Introduction of new datasets of drought indices based on multivariate methods in semi-arid regions

Nastaran Chitsaz and Seyed-Mohammad Hosseini-Moghari

ABSTRACT

Drought is a slow and creeping worldwide phenomenon which has adversely affected arid and semi-arid regions of the world. Drought indices like Streamflow Drought Index (SDI) and Standardized Precipitation Index (SPI) offer quantitative methods for combating probable consequences of drought. In this article, the results of the drought indices trend showed that the case study suffers from hydrological drought more than meteorological drought. The correlation analysis between hydrological and meteorological drought was assessed in monthly and seasonal time scales. To this end, some multivariate techniques were used to summarize the SPI and SDI series of all stations into one new dataset. Three assessment criteria involving higher correlation among drought indices, higher eigenvalue in expansion coefficients, and following fluctuation and variation of original data were used to find the best new datasets and the best multivariate method. Results asserted the superiority of singular value decomposition (SVD) over other multivariate methods. EC1 in the SVD method was able to justify about 80% of the variability in drought indices for monthly time scales, as well as summer and spring for seasonal time series, which followed all fluctuations in original datasets. Therefore, the SVD method is recommended for aggregating drought indices.

Key words | drought monitoring, multivariate methods, singular value decomposition (SVD), Standardized Precipitation Index (SPI), Streamflow Drought Index (SDI)

Nastaran Chitsaz (corresponding author)
Seyed-Mohammad Hosseini-Moghari
University of Tehran,
Tehran,
Iran
E-mail: *nastaranchitsaz@ut.ac.ir*

INTRODUCTION

Drought, the greatest natural phenomenon threatening the world's human population, has extensive negative impacts spanning environmental, economic and social aspects (Bernard *et al.* 2013; Heudorfer & Stahl 2016). Drought is conventionally grouped as meteorological, hydrological, agricultural, or socioeconomic (Yang 2010; Li *et al.* 2017a, 2017b). Among these, the hydrological element is the most important part and is highly influenced by activities such as industry, agriculture, urban water supply and hydropower generation (Vasiliades *et al.* 2011; Ye *et al.* 2016).

Many researchers claim that drought is an abnormal event with complexity related to climate parameters, water balance status, and spatial and temporal dimensions (Heim 2002; Lennard *et al.* 2016; Li *et al.* 2017a, 2017b). Thus, it is vital to focus on the definition of drought and the quantification of its intensity and duration. To address this issue, it is necessary to perceive drought's characteristics and indices (duration, severity, spatiotemporal variability and frequency) because they offer a quantitative method for determining the onset of a drought event (Yoo *et al.* 2012; Tabari *et al.* 2013; Sung & Chung 2014; Khadr 2017; Ling *et al.* 2017).

Many indices have been devised to characterize hydrological and meteorological droughts, including Surface Water Supply Index (Shafer & Dezman 1982), Palmer Drought Severity Index (PDSI; Palmer 1965), the China-Z index, which is used by the National Metrological Centre of China (Wu *et al.* 2001), Aggregate Drought Index

doi: 10.2166/nh.2017.254

(Keyantash & Dracup 2004) and Effective Drought Index (Byun & Wilhite 1999). Each of these indices has its own strengths and weaknesses. Most of them are data demanding and have complex computational processes (Nalbantis & Tsakiris 2009).

Drought is defined as the lack of precipitation in all its forms (streamflow, snowmelt, reservoir level and groundwater level) which is observed in the land phase of the hydrological cycle. Among these variables, streamflow is the most remarkable variable from the viewpoint of water quantity (Nalbantis & Tsakiris 2009). Thus, the role of precipitation and streamflow, as the most common variables in drought research, must be highlighted (Kazemzadeh & Malekian 2016). Considerable research in the realm of hydrological drought assessment has been conducted with the application of Streamflow Drought Index (SDI) introduced by Nalbantis & Tsakiris (2009) and Standardized Precipitation Index (SPI) developed by McKee *et al.* (1993), which are very simple in calculation and prevalent in on-site drought assessments (Moreira *et al.* 2008; Hosseini-Moghari & Araghinejad 2015; Portela *et al.* 2015; Zhu *et al.* 2016). Furthermore, Guttman (1999) compared the PDSI and SPI through a spectral analysis and recommended SPI as a more useful drought index because it has the capability to recognize the importance of time scales in the analysis of water availability and water use, so it can be used in risk assessment and decision-making. Tabari *et al.* (2013, 2015) and Nalbantis & Tsakiris (2009) used SDI in monthly time scales of 3, 6, 9, and 12 months and assessed drought severity based on streamflow data. Arabzadeh *et al.* (2015) concluded that SDI can accurately identify the expected frequency of drought in seasonal time scales.

As drought has complicated spatiotemporal patterns, the technique of principal component analysis (PCA) is vital to assess these patterns and summarize the existing variability in several dependent variables into fewer principal components (Sigdel & Ikeda 2010; Raziei *et al.* 2013). Arabzadeh *et al.* (2015) used PCA method to summarize the SDI index and concluded that the first principal component has the highest eigenvalue for seasonal time scales. Similarly, Bazrafshan *et al.* (2014) gained the same results in PCA in monthly time scales. PCA has been used widely in different studies as a multivariate method, while it merely removes the covariance in multivariate time series, without necessarily ensuring complete independence (Oja 2004). Independent component analysis, (ICA), introduced by Herault & Jutten (1986), is considered as an extension of PCA (Oja 2004). It has the capability to use an approximation of mutual information to minimize the higher-order dependence between the transformed series and to open up the opportunity to search for statistically mutually independent components (Ndehedehe *et al.* 2016). Westra *et al.* (2007) compared PCA and ICA for modelling multivariable hydrological series in Colombia. They concluded that significant benefits exist in maximizing statistical independence by ICA, in comparison to merely removing correlation by the PCA method. The singular value decomposition (SVD) technique aims to relate the sets of data by decomposing their covariance matrix into singular values and define a set of paired-orthogonal vectors based on their corresponding spatial and temporal patterns (Barnston & Smith 1996; Von Storch & Zwiers 2001). In hydrology, it has been significantly applied to evaluate the inter-relationships between various variables with spatio-temporal features (Shabbar & Skinner 2004; Lipovetsky & Conklin 2005; Lipovetsky 2009; Chitsaz *et al.* 2015).

This literature review illustrates a gap in assessing the correlation between SPI and SDI as the most common drought indices in both monthly and seasonal time scales by means of aggregation methods. In addition, a comparison among different multivariate methods is a crucial prerequisite for data decomposing, according to the diversity of results in finding the best multivariate method in previous research. Therefore, the major challenges in this study are attributed to assessing drought indices correlation along with compressing the large datasets into independently standardized components. In this regard, SPI and SDI indices were used, which are most frequently applied in different fields of drought. Eleven rain gauge stations and nine hydrometric stations were used for precipitation and streamflow data, and were decomposed to new datasets by multivariate methods.

The major innovation of this study is related to evaluating the capabilities of some well-known multivariate methods such as PCA, ICA, SVD, and averaging method, not only for determining the coupled relationship between SPI and SDI indices but also for decomposing the dimensions of input data and determining the best multivariate method. Furthermore, multivariate methods are compared by means of eigenvalues in different components to define the method which can accurately reflect the fluctuation and variation in the original data and confirm higher correlation among SPI and SDI indices. Also, sensitivity analysis of the proposed indices in different sets of SPI and SDI time scales (monthly and seasonal) has been performed.

STUDY AREA AND DATA

Case study

Iran is located in the mid-latitude belt of arid and semi-arid regions. Thus, drought periods and their effects are of great concern to politicians and planners (Raziei *et al.* 2009). The case study of this article, Karkheh River basin, is located in the southwest of Iran and covers an area of about 51,000 km^2 bounded by 30–35°N latitude and 46–49°E longitude (Figure 1). The average annual discharge changes from 3.3 to 86 m^3/s and the maximum discharge is 190.6. The average annual precipitation varies between 300 and 800 mm. The rivers Gamasiab, Qarasou, Saymareh and Kashkan are tributaries of the Karkheh River (Masih *et al.* 2009).

The datasets for the current research include precipitation in 11 rain gauge stations (six in the Saymareh sub-basin and the remainder in the Karkheh sub-basin) and the natural streamflow to the Karkheh Dam in nine hydrometric stations with 39 years of statistical periods from 1968–69 to 2006–2007. Water in the basin is mainly used

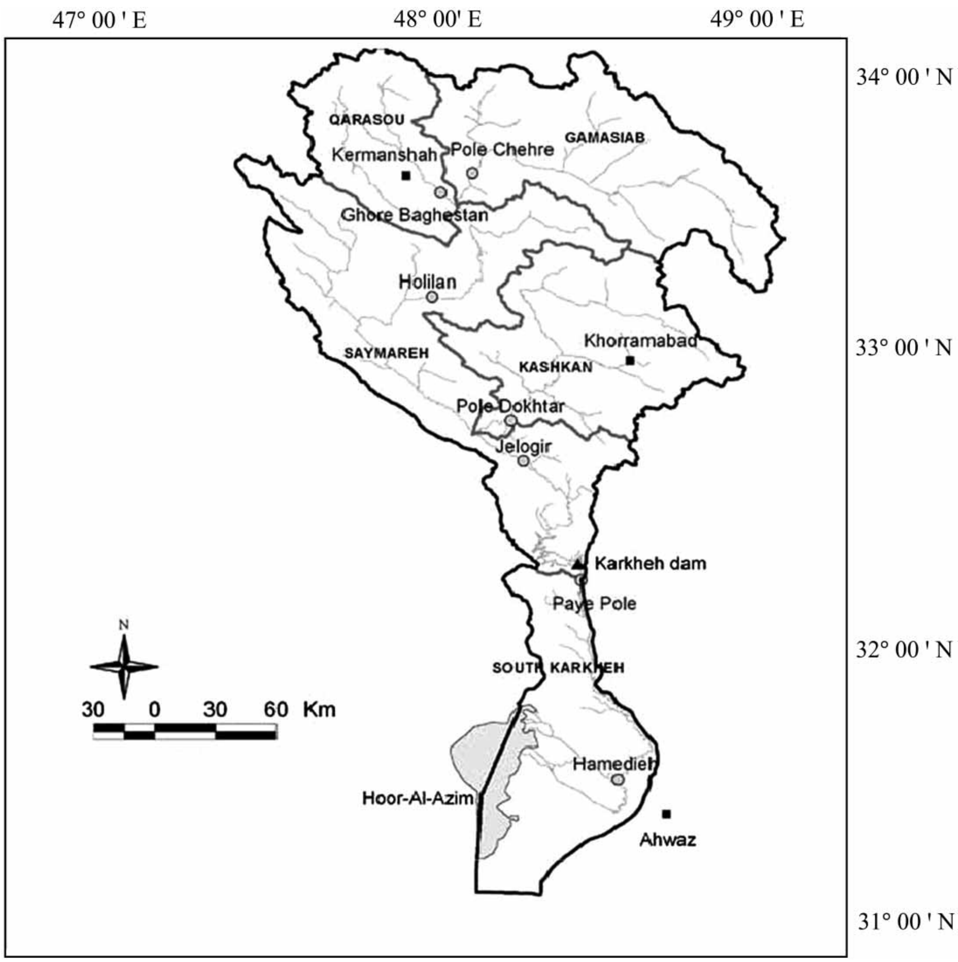

Figure 1 | Map of the case study.

for agriculture production, domestic supplies, fish farming, and serves to sustain the environment.

METHODS

The following subsections define the material and methods and involve a brief description of SPI and SDI drought indices for monthly and seasonal time scales, some stochastic tests to assess the trend in datasets, using multivariate methods intended for introduction of new datasets, and choosing the best multivariate method based on assessment parameters. Figure 2 illustrates the flowchart of the computational framework.

Drought indices

SPI

To calculate the SPI, precipitation dataset was prepared for 39 years at different time scales. Each of the datasets was fitted to the gamma distribution which fits well to the climatological precipitation time series. The two-parameter gamma probability density function is calculated as Equation (1):

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad x \succ 0 \tag{1}$$

where α and β are the shape and scale parameters, respectively. Parameter x is the precipitation amount and $\Gamma(\alpha)$ is the gamma function as Equation (2):

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy \tag{2}$$

Parameters α and β are estimated for each station in each time scale. Maximum likelihood estimations of α and β are according to Equation (3):

$$\widehat{\alpha} = \frac{1}{4A}\left[1 + \sqrt{1 + \frac{4A}{3}}\right], \quad \widehat{\beta} = \frac{\bar{x}}{\widehat{\alpha}} \tag{3}$$

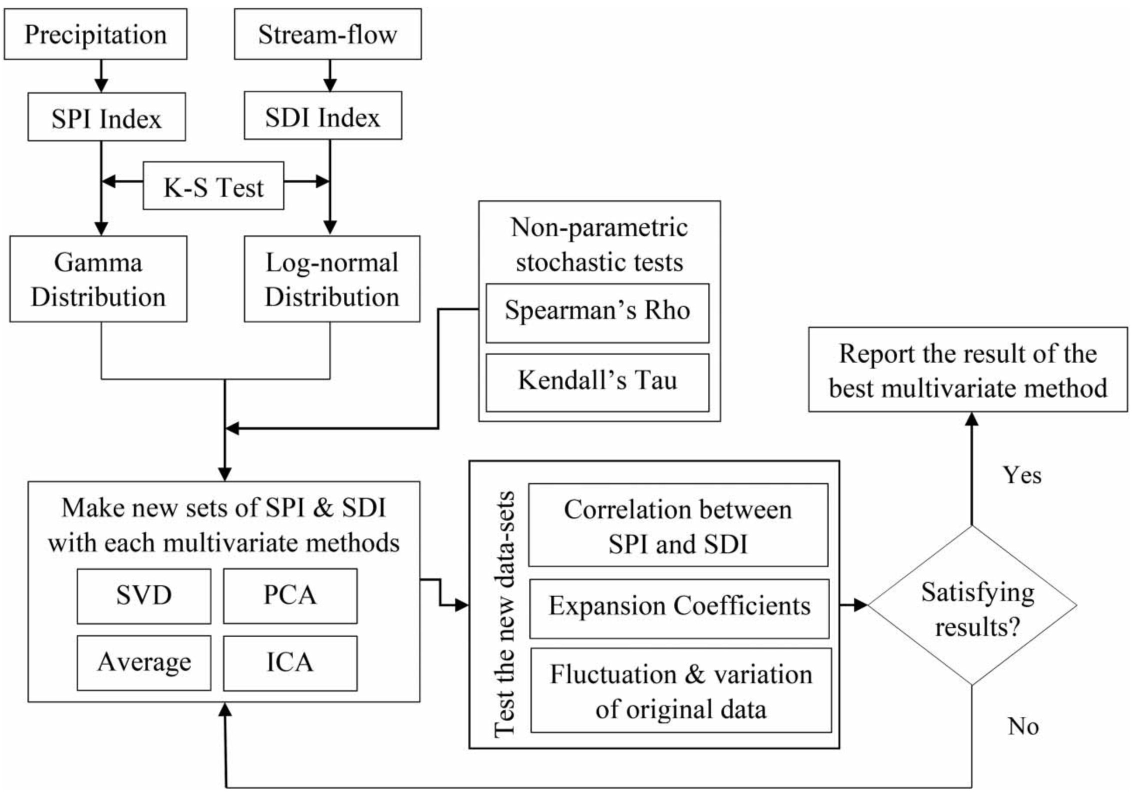

Figure 2 | Flowchart of the computational framework.

where $A = \ln(\bar{x}) - \sum \ln(x)/n$ and n is the number of precipitation dataset. The resulting parameters are used to find the cumulative probability. If the gamma function is undefined for $x = 0$ and precipitation distribution may contains zeroes, the cumulative probability is calculated as Equation (4):

$$H(x) = q + (1 - q)F(x) \tag{4}$$

where q is the probability of zero precipitation and $F(x)$ is the cumulative probability of the incomplete gamma function which is transformed to the standard normal random variable z with mean and variance equal to zero and one, respectively (Thom 1958; Abramovitz & Stegun 1965). Table 1 shows the classification of drought severity based on the SPI (Mckee *et al.* 1993).

SDI

In calculation of SDI, Log-normal distribution fits well to the hydrological streamflow time series as Equation (5):

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{\frac{-(\ln x - \mu)^2}{2\sigma^2}\right\} \tag{5}$$

where $\mu = \ln\left(m/\sqrt{1 + v/m^2}\right)$, $\sigma = \sqrt{\ln(1 + v/m^2)}$, m and v are mean and variance of streamflow dataset, respectively.

Table 1 | Drought classification based on the SPI and SDI (Mckee *et al.* 1993; Nalbantis & Tsakiris 2009)

Category	**SPI value**
Extremely wet	$2 \leq$ SPI
Very wet	$1.5 \leq$ SPI ≤ 1.99
Moderately wet	$1 \leq$ SPI ≤ 1.49
Near normal	$-0.99 \leq$ SPI ≤ 0.99
Moderately dry	$-1.0 \leq$ SPI ≤ -1.49
Severely dry	$-1.5 \leq$ SPI ≤ -1.99
Extremely dry	SPI ≤ -2
Category	**SDI value**
No drought	$0 \leq$ SDI
Mild drought	$-1 \leq$ SDI ≤ 0
Moderate drought	$-1.5 \leq$ SDI ≤ -1
Severe drought	$-2 \leq$ SDI ≤ -1.5
Extreme drought	SDI ≤ -2

Classification of SDI values is illustrated in Table 1 (Nalbantis & Tsakiris 2009).

Non-parametric stochastic tests

Spearman's rho test

The Spearman correlation coefficient of the linear regression between series of i and j is obtained as Equation (6):

$$r = 1 - \frac{[6\sum(i - j)^2]}{n(n^2 - 1)} \tag{6}$$

where n is the number of data items. For n (sample size) >30, the distribution will be normal and the test statistic z is computed as Equation (7):

$$z = r\sqrt{n - 1} \tag{7}$$

If $|z| \succ z_a$ at the significance level of a, the null hypothesis of no trend is rejected (Yue *et al.* 2002; Kahya & Kalaycı 2004).

Kendall's tau test

Kendall's correlation is commonly used to assess the significance trends in hydrometeorological time series (Kendall *et al.* 1968). According to this test, the null hypothesis *H0* assumes that there is no trend and this is tested against the alternative hypothesis *H1*, which assumes that there is a trend. The Kendall s statistic is computed as Equation (8):

$$s = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(T_j - T_i),$$
$$sign(T_j - T_i) = \begin{cases} 1 & \text{if } T_j - T_i \succ 0 \\ 0 & \text{if } T_j - T_i = 0 \\ -1 & \text{if } T_j - T_i \prec 0 \end{cases} \tag{8}$$

The standard test statistic z is calculated as Equation (9):

$$z = \begin{cases} \frac{s - 1}{\sigma} & \text{for } s \succ 0 \\ 0 & \text{for } s = 0 \\ \frac{s + 1}{\sigma} & \text{for } s \prec 0 \end{cases} \tag{9}$$

where σ^2 is the variance for s. This test statistic is used to test the null hypothesis, $H0$. If $|z| \succ z_{a/2}$ then the null hypothesis is invalid and the trend is significant.

Multivariate methods

ICA

ICA method is based on maximization of the output entropy or minimization of mutual information between the outputs. It aims to decompose the time series of data in matrix $x_i(t)$, which consists of a mixing matrix A and a number of statistically independent source signals $s_j(t)$ as Equation (10) (Herault & Jutten 1986; Ziehe 2005):

$$x_i(t) = \sum_{j=1}^{m} A_{i,j} s_j(t) \quad i = 1, \ldots, N, \quad j = 1, \cdots M \tag{10}$$

where t is the time index and x_i is the mixing model which can be represented as the matrix in Equation (11):

$$X = AS \tag{11}$$

where data in matrix X are samples of the $x_i(t)$, the $N \times M$ matrix A has elements $A_{i,j}$ and matrix S is analogous to the construction of X. According to Equation (11), ICA is based on factoring the observed signals data matrix X into the mixing matrix A and the source signals matrix S.

PCA

The introduction of PCA follows the familiar route of making successive orthogonal linear combinations of the variables with maximum variance. The main goal is to reduce the dimensionality of a dataset with a large number of interrelated variables, provided that the variation in the dataset remains as much as possible. The PCA approach uses all of the original variables to obtain a set of principal components in lower dimensions which are uncorrelated and ordered but can be used to approximate the original variables (Hyvärinen *et al.* 2004; Westra *et al.* 2007). Consider a dataset as matrix X and the linear transformed orthogonal matrix Z according to Equation (12):

$$Z = XA \tag{12}$$

where Z has (i, j) dimension with i_{th} observation and j_{th} principal component; A is $p \times p$ matrix with eigenvector elements of the covariance of X, and having $A^T A = AA^T = I$ and $X^T X = A \Lambda A^T$ where Λ is a diagonal matrix whose non-negative entries are the eigenvalues of $X^T X$.

SVD

The SVD method can decompose any $n \times m$ matrix as A. The first step is to find U, S and V via Equation (13) (Bjornsson & Venegas 1997):

$$A = U \times S \times V^t \tag{13}$$

where U is an $n \times n$ orthonormal matrix, V is an $m \times m$ orthonormal matrix and S is a diagonal $n \times m$ matrix with ρ elements down the diagonal and $\rho \leq \min(n, m)$. V is the matrix of column vectors which are eigenvectors of $C = A^t A$. U is the matrix of projections of X onto the eigenvector of C. S is a non-square matrix with zero entries everywhere, except on the leading diagonal with elements S_i in descending order of magnitude. Each S_i is equal to the square root of the eigenvalues of $C = A^t A$.

Based on the above information, all of the multivariate methods have the goal of finding a set of low dimensional and uncorrelated data which are independent to simplified computational structures. There are some differences between multivariate methods. ICA uses a kurtosis measure (the fourth certain moment), which is based on maximizing non-Gaussianity to maximize the independency (Westra *et al.* 2007). In PCA, an orthogonal transformation based on variance (the second order moment) is used to change data into linearly uncorrelated elements. It is possible to combine two or more variables in the same PCA but some research has proven that SVD is superior to combined PCA. SVD is a robust statistic technique, which isolates linear combinations of variables to obtain coupled relationships between two spatiotemporal fields that tend to be linearly related to one another. It is a fundamental matrix operation, a generalization of the diagonalization procedure in PCA to matrices that are not square or symmetric (Lipovetsky 2009).

Expansion coefficient assessment

Expansion coefficient (EC) is used to find the spatial patterns of variability and gives a measure of the importance of each pattern as assessed by Equation (14) (Bjornsson & Venegas 1997):

$$EC_i = E_i^T X = \sum_{j=1}^{k} e_{ij} X_k \tag{14}$$

where EC_i is the i_{th} expansion coefficient, E_i^T is the i_{th} eigenvector, X_k is the k_{th} original variable, and e_{ij} is the k_{th} element of the i_{th} eigen component. The first expansion coefficient equivalent to EC_1, justifies a large amount of variation of hyper-cloud buildup by the variables (Sharma 1996). The eigenvalues corresponding to each EC are calculated using Equation (15):

$$|C - \lambda I| = 0 \tag{15}$$

where C is correlation matrix of original data, λ is eigenvalue and I is identity matrix. As the eigenvalues were calculated, the eigenvector will be formed to correspond to each eigenvalue.

RESULTS AND DISCUSSION

Trend and correlation tests for meteorological and hydrological drought

Due to the existence of perennial and seasonal rivers in the case study watershed, meteorological and hydrological droughts were evaluated based on SPI and SDI, respectively. These analyses were assessed in the seasonal time scales (fall, winter, spring and summer) and five monthly time series (3, 6, 9, 12 and 24 months). According to the Kolmogorov–Smirnov (K-S) test, at the 0.05 significance level, gamma and log-normal distributions provided an adequate fit to the precipitation and streamflow series which are shown via Equations (1)–(5) for SPI and SDI, respectively.

In this section, the trends of meteorological and hydrological drought are analysed by means of two non-parametric stochastic tests, Spearman's rho and Kendall's tau, over the period of 1968–69 to 2006–07 in both monthly and seasonal time series, as shown in Table 2. The trends of SPI and SDI are illustrated based on their p value and correlation coefficient in both statistical tests. Trend analysis for SPI indicates that the 3-month and 24-month periods have the lowest and the highest trend, respectively, based on their p values. In other words, this negative trend in long time scales of SPI is higher than short time scales and the correlation is significant at the 0.01 intervals for the 24-month period. The p value for SDI suggests a negative trend for all monthly time series at 0.01 significance level, which shows a severe decreasing trend based on Spearman's rho correlation coefficient. A comparison between p values for SPI and SDI trends in monthly time series proves that the negative trend of the hydrological drought is stronger than the meteorological drought. The p values of the statistical methods for seasonal time series showed that the trend in meteorological drought does not have a significant correlation at 0.01 intervals, while a severe trend occurs in hydrological drought in summer. The highest trend for SPI and SDI in monthly time series (24 months) is illustrated as an example in Figure 3.

SDI and SPI have reflected large fluctuations on one hand, as shown in Figure 3, and a huge amount of input data as 11 rain gauge stations and nine hydrometric stations on the other hand. Owing to this, the multivariate methods (ICA, PCA and SVD) were used in addition to average values of input data to reduce the dimensionality of input data. The new data series which is released by the best multivariate method institutes a higher correlation between SDI and SPI as drought indices and reflects the exact fluctuation and trend of the original data.

Due to seasonal and monthly time scales, these assessments were done based on two scenarios. In the first scenario, the correlation coefficients between SDI and SPI were assessed for seasonal time scales in each hydrological year, and for the second scenario this stochastic relationship was used for drought indices in overlapping periods of 3, 6, 9, 12 and 24 months.

Table 2 | The *p* values of Spearman's rho and Kendall's tau for SDI and SDI

	3 Month	6 Month	9 Month	12 Month	24 Month	Fall	Winter	Spring	Summer
SPI									
Kendall's tau									
R	−0.025	−0.031	−0.044	−0.049	−0.120**	0.155	−0.131	0.049	0.283*
p value	0.207	0.155	0.076	0.058	0	0.164	0.241	0.663	0.030
Spearman's rho									
R	−0.038	−0.046	−0.067	−0.082*	−0.170**	0.2	−0.17	0.063	0.378*
p value	0.205	0.162	0.075	0.038	0	0.223	0.301	0.702	0.018
SDI									
Kendall's tau									
R	−0.188**	−0.187**	−0.177**	−0.170**	−0.254**	−0.174	−0.077	−0.163	−0.287**
p value	0	0	0	0	0	0.059	0.245	0.072	0.005
Spearman's rho									
R	−0.280**	−0.282**	−0.281**	−0.283**	−0.386**	−0.276*	−0.132	−0.257	−0.413**
p value	0	0	0	0	0	0.044	0.211	0.057	0.004

*Correlation is significant at the 0.05 intervals.
**Correlation is significant at the 0.01 intervals.

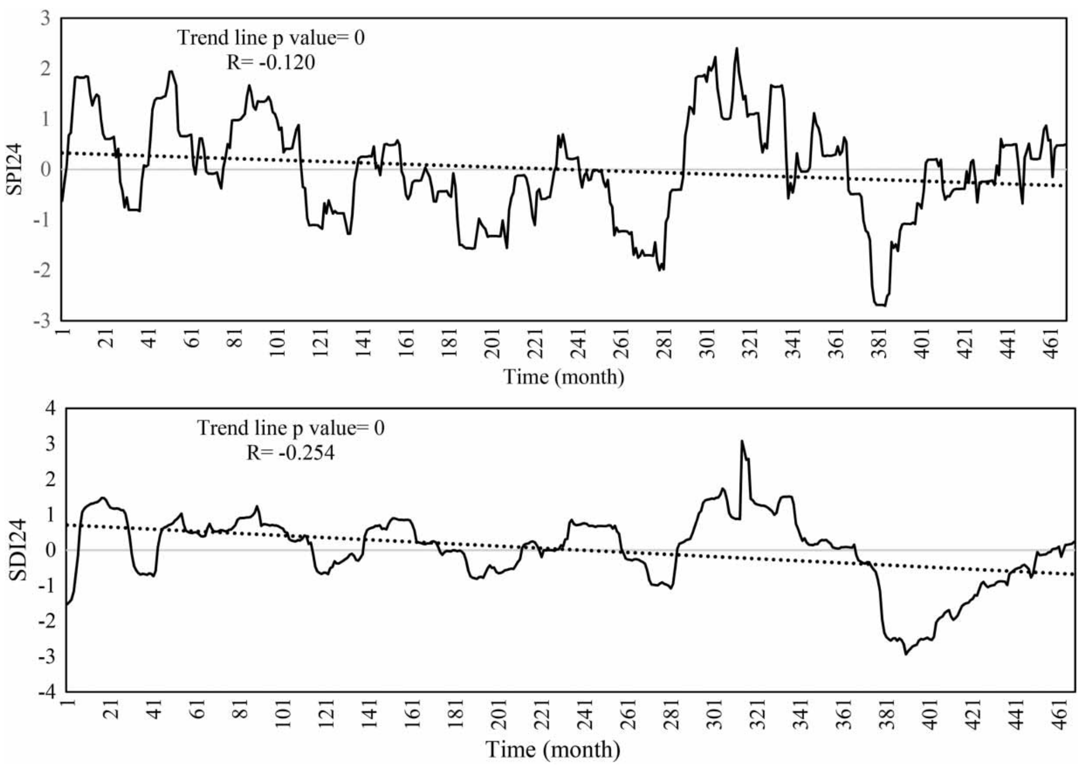

Figure 3 | Trend in monthly time series for SPI24 and SDI24.

Multivariate methods for seasonal time scales

Table 3 illustrates the results of correlation coefficient between SPI and SDI in different seasons for average and multivariate methods. SPI and SDI data in all stations were aggregated to one series of SPI and SDI for each time period with multivariate methods. Results show that performing multivariate methods to derive aggregated indices in all of the time series was effective. The most prominent results to emerge from Table 3 are related to finding the suitable multivariate methods for aggregating drought indices. According to the correlation values, SVD and PCA results outweigh average and ICA methods.

Another striking point is that specific SPI with higher correlation with each SDI in different seasons demonstrates the time delay in meteorological and hydrological droughts. For fall, winter and spring the highest correlation between SDI and SPI is in the same seasons, which means that meteorological drought and hydrological drought occur with no lag time in seasonal time scale. There is an exception in released data for summer SDI, which has the highest correlation with winter and spring SPI. This outcome stems from lack of precipitation in summer. Thus, summer hydrological drought is related to meteorological drought in winter and spring. Throughout the seasonal drought assessment in Sefid-Rud basin in the northwest of Iran, Arabzadeh *et al.* (2015) asserted that SDI and multivariate SDI have lower correlation for spring among other seasons, but there was no correlation assessment of SPI and SDI.

EC for seasonal drought indices

The results of ECs based on eigenvalue for two selected multivariate methods, SVD and PCA, for seasonal drought indices are shown in Figure 4. Figure 4(c) shows the contribution of each variant in forming the SVD results. The ECs in the SVD method come from squared diagonals of matrix S (matrix of singular eigenvalues). As the SVD method is based on the eigenvalues of the covariance matrix $SPI \times SPI$, the squared matrix S is in a 9×9 dimension. EC1 has the highest eigenvalue for all seasons; in consequence, it is able to justify a great part of the total variance in the original variables and can explain the most percentage of variability in the original SPI and SDI data. According to Figure 4(c), EC1 has less contribution for fall and winter, of about 70% and 64%, respectively, while spring and summer have a high eigenvalue of about 81%. The reason for the increase in the EC1 for summer and spring is the lower variety of SDI and SPI indices existing in this range, which represents the lower variability in seasonal precipitation and streamflow data, as shown in Figure 5. In contrast, the ECs in the PCA method cannot justify the fluctuations in SPI and SDI data. Based on Figure 4(a) and 4(b), the higher eigenvalues for SPI are in fall and summer, of about 73% and 75%, and the lower values are in winter and spring, of about 60%. Moreover,

Table 3 | SPI and SDI correlation coefficient in seasonal time scales

	SVD				**PCA**			
	SPI Fall	**SPI Winter**	**SPI Spring**	**SPI Summer**	**SPI Fall**	**SPI Winter**	**SPI Spring**	**SPI Summer**
SDI Fall	0.690	0.207	0.053	0.023	0.677	0.224	0.049	0.006
SDI Winter	0.324	0.650	0.168	0.090	0.312	0.632	0.161	0.051
SDI Spring	0.122	0.648	0.755	0.103	0.118	0.645	0.748	0.105
SDI Summer	0.172	0.617	0.575	0.020	0.166	0.609	0.568	0.090
	Average				**ICA**			
SDI Fall	0.609	0.259	0.094	0.026	0.367	0.005	0.077	0.028
SDI Winter	0.326	0.483	0.294	0.053	0.234	0.244	0.151	0.076
SDI Spring	0.011	0.564	0.554	0.078	0.028	0.291	0.607	0.066
SDI Summer	0.079	0.561	0.474	0.055	0.039	0.360	0.471	0.061

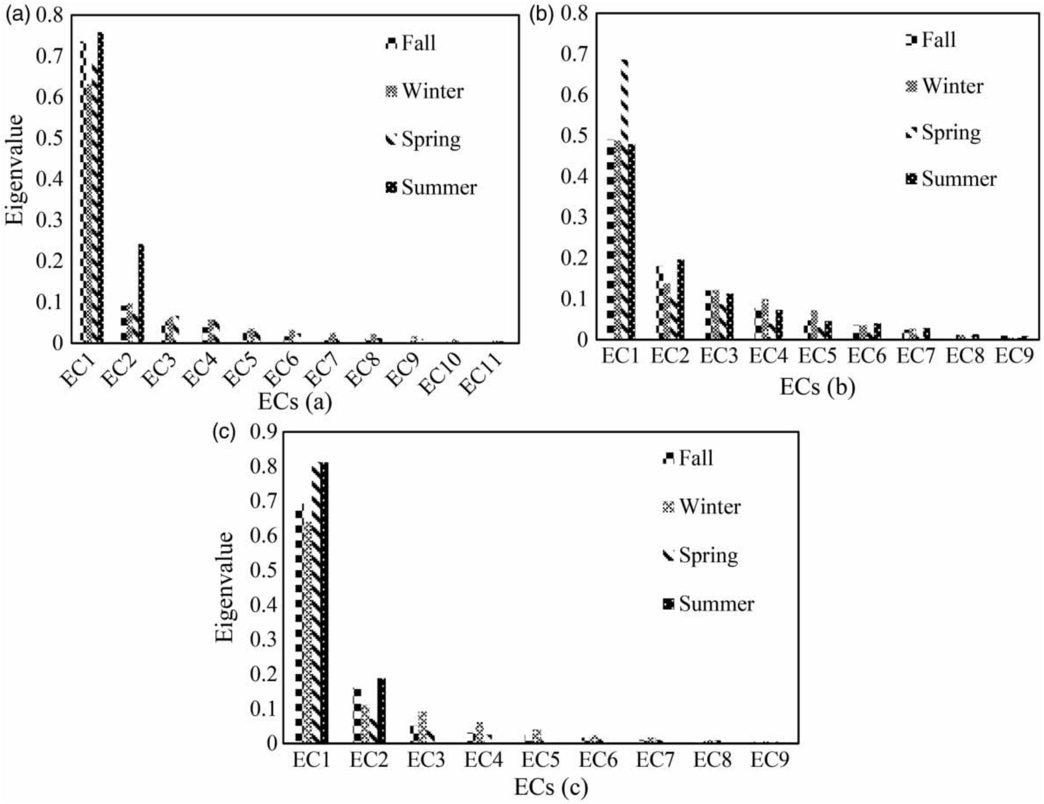

Figure 4 | Eigenvalues for ECs in seasonal time scale: (a) PCA in SPI, (b) PCA in SDI, (c) SVD.

most of the eigenvalues for SDI are in a low range of about 47% to 68%.

In Shabbar & Skinner's (2004) research, most of the variance of drought indices is concentrated in the first mode of the SVD expansion. The first principal component of the PCA method has the greatest part of information in a set of SPI and SDI data in some studies (Bazrafshan *et al.* 2014; Arabzadeh *et al.* 2015; Azmi *et al.* 2016). Thus, the diversity of results in finding the best aggregated dataset among the aforementioned studies shows the comparison of different aggregation methods based on the eigenvalues as an appropriate idea.

Multivariate methods for monthly time series

Table 4 illustrates the results of correlation coefficient between SPI and SDI for 3, 6, 9, 12 and 24 months for average and multivariate methods. SVD, PCA and average methods have a remarkable correlation (more than 0.7) for each SDI, while ICA has no satisfaction results (lower than 0.5). The higher correlation for each SDI asserts that SDI3 has the highest correlation with SPI9, while SDI6 and SDI9 have the highest correlation with SPI12. The correlation for SDI12 and SDI24 has the highest value in the same time period for SPIs. In other words, a lower time scale such as SDI3 has more correlation with SPI9, and SPI12 has the highest correlation for higher time series such as SDI6 to SDI12. Nalbantis & Tsakiris (2009) reported the same results, high correlation between SPI and SDI in the same time series for 9 and 12 months.

Unlike the current research, Westra *et al.* (2007) concluded that the ICA method performed better than the PCA method based on discrepancy between the synthetically generated data and the original data in monthly datasets of reservoir inflows from Colombia. Ndehedehe *et al.* (2016) employed the ICA method to decompose SPI in drought spatiotemporal variability monitoring. However, PCA delivered appropriate results in Bazrafshan *et al.'s* (2014) study. Considering these three common techniques

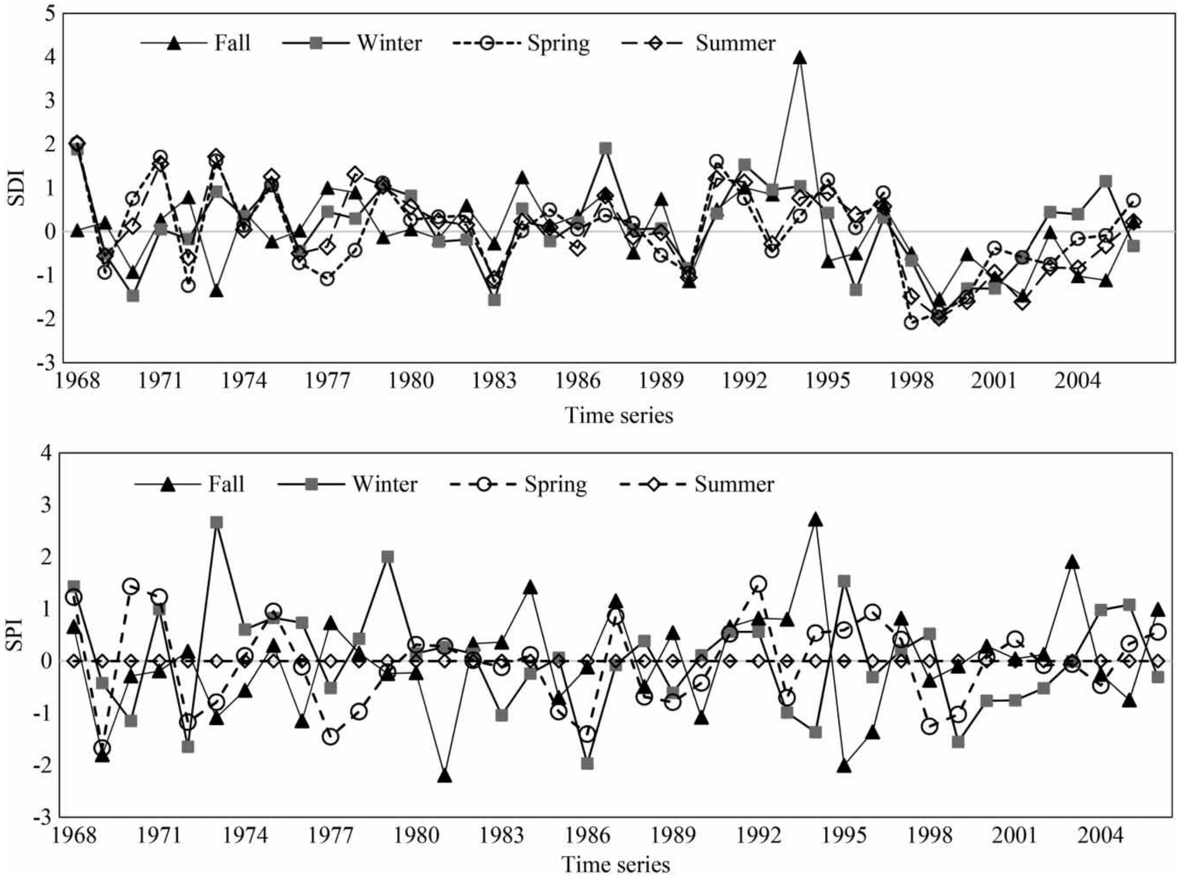

Figure 5 | Fluctuations in SDI and SPI in seasonal time scale.

Table 4 | SPI and SDI correlation coefficient in monthly time scales

	SVD					**PCA**				
	SPI3	**SPI6**	**SPI9**	**SPI12**	**SPI24**	**SPI3**	**SPI6**	**SPI9**	**SPI12**	**SPI24**
SDI3	0.562	0.719	0.776*	0.759	0.627	0.549	0.710	0.769*	0.751	0.615
SDI6	0.388	0.682	0.792	0.805*	0.672	0.376	0.672	0.785	0.796*	0.659
SDI9	0.303	0.521	0.738	0.828*	0.707	0.293	0.512	0.732	0.821*	0.695
SDI12	0.211	0.397	0.572	0.771*	0.753	0.200	0.387	0.565	0.762*	0.739
SDI24	0.130	0.225	0.316	0.423	0.801*	0.118	0.213	0.307	0.411	0.784*
	Average					**ICA**				
SDI3	0.525	0.696	0.750*	0.720	0.564	0.416	0.316	0.480	0.467	0.227
SDI6	0.349	0.656	0.762	0.763*	0.604	0.280	0.278	0.491	0.495	0.239
SDI9	0.280	0.499	0.711	0.790*	0.647	0.208	0.197	0.462	0.483	0.210
SDI12	0.199	0.376	0.544	0.727*	0.684	0.124	0.139	0.340	0.459	0.227
SDI24	0.124	0.199	0.286	0.380	0.716*	0.062	0.024	0.139	0.308	0.249

*Data with high correlation coefficients.

along with an averaging method, one can determine the most robust technique in this case study.

EC for monthly drought indices

The results of ECs based on eigenvalue for two selected multivariate methods, SVD and PCA, for monthly drought indices are shown in Figure 6. It is obvious that most of the dataset features can be explained by EC1 and other ECs only cover some of the data. In the SVD technique, EC1 has a high eigenvalue for all monthly time series of around 80%, while PCA has a lower eigenvalue for both SPI and SDI data with a range of eigenvalues from 53% to 70% according to Figure 6(a) and 6(b). A glance at Figure 7, which shows the original SDI and SPI data fluctuations, reveals a clearer assessment for these two multivariate methods. Since SPI and SDI data have the same fluctuation for all five monthly time series, the superiority of SVD over PCA is asserted. Hence, due to its higher eigenvalue, SVD again shows more compatibility with SPI and SDI data in monthly time series as well as seasonal time scales.

CONCLUSION

In the current research, multivariate methods (SVD, PCA and ICA) and averaging technique are used to create new data series for drought indices. The innovation of this research is a comparative framework to evaluate multivariate methods based on different criteria assessment. The main features of these multivariate evaluations are: (1) making a standardized data series which simplifies the ambiguity of numerous data series of SPI and SDI to obtain new drought indices; (2) creating new drought data series which can accurately demonstrate fluctuations and variations in the original data and confirm higher correlation between SPI and SDI as the most common drought indices; and (3)

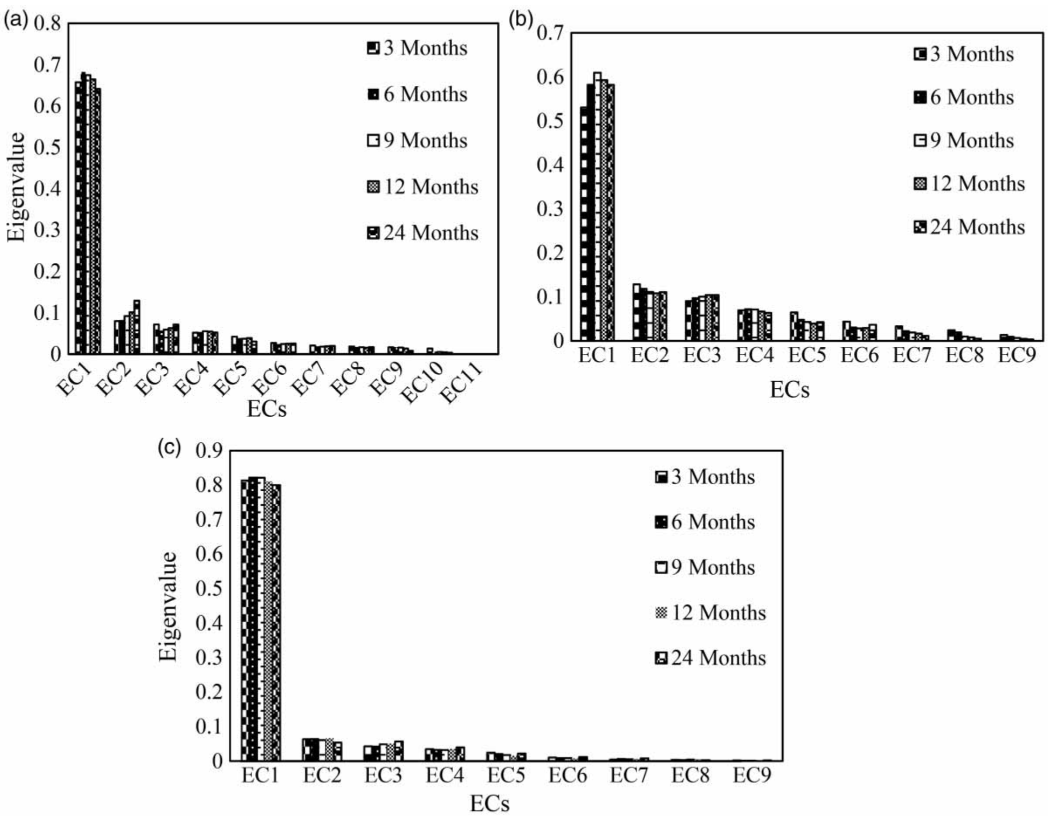

Figure 6 | Eigenvalues for ECs in monthly time scale: (a) PCA in SPI, (b) PCA in SDI, (c) SVD.

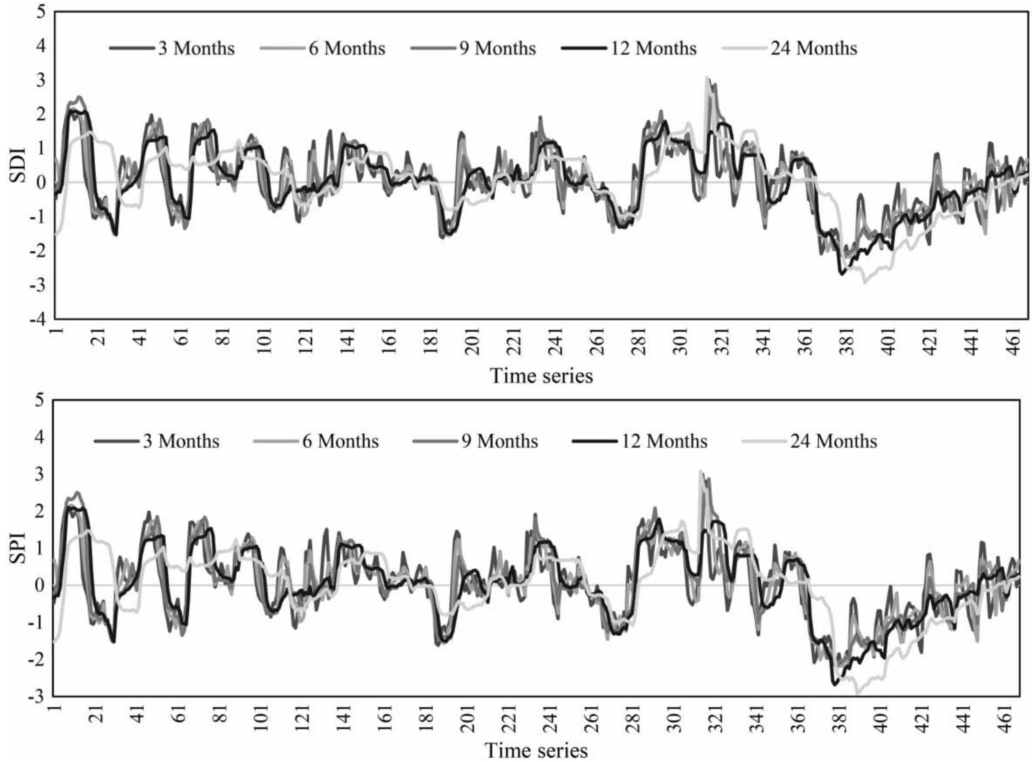

Figure 7 | Fluctuations in SDI and SPI in monthly time scale.

including the EC's operational based standardized values, which are expected to have higher eigenvalues in order to contain a large part of the original data. This study asserts the superiority of the SVD technique which is aptly suited for the above expected features over other multivariate methods.

The *p* values of Spearman's rho and Kendall's tau showed that the study area has suffered from hydrological drought more than meteorological drought. The SVD technique showed more capability of finding the coupled pattern for SPI and SDI, which has the highest correlation coefficients for both monthly and seasonal time scales. The higher eigenvalue of about 80% in spring, summer and all monthly time scales for the SVD method asserts its dominance over the PCA method, with a varied eigenvalue from 53% to 70%. Additionally, the SVD technique reflected the higher fluctuation of fall and winter datasets by lower eigenvalues for these seasons, however PCA could not demonstrate these fluctuations. Thus, SVD explains the greatest percentage of variability in the original SPI and SDI data, better than PCA as the most common multivariate method in drought indices' assessment research.

ACKNOWLEDGEMENTS

The authors express their appreciation to the Associate Editor Prof. Chong-Yu Xu and two anonymous reviewers for their constructive suggestions and helpful comments on a previous version of the manuscript and for improving the quality of this paper.

REFERENCES

Abramovitz, M. & Stegun, I. 1965 *Handbook of Mathematical Functions, Applied Mathematics Series 55*. National Bureau of Standards, Dover Publications, New York.

Arabzadeh, R., Kholoosi, M. M. & Bazrafshan, J. 2015 Regional hydrological drought monitoring using principal components analysis. *Journal of Irrigation and Drainage Engineering* **142** (1), 04015029. doi: 10.1061/(ASCE)IR.1943-4774.0000925, 04015029

Azmi, M., Rüdiger, C. & Walker, J. P. 2016 A data fusion-based drought index. *Water Resources Research* **52** (3), 2222–2239. doi: 10.1002/2015WR017834

Barnston, A. G. & Smith, T. M. 1996 Specification and prediction of global surface temperature and precipitation from global SST using CCA. *Journal of Climate* **9** (11), 2660–2697.

Bazrafshan, J., Hejabi, S. & Rahimi, J. 2014 Drought monitoring using the multivariate standardized precipitation index (MSPI). *Water Resources Management* **28** (4), 1045–1060. doi:10.1007/s11269-014-0533-2

Bernard, B., Vincent, K., Frank, M. & Anthony, E. 2013 Comparison of extreme weather events and streamflow from drought indices and a hydrological model in River Malaba, Eastern Uganda. *International Journal of Environmental Studies* **70** (6), 940–951. doi: 10.1080/00207233.2013.862463

Bjornsson, H. & Venegas, S. A. 1997 A manual for EOF and SVD analyses of climatic data. *CCGCR Report* **97** (1), 112–134.

Byun, H. R. & Wilhite, D. A. 1999 Objective quantification of drought severity and duration. *Journal of Climate* **12** (9), 2747–2756.

Chitsaz, N., Azarnivand, A. & Araghinejad, S. 2015 Pre-processing of data-driven river flow forecasting models by singular value decomposition (SVD) technique. *Hydrological Sciences Journal* **61** (12), 2164–2178. doi:org/10.1080/02626667.2015.1085991

Guttman, N. B. 1999 Accepting the standardized precipitation index: a calculation algorithm. *Journal of the American Water Resources Association* **35** (2), 311–322.

Heim Jr, R. R. 2002 A review of twentieth-century drought indices used in the United States. *Bulletin of the American Meteorological Society* **83** (8), 1149–1165.

Herault, J. & Jutten, C. 1986 Space or time adaptive signal processing by neural network models. *Neural Networks for Computing* **151** (1), 206–211.

Heudorfer, B. & Stahl, K. 2016 Comparison of different threshold level methods for drought propagation analysis in Germany. *Hydrology Research* doi: 10.2166/nh.2016.258

Hosseini-Moghari, S. M. & Araghinejad, S. 2015 Monthly and seasonal drought forecasting using statistical neural networks. *Environmental Earth Sciences* **74** (1), 397–412. doi:10.1007/s12665-015-4047-x

Hyvärinen, A., Karhunen, J. & Oja, E. 2004 *Independent Component Analysis*, Vol. 46. John Wiley & Sons, New York.

Kahya, E. & Kalaycı, S. 2004 Trend analysis of streamflow in Turkey. *Journal of Hydrology* **289** (1), 128–144. doi:10.1016/j.jhydrol.2003.11.006

Kazemzadeh, M. & Malekian, A. 2016 Spatial characteristics and temporal trends of meteorological and hydrological droughts in northwestern Iran. *Natural Hazards* **80** (1), 191–210. doi:10.1007/s11069-015-1964-7

Kendall, M. G., Stuart, A. & Ord, J. K. 1968 *The Advanced Theory of Statistics*, Vol. 3. Wiley-Blackwell, London, pp. 306–311.

Keyantash, J. A. & Dracup, J. A. 2004 An aggregate drought index: assessing drought severity based on fluctuations in the hydrologic cycle and surface water storage. *Water Resources Research* **40** (9), 1–13. doi:10.1029/2003WR002610.

Khadr, M. 2017 Temporal and spatial analysis of meteorological drought characteristics in the upper Blue Nile river region. *Hydrology Research* **48** (1), 265–276. doi: 10.2166/nh.2016.194.

Lennard, A. T., Macdonald, N., Clark, S. & Hooke, J. M. 2016 The application of a drought reconstruction in water resource management. *Hydrology Research* **47** (3), 646–659. doi: 10.2166/nh.2015.090

Li, X. H., Zhang, Q., Zhang, D. & Ye, X. C. 2017a Investigation of the drought–flood abrupt alternation of streamflow in Poyang Lake catchment during the last 50 years. *Hydrology Research* **48** (5), 1402–1417. doi: 10.2166/nh.2016.266.

Li, Y., Chen, C. Y. & Sun, C. F. 2017b Drought severity and change in Xinjiang, China, over 1961–2013. *Hydrology Research* **48** (5), 1343–1362. doi: 10.2166/nh.2016.026.

Ling, H. B., Deng, X. Y., Long, A. H. & Gao, H. F. 2017 The multi-time-scale correlations for drought–flood index to runoff and North Atlantic Oscillation in the headstreams of Tarim River, Xinjiang, China. *Hydrology Research* **48** (1), 253–264. doi: 10.2166/nh.2016.166.

Lipovetsky, S. 2009 PCA and SVD with nonnegative loadings. *Pattern Recognition* **42** (1), 68–76. doi: 10.1016/j.patcog.2008.06.025

Lipovetsky, S. & Conklin, W. M. 2005 Singular value decomposition in additive, multiplicative, and logistic forms. *Pattern Recognition* **38** (7), 1099–1110. doi: 10.1016/j.patcog.2005.01.010

Masih, I., Uhlenbrook, S., Turral, H. & Karimi, P. 2009 Analysing streamflow variability and water allocation for sustainable management of water resources in the semi-arid Karkheh river basin, Iran. *Physics and Chemistry of the Earth, Parts A/B/C* **34** (4), 329–340. doi:10.1016/j.pce.2008.09.006

McKee, T. B., Doesken, N. J. & Kleist, J. 1993 The relationship of drought frequency and duration to time scales. In: *Proceedings of the 8th Conference on Applied Climatology*, American Meteorological Society, Boston, MA, 17 (22), pp. 179–183.

Moreira, E. E., Coelho, C. A., Paulo, A. A., Pereira, L. S. & Mexia, J. T. 2008 SPI-based drought category prediction using loglinear models. *Journal of Hydrology* **354** (1), 116–130.

Nalbantis, I. & Tsakiris, G. 2009 Assessment of hydrological drought revisited. *Water Resources Management* **23** (5), 881–897. doi:10.1007/s11269-008-9305-1

Ndehedehe, C. E., Agutu, N. O., Okwuashi, O. & Ferreira, V. G. 2016 Spatio-temporal variability of droughts and terrestrial water storage over Lake Chad Basin using independent component analysis. *Journal of Hydrology*. doi:org/10.1016/j.jhydrol.2016.05.068.

Oja, E. 2004 Applications of independent component analysis. In: *International Conference on Neural Information Processing*, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 1044–1051. doi: 10.1007/978-3-540-30499-9_162.

Palmer, W. C. 1965 *Meteorological Drought*, Vol. 30. US Department of Commerce, Weather Bureau, Washington, DC.

Portela, M. M., Santos, J. F., Silva, A. T., Benitez, J. B., Frank, C. & Reichert, S. M. 2015 Drought analysis in southern Paraguay, Brazil and northern Argentina: regionalization, occurrence rate and rainfall thresholds. *Hydrology Research* **46** (5), 792–810. doi:10.2166/nh.2014.074

Raziei, T., Saghafian, B., Paulo, A. A., Pereira, L. S. & Bordi, I. 2009 Spatial patterns and temporal variability of drought in western Iran. *Water Resources Management* **23** (3), 439–455.

Raziei, T., Bordi, I. & Pereira, L. S. 2013 Regional drought modes in Iran using the SPI: the effect of time scale and spatial resolution. *Water Resources Management* **27** (6), 1661–1674.

Shabbar, A. & Skinner, W. 2004 Summer drought patterns in Canada and the relationship to global sea surface temperatures. *Journal of Climate* **17** (14), 2866–2880. doi: http://dx.doi.org/10.1175/1520-0442(2004)017 < 2866: SDPICA > 2.0.CO;2.

Shafer, B. A. & Dezman, L. E. 1982 Development of a surface water supply index (SWSI) to assess the severity of drought conditions in snowpack runoff areas. In: *Proceedings of the Western Snow Conference*, Vol. 50. Colorado State University, Fort Collins, CO, pp. 164–175.

Sharma, S. 1996 *Applied Multivariate Techniques*. Wiley, New York.

Sigdel, M. & Ikeda, M. 2010 Spatial and temporal analysis of drought in Nepal using standardized precipitation index and its relationship with climate indices. *Journal of Hydrology and Meteorology* **7** (1), 59–74.

Sung, J. H. & Chung, E.-S. 2014 Development of streamflow drought severity–duration–frequency curves using the threshold level method. *Hydrology and Earth System Sciences* **18**, 3341–3351. doi:10.5194/hess-18-3341-2014

Tabari, H., Nikbakht, J. & Talaee, P. H. 2013 Hydrological drought assessment in Northwestern Iran based on streamflow drought index (SDI). *Water Resources Management* **27** (1), 137–151.

Tabari, H., Zamani, R., Rahmati, H. & Willems, P. 2015 Markov chains of different orders for streamflow drought analysis. *Water Resources Management* **29** (9), 3441–3457. doi: 10.1007/s11269-015-1010-2

Thom, H. C. 1958 A note on the gamma distribution. *Monthly Weather Review* **86** (4), 117–122.

Vasiliades, L., Loukas, A. & Liberis, N. 2011 A water balance derived drought index for Pinios River Basin, Greece. *Water Resources Management* **25** (4), 1087–1101.

Von Storch, H. & Zwiers, F. W. 2001 *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge.

Westra, S., Brown, C., Lall, U. & Sharma, A. 2007 Modeling multivariable hydrological series: principal component analysis or independent component analysis? *Water Resources Research* **43** (6), 1–11. doi:10.1029/2006WR005617

Wu, H., Hayes, M. J., Weiss, A. & Hu, Q. 2001 An evaluation of the standardized precipitation index, the China-Z index and the statistical Z-Score. *International Journal of Climatology* **21** (6), 745–758.

Yang, W. 2010 Drought Analysis Under Climate Change by Application of Drought Indices and Copulas. MS thesis, Portland State University, Portland, OR.

Ye, X. C., Li, X. H., Xu, C.-Y. & Zhang, Q. 2016 Similarity, difference and correlation of meteorological and hydrological drought indices in a humid climate region – the Poyang Lake catchment in China. *Hydrology Research* doi: 10.2166/nh.2016.214.

Yoo, J., Kwon, H. H., Kim, T. W. & Ahn, J. H. 2012 Drought frequency analysis using cluster analysis and bivariate probability distribution. *Journal of Hydrology* **420**, 102–111.

Yue, S., Pilon, P. & Cavadias, G. 2002 Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology* **259** (1), 254–271.

Zhu, Y. L., Chang, J. X., Huang, S. Z. & Huang, Q. 2016 Characteristics of integrated droughts based on a nonparametric standardized drought index in the Yellow River Basin, China. *Hydrology Research* in press. doi:10.2166/nh.2015.287.

Ziehe, A. 2005 Blind Source Separation Based on Joint Diagonalization of Matrices with Applications in Biomedical Signal Processing. Doctoral dissertation, University of Potsdam, Potsdam, Germany.

First received 31 October 2016; accepted in revised form 8 February 2017. Available online 27 March 2017