likely the occurrence of flooding due to subsequent peak flows. The range of the index can be analogous to PDSI values of −3 to −4 as a severe drought, −2 to −3 moderate drought, 2 to 3 moderately wet, 3 to 4 extremely wet, and so on.

3.3 Spatial variability of LSHDI

To evaluate the spatial variability of the LSHDI, EOF analysis was first applied using the covariance matrix of LSHDI to identify the primary modes of spatial hydrologic drought in the United States and a correlation analysis between MEI and LSHDI was then conducted to determine significant relationships.

As a first step, eigenvalue analysis is utilized to retain an appropriate number of significant EOFs to represent a sufficient fraction of the variances in the data. As a rule of thumb, the method of North et al. (1982) was adopted to differentiate signal from noise. The 95% confidence error in the estimation of the eigenvalues is approximately

$$\lambda^* = \lambda\sqrt{\frac{2}{N^*}} \tag{4}$$

where λ^* is the sampling error of a particular eigenvalue λ, and N^* is the number of realizations (degrees of freedom) in the dataset (North et al. 1982). Figure 2 illustrates the first 10 eigenvalues with the standard errors λ^* due to sampling for each eigenvalue λ. As shown in the figure, the first 3–5 eigenvalues dominate the others in most of the LSHDI because the eigenvalues drop sharply. Note that the second and third eigenvalues are indistinguishable in most of the LSHDI except the 1-month LSHDI since their eigenvalues are the same to within the statistical uncertainty. The existence of the significance at 95% confidence level in this analysis could be contentious due to indistinguishable eigenvalues after the first month LSHDI. However, since the moving average flows carry precedent hydrologic information over the following stage of hydrologic condition via current condition, eigenvalues are readily tied together and this phenomenon can be explained by physical based hydrologic processes, the second and third eigenvalues are also considered as a dominant mode for the study. The percentage of variance and the cumulative percentage of variance explained by the first five EOFs are listed in Table 1. Note that italicized numbers in the table represent cumulative percentage of variance with an explained variance of more than 40% of the total variance.

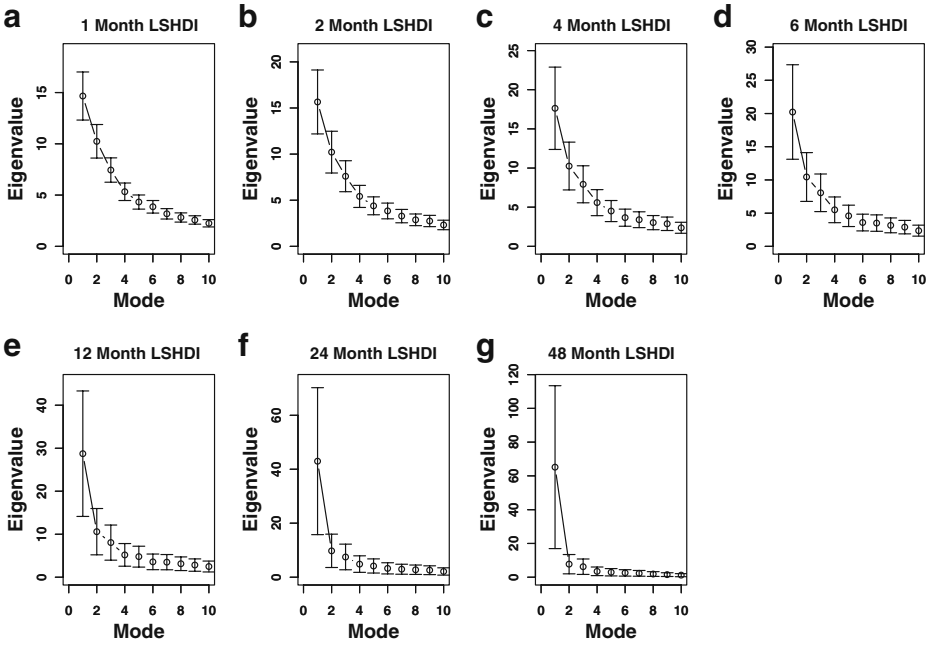

Fig. 2 Eigenvalue spectrum of the Log-Standardized Hydrologic Drought Index

3.4 Temporal variability of LSHDI

While the EOF analysis is used to identify spatial patterns among space–time variables, the principal component analysis (PCA) is also widely used to identify temporal patterns in datasets. For EOF/PCA, the eigenvalue-eigenvector pairs derived from correlation or covariance matrices are required, and the selection of matrices is dependent on the purpose of the study. However, it is suggested to compute eigenvalue–eigenvector pairs based on the covariance matrix if the unit of all data used is identical (e.g. precipitation measurements from many different locations). Otherwise, it is preferable to conduct PCA using the correlation matrix (Wilks 2006). For instance, measurements of many different fields, including precipitation, temperature, wind speed, and sea surface temperature, may require arbitrary relative scaling factors rather than their usual physical magnitude, which occurs in different units.

Like EOF, the key aspect of the PCA framework is that it provides useful insights for the analyst to present an overall statistical structure with fewer critical variables, which is a subset of all of the variables contained in the original data. Furthermore, the PCA is able to reconstruct the original time series of the data set through a linear combination of decomposed orthogonal and independent vectors. The simplified mathematical formulation of PCA is available at Eqs. 8 and 9 in Appendix.

3.5 Lagged correlation analysis

To identify distinct phase features of LSHDIs before and/or after the ENSO cycle, lagged correlation analysis was also utilized. Correlation analysis is commonly used to represent the strength and direction of a linear relationship between two variables.

Table 1 Percentage of variance and cumulative percentage of variance explained by the first five EOFs of the LSHDI

Number of EOF	1-Month LSHDI			2-Month LSHDI			4-Month LSHDI			6-Month LSHDI			12-Month LSHDI			24-Month LSHDI			48-Month LSHDI		
	PV	CPV	SE	PV	CPV	SE	PV	CPV	SE	PV	CPV	SE	PV	CPV	SE	PV	CPV	SE	PV	CPV	SE
1	14.7	14.7	2.3	15.7	15.7	3.5	17.6	17.6	5.3	20.2	20.2	7.1	28.7	28.7	14.6	43.0	*43.0*	27.2	65.2	*65.2*	48.2
2	10.2	24.9	1.6	10.2	25.9	2.3	10.3	27.9	3.1	10.4	30.6	3.7	10.6	39.3	5.4	9.8	*52.8*	6.2	7.7	*72.9*	5.7
3	7.4	32.3	1.2	7.6	33.5	1.7	7.9	35.8	2.4	8.1	38.7	2.8	8.0	*47.3*	4.1	7.5	*60.3*	4.7	6.2	*79.1*	4.6
4	5.3	37.6	0.8	5.4	38.9	1.2	5.6	*41.4*	1.7	5.5	*44.2*	2.0	5.2	*52.5*	2.6	4.9	*65.2*	3.1	3.5	*82.6*	2.6
5	4.3	*41.9*	0.7	4.4	*43.3*	1.0	4.5	*45.9*	1.3	4.6	*48.8*	1.6	4.8	*57.3*	2.4	4.1	*69.3*	2.6	3.0	*85.6*	2.2

PV percentage of variance, *CPV* cumulative percentage of variance, *SE* standard error from sampling

Lagged correlations, in particular, are another way to analyze the relationship between two sequences shifted by a certain unit of time, either forward or backward. A positive (negative) lag in time refers to a later (earlier) time. The results are discussed in next section.

4 Results and discussion

4.1 Spatial variations of streamflow

Figure 3 shows the first three EOFs of the LSHDI, which are the dominant mode of variation. Two strong dipole patterns appear in the first EOF, indicative of the greatest anomaly from total monthly streamflow. This occurs in three adjacent regions, which are characterized by opposite signs in the Rocky Mountains, the south central United States (around Texas), and the Upper Midwest (e.g., Minnesota, Iowa, and Missouri). Strong negative values prevail in Texas (with the notable exception of weak negative values found in central portions of the Great Plains), while positive values cover parts of the Rocky Mountains and the Upper Midwest. This pattern is notably changed as the timescale increases to the 48-month LSHDI, such that the positive anomalies in the Upper Midwest region expand to encompass much of the eastern United States.

Dipole patterns are also discernable in the second and third modes of the EOFs, but with different regional profiles: northwest (positive) versus southeast (negative) anomalies, and east (positive) versus southwest (negative) anomalies, respectively. Unlike the first EOF, however, the second and the third EOFs are not changed dramatically in terms of areal extent across the various timescales.

The dipole patterns from spatial analysis using EOF are identified above. However, sometimes, it is challenging to understand what the positive and negative values mean unless the physical relationships between data series are clearly known. For instance, such dipole patterns can simply be interpreted as opposite hydrologic events (e.g., negative values referring to dry-prone areas, positive values indicating wet-prone areas). Given the observed dipole pattern, therefore, there may be potential for predicting hydrologic anomalies in a given region based on opposite conditions in an adjacent region.

To understand the types of hydrologic events that correspond to the dipole anomalies identified in Fig. 3, an additional analysis was conducted using historical streamflow records. First, two significant regions where dipole patterns are identified were selected (e.g., Climate Division 75, near Seattle, Washington and Climate Division 33, in central Iowa). Then, three categorical flows (normal, drought, and wet flows) are identified out of monthly streamflow sequences from 1950 to 2005. For example, if the average streamflow for a given month in a particular year is above (below 33%) 66% of all flow sequences for a given month, it is defined as a ("dry-prone" month) "wet-prone" month, respectively, while flows in between these ranges are defined as "normal" months. Once hydrologic condition in each month is identified, a tally is then made to evaluate whether or not which hydrologic conditions (e.g. "wet-prone" or "dry-prone") is dominated in a particular climate division. As a result, it appears that Climate Division 75 is "wet-prone" region and Climate Division 33 is "dry-prone" region.

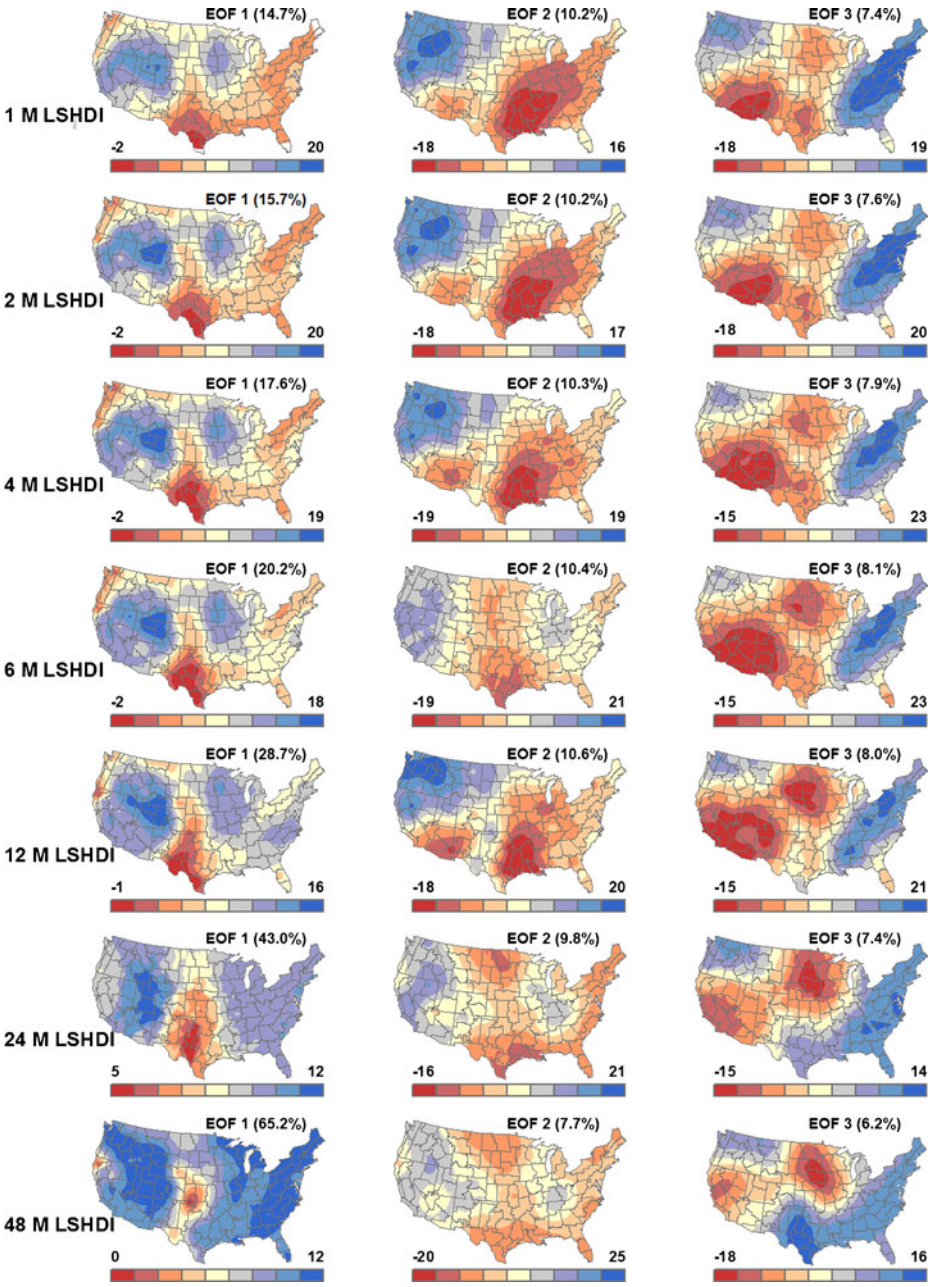

Fig. 3 Loadings (×100) of the first three EOFs for different timescales of the LSHDI

The basic shape of these EOFs closely agrees with the shape of the distribution of the unrotated principal components of annual streamflow identified by Lins (1985), which explains low flow characteristics of the southwest drought years of the early to mid-1950s and early 1970s. Sometimes, however, the orthogonality constraints limit

physical interpretation of the corresponding eigenvectors because the subsequent eigenvectors after the first eigenvector must be orthogonal to previously determined eigenvectors, regardless of the nature of the physical processes nested in the data. To reduce the effect of the orthogonality constraint and account for this effect, the Varimax approach has been utilized (Horel 1981; Lins 1985). Figure 4 represents loadings of the first three orthogonally rotated EOFs of the LSHDI. Unlike the first rotated EOF, the general patterns and loading magnitude of the second and third rotated EOFs are quite similar to their counterparts in different time horizons. It also appears that the effect of the rotations is to isolate coherent modes and enhance areas of strong streamflow variability where physical processes are less well defined because of orthogonal constraint embedded in the unrotated EOFs (Lins 1985).

The spatial distribution map of R-EOFs, correlation map between the MEI and LSHDI, and scatterplots of the correlation between the R-EOF and MEI time series appear in Fig. 5a–c, respectively. All three figures illustrate the close correspondence between these variables through time. The correlation map of the 1-month LSHDI with the MEI shows coefficients of −0.2 in the Pacific Northwest (PNW) and 0.4 in southern California and the northeast corner of the Great Plains. These findings also agree with the recent assessment of the North American climate patterns associated with ENSO in the United States during El Niño (warm) events (Green et al. 1997). More importantly, however, the spatial consistency of the dipole pattern shown in Fig. 5a and b with different time horizons (from 1 to 48 month) indicates that the regional characteristics of the core area isolated by two different analyses imply that the streamflow variability is related to ENSO events, thereby identifying potential candidate locations for streamflow forecast sites. For instance, as shown in Fig. 5a, there exist common characteristics wherein R-EOF show standing oscillatory signals taking the shape of a spatial dipole structure with different LSHDI time horizons up to 12 months. Temporal correlation between the MEI and LSHDI also shows variability in the Pacific Northwest that has a dipole pattern relative to the Great Plains Regions (Fig. 5b).

In the western United States, such as California and Washington, the impacts of ENSO have been studied extensively and utilized as an indicator to identify precipitation and temperature patterns from regional scales (Andrews et al. 2004) to global scales (Ropelewski and Halpert 1987), as well as the nature and magnitude of streamflow variability (Kahya and Dracup 1993; Piechota and Dracup 1996; Redmond and Koch 1991). These results also show evidence of teleconnections between oceanic indices and hydrologic variability, which could perhaps lead to improvements in mid-range (up to 3–6 months) streamflow forecast capabilities and provide useful insights for water resource managers in making their systems less vulnerable to climate change and variability in the western states (Hamlet et al. 2002; Wood and Lettenmaier 2006; Yao and Georgakakos 2001).

In contrast to the western United States, the effects of climate change and global warming on irrigation management, water system operation, and the economy of the Great Plains region, which stretches from the Canadian prairies to Texas, has been less studied and highlighted. In the past, a combination of water resources, including rainfall, streamflow, and ground water have provided a relative measure of security for the region's water needs, except during exceptional periods of drought. Recently, however, the availability of water in this region has been affected by several developments, including over-appropriation by municipal, industrial, and

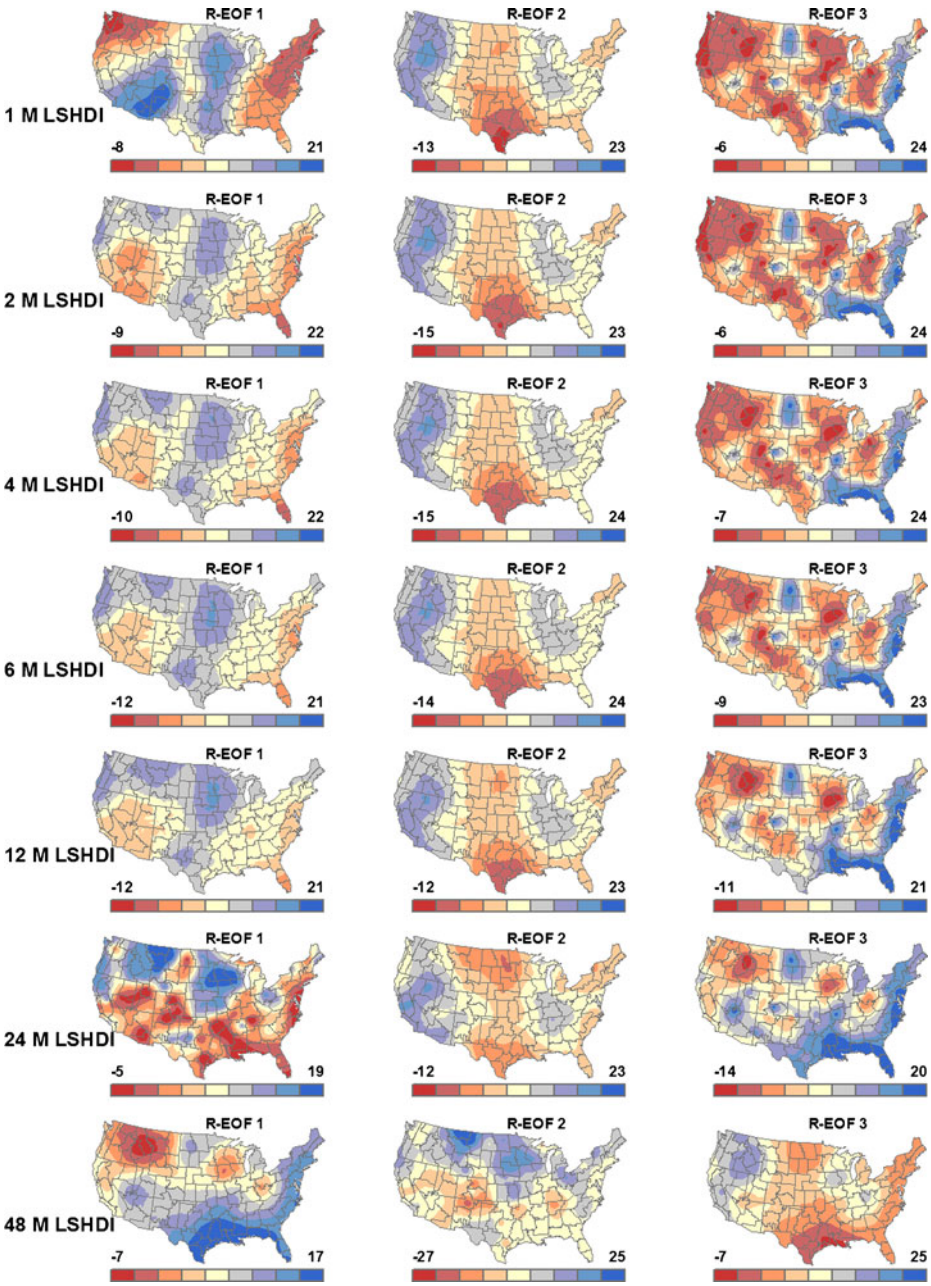

Fig. 4 Loadings (×100) of the first three rotated EOFs (R-EOF) for different timescales of the LSHDI using the Varimax solution

agricultural sectors; increasing recognition of the importance of water for ecosystems needs; reductions in water quality; and water compacts between states. Additionally, drought has been a recurrent challenge during the last decade for much of the

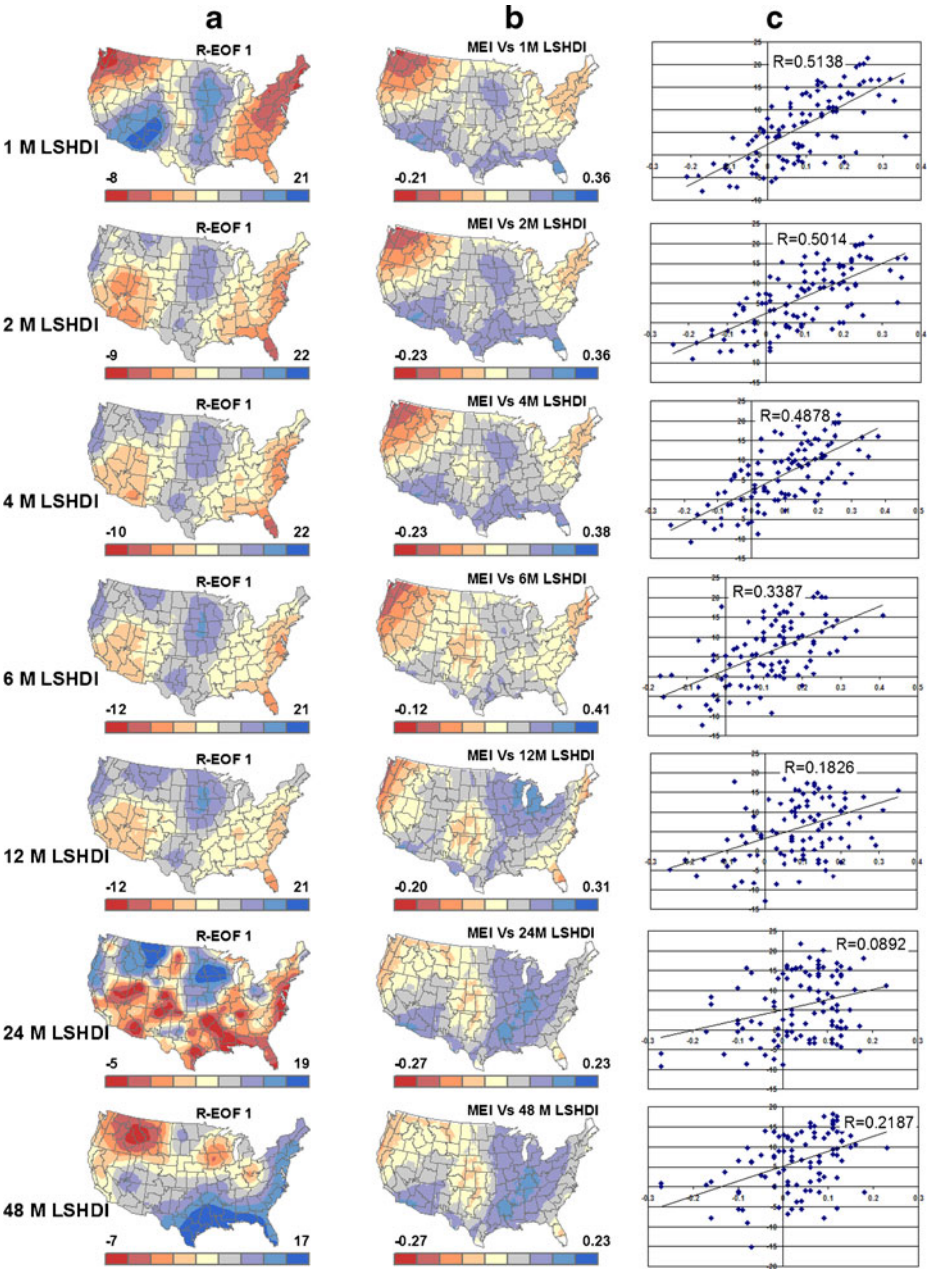

Fig. 5 **a** Spatial distribution of the first rotated EOF (R-EOF1) of the LSHDI, **b** temporal correlation map between the MEI and LSHDI, and **c** scatterplot showing the temporal correlation between the timeseries of R-EOF1 (y-axis) and the MEI (x-axis)

region. These stresses have reduced ground water and surface water availability and increased concerns about the sustainability of water resources in many locations in the region.

These concerns have spurred an increasing amount of interest in water resources, drought, and climate studies in the region. As shown in Fig. 5c, the R-EOFs and the MEI are correlated such that the ENSO signals tend to explain hydrologic variability in the Great Plains. Interestingly, since the spatial distribution and temporal correlation between the ENSO events and hydrologic conditions are relatively significant (for 1-, 2-, and 4 month periods), it appears that there is potential to predict hydrologic drought in the central U.S. using ENSO indices, which would be especially useful given the region's current water and drought planning needs.

4.2 Forecasting potential in selected areas

Based on the dipole patterns identified in the EOF analysis, an additional analysis was performed to investigate a possible link between hydrologic variability in the Great Plains region and surrounding areas. We hypothesize that if the one region is "dry-prone (e.g. CD 33, in Iowa)", another region with similar EOF loading values (e.g. CD 98, in Arizona) will also be "dry-prone". To corroborate this, we reconstructed the 1-month LSHDI of CD 33 and CD 98 using the simplified mathematical formulation of PCA at Eqs. 8 and 9 in Appendix. Figure 6a and b show observed time series of the 1-month LSHDI for CD 33 and CD 98 and the time series of the first three EOF vectors for the 1-month LSHDI. The correlation coefficient between the reconstructed 1-month LSHDI for CD 33 and CD 98 using the first three

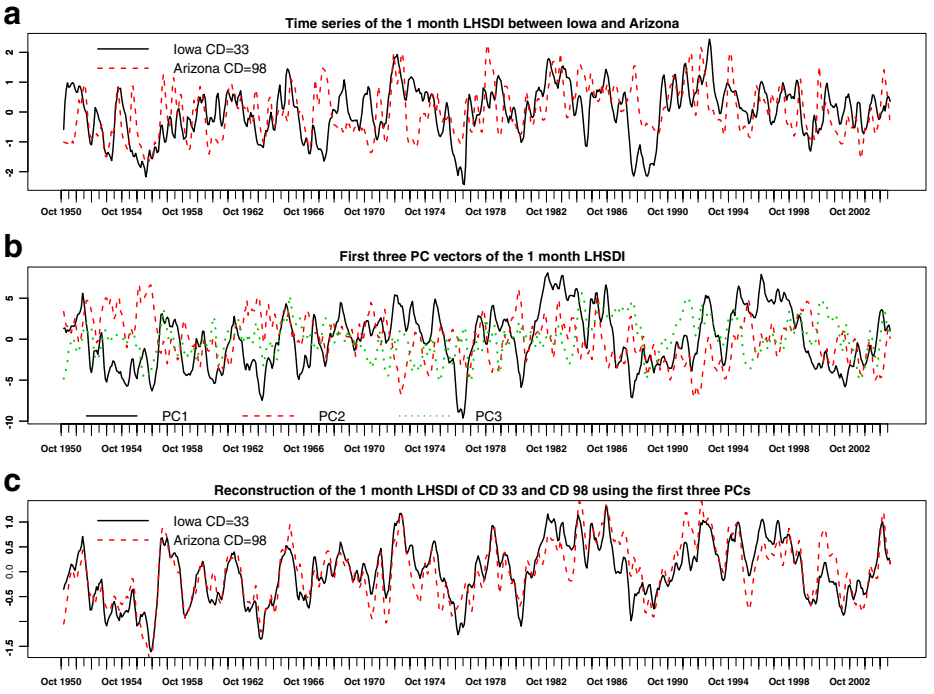

Fig. 6 The **a** observed timeseries, **b** first three EOF timeseries, and **c** reconstruction of the 1-month LSHDI for climate division 33 (Iowa) and climate division 98 (Arizona)

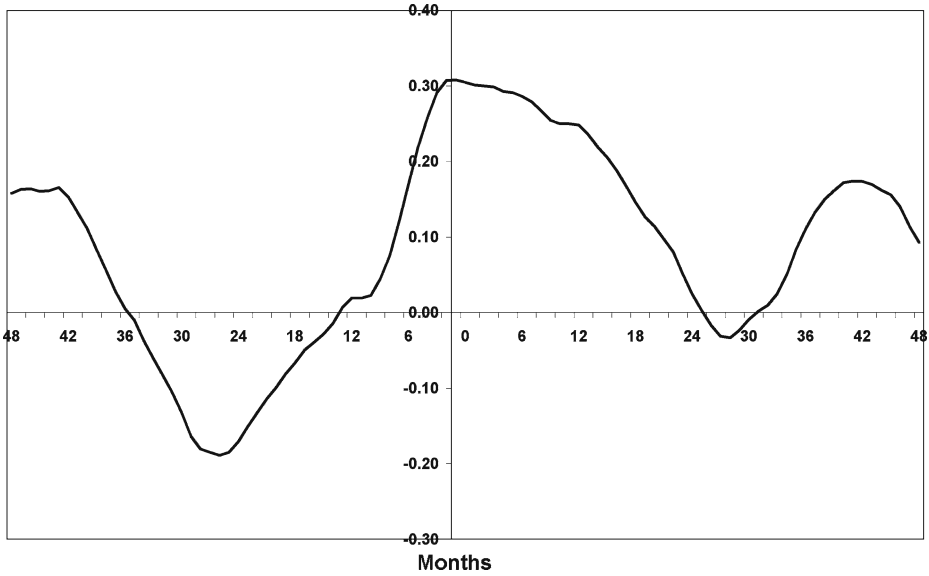

Fig. 7 Plot of the lag correlation between the MEI and 1-month LSHDI for Climate Division 33 (Iowa)

PCs is 0.84, which indicates that drought forecasts for CD 33 may be possible using climatological information from CD 98 (where the drought teleconnection pattern is found to extend into southern Arizona (Rajagopalan et al. 2000). This approach can also be extended to other regions accordingly using the simplified mathematical formulation of PCA at Eqs. 8 and 9. Note that the reconstructed values and the original time series should be identical if all eigenvectors are utilized to reconstruct the original data sets.

Lagged correlations are another way to analyze the relationship between two sequences shifted by a certain unit of time, either forward or backward. Figure 7 shows lagged correlations (the right side of the Y axis is a forward LSHDI lag and the left side of the Y axis is a backward LSHDI lag) between the MEI and 1-month LSHDI for climate division 33. There are no notable correlations between the two variables except the first couple of months lag. This implies that streamflow responds very quickly to ENSO events within short periods of time, perhaps 1 to 4 months. Based on these findings, it is suggested that more rigorous analysis of the mid-range streamflow forecast may aid water resources management in regionally specific applications associated with ENSO events.

5 Conclusion and future work

This research analyzed streamflow variability and correlations between ENSO and hydrologic drought indices across the United States. Three statistically significant modes of variation in monthly streamflow during water years 1950–2005 were

identified using empirical orthogonal function analysis. Inspection of the monthly correlation of the ENSO and hydrologic drought indices shows some relatively strong positive signals, namely in the northern Great Plains and the southwestern United States (e.g., around southern California), and negative signals around the Pacific Northwest. In other words, when drought-prone months are identified in the northern Great Plains, southern California often experiences drought at the same time, while wet conditions are dominant in the Pacific Northwest.

Spatial patterns of the orthogonal rotation of the EOFs of streamflow (using the Varimax solution) are very similar to the spatial correlation between MEI and streamflow. Thus, it appears that this approach will provide a regional profile of the large-scale spatial and temporal variation in hydrologic drought across the United States (as represented by streamflow). However, additional research efforts are needed to explore other large scale patterns beyond ENSO, including additional oceanic indices, the Southern Oscillation Index (SOI), Pacific Decadal Oscillation (PDO), Atlantic Multi-decadal Oscillation (AMO), Pacific/North American index (PNA), Arctic Oscillation (AO), and Madden-Julian Oscillation (MJO).

Our current understanding of ENSO relationships alone are not strong enough for predicting regional hydrologic droughts, but the findings from this study suggest that by applying knowledge of both teleconnections and intercorrelations between climate regions, ENSO events could help provide water and drought managers with useful insights in order to make better decisions if such events continue to be dominant in our future climate.

A number of improvements, enhancements, and new directions can be taken to improve the predictability of regional hydrologic drought associated with climate variability in the United States. These new directions may include: 1) correlate the time components with global sea surface temperature (SST) and/or sea level pressure (SLP) to identify large scale patterns that go beyond ENSO, 2) Analyze individual seasons to highlight important ENSO contributions to specific severe events, since it is clear that ENSO has different influences in different seasons, and 3) pursue an enhanced statistical analysis, such as canonical correlation analysis with precipitation and streamflow, and identify components of hydrologic events that are correlated with ENSO and other oceanic indices.

Finally, although the authors presume that the streamflow data used in this study have been rigorously quality controlled by USGS, it should be noted that in many locations, the unregulated streamflow data are not available during the winter season due to ice cover. Evaluating EOF analysis using these dataset as well as snow data, such as Snowpack Telemetry (SNOTEL) information, would be very interesting future research for enhancing the potential predictability of future droughts. While rainfall-dominated runoff regimes may respond relatively quickly to climate variability, basins with snowmelt or groundwater-dominated runoff responses may have a much longer lag time between precipitation and runoff. These differences in runoff regimes could be important controls on spatial and temporal patterns of drought indices.

Acknowledgements This work was partially supported by the US Department of Agriculture through the Risk Management Agency and the National Aeronautics and Space Administration under award No NNX08AL94G. The authors thank anonymous reviewers for insightful comments and suggestions, which helped to improve the quality of the manuscript substantially.

Appendix

Table 2 Station list: selected 102 hydro climate data network

CD	USGS station no.	Gauging station name	State	Latitude	Longitude
1	1055000	Swift River near Roxbury	Maine	44.6422	−70.5881
2	1038000	Sheepscot River at North Whitefield	Maine	44.2231	−69.5939
3	4256000	Independence River at Donnattsburg	New York	43.7472	−75.3347
4	1127500	Yantic River at Yantic	Connecticut	41.5586	−72.1219
5	3011020	Allegheny River at Salamanca	New York	42.1564	−78.7156
6	3219500	Scioto River near Prospect	Ohio	40.4194	−83.1972
7	1580000	Deer Creek at Rocks	Maryland	39.6303	−76.4036
8	1541500	Clearfield Creek at Dimeling	Pennsylvania	40.9717	−78.4061
9	3051000	Tygart Valley River at Belington	West Virginia	39.0292	−79.9361
10	2044500	Nottoway River near Rawlings	Virginia	36.9833	−77.8000
11	3465500	Nolichucky River at Embreeville	Tennessee	36.1764	−82.4575
12	2134500	Lumber River at Boardman	North Carolina	34.4422	−78.9606
13	2126000	Rocky River near Norwood	North Carolina	35.1483	−80.1758
14	4040500	Sturgeon River near Sidnaw	Michigan	46.5842	−88.5758
15	5130500	Sturgeon River near Chisholm	Minnesota	47.6736	−92.9000
16	5056000	Sheyenne River near Warwick	North Dakota	47.8056	−98.7158
17	6339500	Knife River near Golden Valley	North Dakota	47.1611	−102.0608
18	6131000	Big Dry Creek near Van Norman	Montana	47.3494	−106.3572
19	6099500	Marias River near Shelby	Montana	48.4272	−111.8889
20	6192500	Yellowstone River near Livingston	Montana	45.5972	−110.5653
21	12330000	Boulder Creek at Maxville	Montana	46.4722	−113.2331
22	4121500	Muskegon River at Evart	Michigan	43.8992	−85.2553
23	4191500	Auglaize River near Defiance	Ohio	41.2375	−84.3992
24	5525000	Iroquois River at Iroquois	Illinois	40.8236	−87.5819
25	5430500	Rock River at Afton	Wisconsin	42.6092	−89.0706
26	5399500	Big Eau Pleine River near Stratford	Wisconsin	44.8219	−90.0794
27	5280000	Crow River at Rockford	Minnesota	45.0867	−93.7339
28	5291000	Whetstone River near Big Stone City	Minnesota	45.2922	−96.4872
29	6452000	White River near Oacoma	South Dakota	43.7483	−99.5561
30	6425500	Elk Creek near Elm Springs	South Dakota	44.2483	−102.5028
31	6311000	North Fork Powder River near Hazelton	Wyoming	44.0278	−107.0803
32	6207500	Clarks Fork Yellowstone River near Belfry	Montana	45.0103	−109.0647
33	5451500	Iowa River at Marshalltown	Iowa	42.0658	−92.9075
34	6600500	Floyd River at James	Iowa	42.5767	−96.3119
35	6464500	Keya Paha River at Wewela	South Dakota	43.0289	−99.7803
36	6800500	Elkhorn River at Waterloo	Nebraska	41.2903	−96.2847
37	6677500[a]	Horse Creek near Lyman	Nebraska	41.9392	−103.9869
38	3253500	Licking River at Catawba	Kentucky	38.7103	−84.3108
39	3320500	Pond River near Apex	Kentucky	37.1222	−87.3194
40	7061500	Black River near Annapolis	Missouri	37.3361	−90.7886
41	5501000	North River at Palmyra	Missouri	39.8183	−91.5203
42	6908000	Blackwater River at Blue Lick	Missouri	38.9922	−93.1967
43	7183000	Neosho River near Iola	Kansas	37.8908	−95.4306
44	6864500	Smoky Hill River at Ellsworth	Kansas	38.7267	−98.2333
45	7157500	Crooked Creek near Nye	Kansas	37.0339	−100.1986

Table 2 (continued)

CD	USGS station no.	Gauging station name	State	Latitude	Longitude
46	6710500	Bear Creek at Morrison	Colorado	39.6531	−105.1953
47	7096000	Arkansas River at Canon City	Colorado	38.4339	−105.2567
48	9132500	North Fork Gunnison River near Somerset	Colorado	38.9258	−107.4336
49	6630000	North Platte River near Sinclair	Wyoming	41.8722	−107.0569
50	3434500	Harpeth River near Kingston Springs	Tennessee	36.1219	−87.0989
51	7029500	Hatchie River at Bolivar	Tennessee	35.2753	−88.9767
52	7197000	Baron Fork at Eldon	Oklahoma	35.9211	−94.8383
53	7331000	Washita River near Dickson	Oklahoma	34.2333	−96.9756
54	8082000	Salt Fork Brazos River near Aspermont	Texas	33.3339	−100.2378
55	7234000	Beaver River at Beaver	Oklahoma	36.8222	−100.5189
56	2217500	Middle Oconee River near Athens	Georgia	33.9467	−83.4228
57	2424000	Cahaba River at Centreville	Alabama	32.9450	−87.1392
58	2484500	Yockanookany River near Ofahoma	Mississippi	32.7056	−89.6722
59	7363500	Saline River near Rye	Arkansas	33.7008	−92.0258
60	8032000	Neches River near Neches	Texas	31.8922	−95.4306
61	8088000	Brazos River near South Bend	Texas	33.0242	−98.6436
62	8172000	San Marcos River at Luling	Texas	29.6650	−97.6497
63	8194500	Nueces River near Tilden	Texas	28.3086	−98.5569
64	8151500	Llano River at Llano	Texas	30.7511	−98.6694
65	8408500	Delaware River near Red Bluff	New Mexico	32.0231	−104.0542
66	2317500	Alapaha River at Statenville	Georgia	30.7039	−83.0333
67	2303000	Hillsborough River near Zephyrhills	Florida	28.1497	−82.2325
68	2256500	Fisheating Creek at Palmdale	Florida	26.9322	−81.3150
69	2361000	Choctawhatchee River near Newton	Alabama	31.3417	−85.6119
70	7375500	Tangipahoa River at Robert	Louisiana	30.5064	−90.3617
71	8030500	Sabine River near Ruliff	Texas	30.3036	−93.7436
72	12409000	Colville River at Kettle Falls	Washington	48.5944	−118.0614
73	13342500	Clearwater River at Spalding	Idaho	46.4486	−116.8264
74	12488500	American River near Nile	Washington	46.9775	−121.1681
75	12134500	Skykomish River near Gold Bar	Washington	47.8375	−121.6656
76	12048000	Dungeness River near Sequim	Washington	48.0144	−123.1314
77	13073000	Portneuf River at Topaz	Idaho	42.6250	−112.0889
78	13313000	Johnson Creek at Yellow Pine	Idaho	44.9622	−115.4994
79	13168500	Bruneau River near Hot Spring	Idaho	42.7711	−115.7194
80	10396000	Donner und Blitzen River near Frenchglen	Oregon	42.7911	−118.8667
81	14185000	South Santiam River below Cascadia	Oregon	44.3931	−122.5097
82	14306500	Alsea River near Tidewater	Oregon	44.3861	−123.8306
83	10128500	Weber River near Oakley	Utah	40.7361	−111.2458
84	9330500	Muddy Creek near Emery	Utah	38.9819	−111.2486
85	9408400[a]	Santa Clara River near Pine Valley	Utah	37.3833	−113.4825
86	10316500	Lamoille Creek near Lamoille	Nevada	40.6908	−115.4756
87	10312000	Carson River near Fort Churchill	Nevada	39.2917	−119.3111
88	11383500	Deer Creek near Vina	California	40.0142	−121.9472
89	11525500	Trinity River at Lewiston	California	40.7194	−122.8025
90	10249300[a]	South Twin River near Round Mountain	Nevada	38.8875	−117.2444
91	11266500	Merced River at Pohono Bridge	California	37.7169	−119.6653
92	11152000	Arroyo Seco near Soledad	California	36.2806	−121.3217
93	11098000	Arroyo Seco near Pasadena	California	34.2222	−118.1767
94	11230500	Bear Creek near Lake T.A. Edison	California	37.3383	−118.9731
95	9415000	Virgin River at Littlefield	Nevada	36.8917	−113.9236

Table 2 (continued)

CD	USGS station no.	Gauging station name	State	Latitude	Longitude
96	9498500	Salt River near Roosevelt	Arizona	33.6194	−110.9208
97	9402000	Little Colorado River near Cameron	Arizona	35.9264	−111.5667
98	9448500	Gila River at head of Safford Valley	Arizona	32.8683	−109.5106
99	8289000	Rio Ojo Caliente at La Madera	New Mexico	36.3497	−106.0436
100	8405500	Black River above Malaga	New Mexico	32.2289	−104.1506
101	9430500	Gila River near Gila	New Mexico	33.0611	−108.5367
102	9431500[a]	Gila River near Redrock	New Mexico	32.7269	−108.6750

CD climate division
[a]Ten percent of monthly data are extended by MOVE techniques

MOVE4 method

The MOVE4 estimates of a and b are obtained from

$$a = \frac{(n_1+n_2)\,\mu_y - n_1\bar{y}_1}{n_2}$$

$$b = \sqrt{\frac{\left[(n_1+n_2-1)\,\sigma_y^2 - (n_1-1)\,\frac{1}{n_1-1}\sum_{i=1}^{n_1}(y_i-\bar{y}_1)^2 - n_1\left(\bar{y}_1-\mu_y\right)^2 - n_2\left(a-\mu_y\right)^2\right]}{\left[(n_2-1)\,\frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}(x_i-\bar{x}_2)^2\right]}} \tag{5}$$

Fundamental calculation procedures are described below. First, the unbiased estimators of the Mean $\hat{\mu}_y$ and the standard deviation $\hat{\sigma}_y$ of the complete extended records are

$$\hat{\mu}_y = \bar{y}_1 + \frac{n_2}{n_1+n_2}\beta\,(\bar{x}_2-\bar{x}_1)$$

$$\hat{\sigma}_y = \frac{1}{n_1+n_2-1}\left\{\begin{array}{l}(n_1-1)\,\frac{1}{n_1-1}\sum_{i=1}^{n_1}(y_i-\bar{y}_1)^2 + (n_2-1)\,\beta^2\frac{1}{n_2-1}\sum_{i=n_1+n_2}^{n_1+n_2}(x_i-\bar{x}_2)^2 \\ +(n_2-1)\,\alpha^2\left(1-\rho^2\right)\frac{1}{n_1-1}\sum_{i=1}^{n_1}(y_i-\bar{y}_1)^2 + \frac{n_1n_2}{(n_1+n_2)}\beta^2\,(\bar{x}_2-\bar{x}_1)^2\end{array}\right\} \tag{6}$$

where,

$$\beta = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x}_1)(y_i - \bar{y}_1)}{\sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2}$$

$$\bar{y}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} y_i$$

$$\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i$$

$$\bar{x}_2 = \frac{1}{n_2} \sum_{i=n_1+1}^{n_1+n_2} x_i$$

$$\rho = \beta \frac{\sqrt{\frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2}}{\sqrt{\frac{1}{n_1 - 1} \sum (y_i - \bar{y}_1)^2}}$$

Then, the minimum variance linear estimator μ_y and σ_y can be obtained from the unbiased estimator of the mean and standard deviation, $\hat{\mu}_y$ and $\hat{\sigma}_y$, respectively.

$$\mu_y = (1 - \theta_1)\, \bar{y}_1 + \theta_1 \hat{\mu}_y$$

$$\sigma_y = \sqrt{(1 - \theta_2) \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (y_i - \bar{y}_1)^2 + \theta_2 \hat{\sigma}_y}$$

$$\text{where, } \theta_1 = \frac{(n_1 - 3)\, \rho^2}{(n_1 - 4)\, \rho^2 + 1}\ ,\ \theta_2 = \frac{(n_1 - 4)\, \rho^2}{(n_1 - 9.5)\, \rho^2 + 4.5} \tag{7}$$

Principal Component Analysis (PCA)

Let **X** be a data series with N observations of variable x over times with a number of M stations, so **X** is a data matrix with N rows and M columns (NXM). The singular value decomposition (SVD) of the covariance matrix of **X** is defined as follows:

$$\mathbf{C} = \mathbf{X}\mathbf{X}^T / N = \mathbf{U} \sum \mathbf{V}^T \tag{8}$$

where **U** is an MXN matrix and $\mathbf{V}^T$ an NXM matrix. The column vectors within matrix **U** comprise the eigenvectors, which represent empirical orthogonal functions,

and those vectors are linearly independent. Likewise, the matrix $\mathbf{V}^T$ is comprised of orthogonal vectors between individual columns. The principal components, $\mathbf{Y}$, are computed as

$$\mathbf{Y} = \mathbf{XU} \tag{9}$$

References

Andrews ED, Antweiler RC, Neiman PJ, Ralph FM (2004) Influence of ENSO on flood frequency along the California Coast. J Climate 17(2):337–348

Barlow M, Nigam S, Berbery EH (2001) ENSO, Pacific decadal variability, and U.S. Summertime precipitation, drought, and stream flow. J Climate 14(9):2105–2128

Battisti DS, Sarachik ES (1995) Understanding and predicting ENSO. Rev Geophys Suppl 33: 1367–1376

Box GEP, Cox DR (1964) An analysis of transformations. J R Stat Soc 26(2):211–252

Green PM, Legler DM, Miranda CJ, O'Brien JJ (1997) The North American climate patterns associated with the El Niño-Southern Oscillation. Center for Ocean-Atmospheric Prediction Studies, Project Report Series 97–1

Hamlet AF, Daniel H, Lettenmaier DP (2002) Economic value of long-lead streamflow forecasts for Columbia River hydropower. J Water Resour Plan Manage 128(2):91–101

Hidalgo HG, Dracup JA (2003) ENSO and PDO effects on hydroclimatic variations of the Upper Colorado River Basin. J Hydrometeorol 4(1):5–23

Hirsch RM (1982) A comparison of four record extension techniques. Water Resour Res 18(4): 1081–1088

Horel JD (1981) A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500 mb height field. Mon Weather Rev 109(10):2080–2092

Kahya E, Dracup JA (1993) U.S. streamflow patterns in relation to the El Niño/Southern oscillation. Water Resour Res 29(8):2491–2504

Karl TR, Koscielny AJ (1982) Drought in the United States: 1895–1981. J Climatol 2:313–329

Lins HF (1985) Streamflow variability in the United States: 1931–78. J Appl Meteorol 24(5):463–470

Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC (1997) A Pacific interdecadal climate oscillation with impacts on salmon production. Bull Am Meteorol Soc 78:1069–1079

North GR, Bell TL, Cahalan RF, Moeng FJ (1982) Sampling errors in the estimation of empirical orthogonal functions. Mon Weather Rev 110(7):699–706

Piechota TC, Dracup JA (1996) Drought and regional hydrologic variation in the United States: Associations with the El Niño-Southern Oscillation. Water Resour Res 32(5):1359–1373

Rajagopalan B, Cook E, Lall U, Ray B (2000) Spatiotemporal variability of ENSO and SST teleconnections to summer drought over the United States during the twentieth century. J Climate 13(24):4244–4255

Redmond KT, Koch RW (1991) Surface climate and streamflow variability in the Western United States and their relationship to large-scale circulation indices. Water Resour Res 27(9):2381–2399

Ropelewski CF, Halpert MS (1986) North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). Mon Weather Rev 114(12):2352–2362

Ropelewski CF, Halpert MS (1987) Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. Mon Weather Rev 115(8):1601–1626

Stedinger JR (1980) Fitting log normal distributions to hydrologic data. Water Resour Res 16(3):481–490

Thompson DWJ, Wallace JM (2000) Annular modes in the extratropical circulation, part I: month-to-month variability. J Climate 13(5):1000–1016

Vogel RM, Stedinger JR (1985) Minimum variance streamflow record augmentation procedures. Water Resour Res 21(5):715–723

Wilhite DA, Glantz MH (1985) Understanding the drought phenomenon: the role of definition. Water Int 10:111–120

Wilks DS (2006) Statistical methods in the atmospheric sciences. Academic, San Diego

Wolter K (1987) The Southern Oscillation in surface circulation and climate over the Tropical Atlantic, Eastern Pacific, and Indian Oceans as captured by cluster analysis. J Appl Meteorol 26(4):540–558

Wolter K, Timlin MS (1998) Measuring the strength of ENSO—how does 1997/98 rank? Weather 53(9):315–324

Wood AW, Lettenmaier DP (2006) A test bed for new seasonal hydrologic forecasting approaches in the Western United States. Bull Am Meteorol Soc 87(12):1699–1712

Yao H, Georgakakos A (2001) Assessment of Folsom Lake response to historical and potential future climate scenarios, 2. Reservoir management. J Hydrol 249(1–4):176–196