# Composite Agrometeorological Drought Index Accounting for Seasonality and Autocorrelation

Mohammad Mehdi M Bateni, Javad Behmanesh, Carlo de Michele, Javad Bazrafshan, Hosein Rezaie

**Journal of Hydrologic Engineering-HE3881 (Accepted Manuscript)**

**doi: 10.1061/(ASCE)HE.1943-5584.0001654**

# Composite Agrometeorological Drought Index Accounting for Seasonality and Autocorrelation

M. M. Bateni[1], J. Behmanesh[2], C. De Michele[3], J. Bazrafshan[4], H. Rezaie[5]

*1* Water Engineering Department, University of Urmia, Urmia, IRAN (corresponding author). E-mail: mm.bateni@urmia.ac.ir

*2* Water Engineering Department, University of Urmia, Urmia, IRAN. E-mail: j.behmanesh@urmia.ac.ir

*3* Department of Civil and Environmental Engineering (DICA), Politecnico di Milano, Milano, Italy. E-mail: carlo.demichele@polimi.it

*4* Department of Irrigation and Reclamation Engineering, University of Tehran, Karaj, IRAN. E-mail: jbazr@ut.ac.ir

*5* Water Engineering Department, University of Urmia, Urmia, IRAN. E-mail: h.rezaie@urmia.ac.ir



**Abstract**

Drought indices are statistical tools used for monitoring the departure from normal conditions of water availability. Recently, the multivariate nature of droughts has been addressed through composite indices, capable of including different factors contributing to the occurrence of a drought. However, some issues (like the auto-correlation or the proper definition of the multivariate index) are still open and need to be addressed to make these indices applicable in the current practice. Here, a composite agro-meteorological drought index (AMDI-SA) has been introduced, accounting for meteorological and agricultural droughts, considering specifically seasonality and auto-correlation. AMDI-SA combines, through the copula concept and the Kendall function, two drought indices (namely Multivariate Standardized Precipitation Index (MSPI) and the Multivariate Standardized Soil moisture Index (MSSI)) in a statistically consistent (normal-distributed) drought indicator. Nonparametric distributions have been used for the variables of interest and the calculation of MSPI and MSSI, while parametric and nonparametric (empirical) copulas to build AMDI-SA. A pre-whitening procedure has been applied to MSPI and MSSI to remove the autocorrelation. An application to Urmia lake basin in Iran has been presented, drought indices compared, and investigated their spatial variability. Results showed that MSPI and MSSI are able to justify 72% and 89% of the variability throughout the year. AMDI-SA reflects the combined effect of soil moisture and precipitation, and has a behavior in between whitened MSPI and MSSI. In addition, having no memory and being composite index, the AMDI-SA is able to clearly detect the temporal variability of recorded droughts to a greater extent than MSPI and MSSI indices.

## 1. Introduction

Droughts are extreme hydrological events and representative of natural hazards, which impose serious challenges to ecosystems and human societies. Droughts may affect a wide variety of sectors such as agriculture or hydroelectric power generation, with diverse geographical and temporal extent. Droughts have multiple aspects and may be classified into four main types: meteorological, agricultural, hydrological and socioeconomic droughts (Dracup et al., 1980; Heim, 2002). Meteorological droughts are related to the deficiency of precipitation over an extended period of time, from which other types of drought originate. Agricultural droughts relate to insufficient water to meet the need of crop production, or plant growth (Heim, 2002).

Drought indices are useful tools to detect, monitor, and evaluate drought events (Zargar et al., 2011). Several indices have been developed for drought monitoring (Mishra and Singh, 2010). Among these, Standardized Precipitation Index (SPI) is one of the most commonly used, applied to local, regional, and global scale studies. SPI is widely used, primarily for its simplicity, standardized nature, and flexibility of use across different time scales (e.g., 1-, 6-, 12-month). On the other hand, SPI has potential limitations. The assumption of one suitable probability distribution function for precipitation data could be inconsistent under different window sizes and could not account for seasonal variability. Moreover, the selection of one single time scale could be too simplistic or misleading in practical applications to water resource managers, decision-makers, and users. Using a novel approach, Bazrafshan et al. (2014) proposed the Multivariate Standardized Precipitation Index (MSPI) using the 1st principal component of SPI aggregates at different time windows. They showed the superiority of MSPI, when the appropriate time window for a drought study had not been identified in advance.

However, drought analyses based on a single variable may not be sufficient because drought phenomena have complex dynamics involving multiple variables (e.g., precipitation, runoff, and soil moisture) (Hao and AghaKouchak, 2013). Moreover, a single drought index may not be sufficient to describe all aspects of drought onset, persistence and termination (see e.g., Dracup, 1980; Kao and Govindaraju, 2010; Hao and Singh, 2015; Waseem et al., 2015; Hao et al., 2016). For example, indices which are used to monitor meteorological droughts usually capture the drought onset earlier (Behrangi et al., 2015), while soil moisture index describes the drought persistence more reliably (Entekhabi et al., 1996; Heim, 2002). The characterization of droughts with a composite perspective is to go over the inadequacy of drought characterization using a single variable. This can be done by developing drought indices combining multiple hydrological variables, or drought indices (Hao and Singh, 2015).

Recently, several composite drought indices have been developed combining different drought indicators to improve drought characterizations from multiple aspects (e.g. Waseem et al., 2015; Rajsekhar et al., 2014; Azmi et al., 2015).

Hao and AghaKouchak (2013, 2014), using theoretical and empirical copula functions, presented parametric and non-parametric versions of the Multivariate Standardized Drought Index (MSDI), respectively. MSDI is an agro-meteorological drought index based on the bivariate distribution of SPI and SSI (standardized soil moisture index). To obtain MSDI, the joint cumulative probability is transformed with the inverse CDF of a standard normal distribution. However, it is important to note that since the joint cumulative probability is not uniformly distributed on [0,1], the transformation will not result in normally-distributed index. In addition, copula function, to be applied, needs input data to be time-independent, which does not hold for SPI and SSI in general.

The objective of the present paper is to develop a composite agro-meteorological drought index, properly defined, copula-based, addressing seasonality and auto-correlation issues. Furthermore, the proposed framework could be used to assimilate two or more standardized drought indices, into a single drought index that may be useful for comprehensive decision-making.

## 2. Methods

In section 2.1, SPI and SSI indices have been briefly recalled. To avoid lack of information in drought monitoring, and include all within-year variations, SPI and SSI are computed at twelve time windows, from 1 to 12 months. In section 2.2, the modified SPI ($\mathrm{SPI}^m$) and the modified SSI ($\mathrm{SSI}^m$), subgrouped by the ending month *m*, have been considered to account the seasonality of the variables. In section 2.3, $\mathrm{SPI}_w^m$ at 12 time windows (*w*=1,...,12 months) combined, through the principal component analysis, in the multivariate standardized precipitation index, MSPI. Similarly, it has been done for $SSI_w^m$ in the multivariate standardized soil moisture index, MSSI. These indices are standardized (subtracting the mean and dividing by the standard deviation) to remove the seasonality. In section 2.4, the autocorrelation within MSPI and MSSI time series has been removed through a whitening procedure, building whitened series indicated by $MSPI^{wh}$ and $MSSI^{wh}$, respectively. In section 2.5, a new agro-meteorological drought index, denominated AMDI-SA, has been introduced combining together $MSPI^{wh}$ and $MSSI^{wh}$ using the concept of copula and the Kendall function. In section 2.6, an interpretation of the proposed index has been presented.

**2.1 SPI and SSI indices**

The most common drought index is the Standardized Precipitation Index (SPI), introduced by McKee et al. (1993). SPI is calculated with reference to different time scales, and can assess drought severity. As it is a standardized index, the frequency of extreme drought events at different locations and time scales are consistent and comparable. Let $X_w$ and $F_{Xw}(x)$ denote the accumulated precipitation at time scale $w$ and its corresponding probability distribution function, respectively. $F_{Xw}(x)$ is transformed into the standard normal precipitation index (SPI) at time scale $w$ as:

$SPI_w=\Phi^{-1}(F_{Xw}(x))$, where $\Phi^{-1}$ is the inverse of the standard normal distribution.

Sometimes it is hard to discriminate among canonical forms of $F_{Xw}(x)$, or they may not provide a good fit to the data (Soláková et al. 2014; Lall et al., 2016). On the other hand, using different distribution functions could lead to different tail behavior and thus inconsistencies in characteristics of extremes across space (Farahmand and AghaKouchak, 2015). Therefore, a nonparametric approach is used to obtain the probability values of $X_w$. The Gringorten plotting position is used to calculate the cumulated frequency of non-zero values, $F_i=(i-0.44)/(n+0.12)$, where $n$ denotes the sample size of non-zero values, and $i$ refers to the rank of the non-zero observation $x_{(i)}$, ordered from the smallest to the largest. Since there are zero values of precipitation data, the frequency of zeros has to be added to the plotting position of non-zero values to estimate $F_{X_w}$. The method of handling zeros proposed by Stagge et al. (2015) is used, which is superior to using the relative frequency for the probability of zero values. Thus, $F_{X_w}(x)$ is calculated as:

$$F_{X_w}(x) = \begin{cases} \frac{n_0}{n_{tot}+1} + \left(1 - \frac{n_0}{n_{tot}+1}\right)\frac{(i-0.44)}{(n+0.12)}, & \text{for } x = x_{(i)} \\ \frac{n_0+1}{2(n_{tot}+1)}, & \text{for } x = 0 \end{cases} \quad (1)$$

where $n_0$ is the number of zero values, and $n_{tot}=n+n_0$.

The standardized soil moisture index (SSI) (e.g., AghaKouchak, 2014) can be defined in a similar way to SPI. Here, SSI has been derived from soil moisture data averaged up to 100 cm depth.

However, SSI, like SPI, has two weaknesses:

1) The index does not take into account the seasonal variability within the annual regime. In other words, it fits all data (whether it has been observed in wet or dry season) to the same probability distribution.

2) Increasing the temporal window ($w$), it increases the temporal overlap between two successive values of the index, introducing more auto-correlation to the time series of the index, and bias in the probability distribution fitting.

**2.2 Modified SPI and SSI indices**

Kao and Govindaraju (2010) proposed the following modification in the calculation of SPI to account for the seasonal variability of data. The aggregated precipitation $X_w$, at a given time window $w$, is grouped according to the ending month $m$ ($m$=1 means January, ..., $m$=12 December). Thus, the series $\{X_w\}$ is subdivided into 12 smaller subseries corresponding to 12 months of the year, $\{X_w^m\}$. $X_w^m(y)$ is the aggregated precipitation over the time window $w$, having $m$ as the ending month and relative to the year $y$. Thus, $X_1^{10}(y)$ is the value of the year

$y$, with a window size 1 having October as ending month, while $X_5^{10}(y)$ is the value of the year $y$, with a window size 5 from May to October. In doing so, observations in each set $\{X_w^m\}$ will not have overlapping information, when $w \leq 12$, and reduce the auto-correlation among the samples $\{X_w^m\}$. In addition, observations in each set $\{X_w^m\}$ are subject to the same seasonal effect, and hence, the seasonal variation is accounted for, properly (Kao and Govindaraju, 2010). Then for each of the 12 variables $X_w^m$, the empirical frequency gives an estimation of the probability distribution, $F_{X_w^m}(x)$, and the modified index $\mathrm{SPI}_w^m = \Phi^{-1}\left(F_{X_w^m}(x)\right)$ is obtained. Similarly, it is possible to calculate the modified index $\mathrm{SSI}_w^m$.

**2.3 MSPI and MSSI indices**

MSPI, recently developed by Bazrafshan et al. (2014), applies the multivariate technique of Principal Component Analysis (PCA) to a set of SPIs referred to different values of $w$. This technique is applied to $\mathrm{SPI}_w^m$ and $\mathrm{SSI}_w^m$.

Suppose that observations are stored in $O$, a vector of $K$ variables with covariance matrix $Cv$. The Principal Component Analysis (PCA) is a linear combination of $K$ variables: $\mathrm{PC}_i = E_i^{\mathrm{T}} O = \sum_{k=1}^{K} e_{ki}\, O_k$ with $k$=1,...,$K$, where $\mathrm{PC}_i$ is the $i$th principal component, $E_i^T$ is the $i$th eigenvector of $Cv$ sorted in descending order of corresponding eigen values and $e_{ki}$ is the $k$th element of the $i$th eigenvector of $Cv$. These components are: firstly, extracted in such a way that the first one ($\mathrm{PC}_1$) justifies the greatest percentage of variance of $K$ original variables mutually uncorrelated; secondly, this linear combination is mutually uncorrelated the components can be at most as many as the original variables; and thirdly, the components are extracted in such a way that the first one ($\mathrm{PC}_1$) justifies the greatest percentage of variance of $K$ original variables (Wilks, 2011). The PCA can be useful if the correlation among the original variables is high.

Here, the PCA technique was applied to each of two sets of variables $\mathrm{SPI}_w^m$ and $\mathrm{SSI}_w^m$ with $w$=1,...,12 months. However, it can be applied to any arbitrary set and may include other time scales depending on the research needs. Thus, the first component of $\mathrm{SPI}_w^m$ is indicated as $\mathrm{P}_1^m$ and is equal to

$$\mathrm{P}_1^m = \sum_{w=1}^{12} ep_{w1}\mathrm{SPI}_w^m \tag{2}$$

where $ep_{w1}$ is the $w$th element of the first eigenvector of covariance matrix of SPIs. Similarly, the first component of $\mathrm{SSI}_w^m$ is indicated as $\mathrm{S}_1^m$ and equal to

$$\mathrm{S}_1^m = \sum_{w=1}^{12} es_{w1}\mathrm{SSI}_w^m \tag{3}$$

where $es_{w1}$ is the $w$th element of the first eigenvector of covariance matrix of SSIs. $\mathrm{P}_1^m$ and $\mathrm{S}_1^m$ are characterized by seasonality and are not comparable among different months or places. Therefore, normalized variables $\widehat{\mathrm{P}}_1$ and $\widehat{\mathrm{S}}_1$ are introduced, respectively for $\mathrm{P}_1^m$ and $\mathrm{S}_1^m$, subtracting the mean and dividing by the standard deviation:

$$\mathrm{MSPI} = \frac{\mathrm{P}_1^m - \mu_{P_1^m}}{\sigma_{P_1^m}} \approx \frac{\mathrm{P}_1^m}{\sigma_{P_1^m}} \tag{4}$$

$$\mathrm{MSSI} = \frac{\mathrm{S}_1^m - \mu_{S_1^m}}{\sigma_{S_1^m}} \approx \frac{\mathrm{S}_1^m}{\sigma_{S_1^m}} \tag{5}$$

where $\mu_{P_1^m}$ and $\sigma_{P_1^m}$ are respectively the mean and standard deviation of $\mathrm{P}_1^m$, while $\mu_{S_1^m}$ and $\sigma_{S_1^m}$ of $\mathrm{S}_1^m$. $\mu_{P_1^m}$ and $\mu_{S_1^m}$ are close to zero, and so can be ignored in the numerator of Eq.s (4-5) (Keyantash and Dracup, 2004).

MSPI and MSSI are normally distributed with zero mean and unit variance. These indices summarize several information avoiding problems connected to the selection of the appropriate time scales of SPI and SSI. Similar to SPI or SSI, negative values of MSPI or MSSI indicate drought conditions, while positive values to wet conditions. Normal conditions are associated to values of MSPI or MSSI close to zero, see Table 1.

### 2.4 Whitening MSPI and MSSI

Since MSPI or MSSI are auto-correlated and input variables of copula function must be free of auto-correlation (statistically “white”), the temporal dependence has been filtered out. A classical whitening procedure (Box and Jenkins, 1970) has been applied to MSPI and MSSI, assuming that these can be described by autoregressive moving-average (ARMA) models. Whitened residuals of MSPI and MSSI are indicated as $\mathrm{MSPI}^{wh}$ and $\mathrm{MSSI}^{wh}$. Details about ARMA models can be found in Box and Jenkins (1970) and Hipel and McLeod (1996). The Ljung–Box test has been used to assess the absence of auto-correlation in $\mathrm{MSPI}^{wh}$ and $\mathrm{MSSI}^{wh}$ time series, at a significance level of 0.05 (Ljung and Box, 1978; Hipel and McLeod, 1996). (Wang et al., 2012).

**2.5 Composite Drought Index AMDI-SA**

To have a comprehensive description of droughts, a drought index has been considered combining whitened residuals $\mathrm{MSPI}^{wh}$ and $\mathrm{MSSI}^{wh}$ together within the copula framework. With reference to the bivariate case, the copula $\mathrm{C}(u,v)$ is a cumulative distribution function of uniform marginals in the unitary interval, $u,v\in[0,1]$ (Joe, 1997; Salvadori et al., 2007; Nelsen, 2013). Thanks to the Sklar's theorem (Sklar, 1959), the joint cumulative distribution function of $\mathrm{MSPI}^{wh}$ and $\mathrm{MSSI}^{wh}$, $\mathrm{F}_{\mathrm{MSPI}^{wh},\mathrm{MSSI}^{wh}}$, can be written in terms of copula as:

$$\mathrm{F}_{\mathrm{MSSI}^{wh},\mathrm{MSPI}^{wh}}(p,s) = \mathrm{C}\left(\mathrm{F}_{\mathrm{MSPI}^{wh}}(p),\mathrm{F}_{\mathrm{MSSI}^{wh}}(s)\right) \tag{6}$$

where $\mathrm{F}_{\mathrm{MSPI}^{wh}}$ and $\mathrm{F}_{\mathrm{MSSI}^{wh}}$ are the marginals and C is the copula.

The Kendall distribution function ($\mathrm{K_C}(t)$), also called Kendall's measure, is the probability measure of the set $\{(\mathrm{F}_{\mathrm{MSPI}^{wh}}, \mathrm{F}_{\mathrm{MSSI}^{wh}})\in[0,1]^2\text{: } \mathrm{C}(\mathrm{F}_{\mathrm{MSPI}^{wh}}, \mathrm{F}_{\mathrm{MSSI}^{wh}})\leqslant t\}$, with $t \in[0,1]$. It is defined as:

$$\mathrm{K_C}(t)=\Pr[\mathrm{C}(\mathrm{F}_{\mathrm{MSPI}^{wh}},\mathrm{F}_{\mathrm{MSSI}^{wh}})\leqslant t] \tag{7}$$

where $\mathrm{K_C}(t)$ is a univariate probability distribution. For some copula families, like Archimedean ones, $\mathrm{K_C}(t)$ has an analytical form; and for others, like elliptical copulas, it has not a closed-form, and thus, it is calculated numerically. For more details on the Kendall distribution function, see Nelsen et al. (2003); Salvadori et al. (2007); Salvadori and De Michele (2010). Every copula $\mathrm{C}(u,v)$ satisfies the relation $\mathrm{W}\leq \mathrm{C} \leq\mathrm{M}$ on $[0,1]^2$, with W (counter-monotonicity copula) and M (co-monotonicity copula), also referred as the upper and lower Frechet-Hoeffding bounds. In particular, the co-monotonicity copula describes the case of perfect positive dependence, and is given by $M(u,v)=\min\{u,v\}$. The relation $W \le C \le M$ is written in terms of Kendall distribution function as $t = K_M(t) \le K_C(t) \le K_W(t) = 1$, representing the bounds of the Kendall function (Nelsen et al., 2003).

In the next, five copula families are considered: Gaussian, Student's t, Frank, Gumbel and Clayton. For the last three families, the Kendall distribution function is explicitly given. For others, empirical Kendall function has been used.

The Maximum Likelihood method is used for the estimation of the copula parameter, and the Akaike Information Criterion (AIC) used to rank the copulas and select the best one, provided that it well describes the empirical copula. To check the adequacy of parametric copula with the empirical one, the bivariate Kolmogorov-Smirnov goodness-of-fit test has been used. The empirical process $\mathbb{C}_n = \sqrt{n}(C_n - C_{\theta_n})$ has been considered, where $C_n$ is the empirical copula with sample size $n$, and $C_{\theta_n}$ is the parametric copula estimated from a sample of size $n$. The statistic of the Kolmogorov-Smirnov test ($T_n^{(C)}$), defined as supremum of $\mathbb{C}_n$, has been used as measure of adequacy. If $T_n^{(C)}$ is smaller than the critical value associated to 5% significance level, the best-fitted parametric copula is used (Genest et al., 2009).

AMDI-SA is defined as the inverse normal transformation of $K_C(t)$:

$$\text{AMDI-SA} = \Phi^{-1}(K_C(t)) \tag{8}$$

The Kendall distribution function, different to the copula C, is a uniform variable in [0,1], thus allowing a proper definition of the drought index, i.e., a normal distributed variable.

AMDI-SA is a multivariate composite drought index, correctly defined, which accounts seasonality and auto-correlation. It can be compared across regions with markedly different climates. Like any other standard indices, e.g. SPI, AMDI-SA can explain drought characteristics. It should be noted that AMDI-SA, similar to univariate SPI and SSI, provides probability of occurrence, and thus, it can be used for risk analysis as well.

In this study, 11 classes for drought severity classification are used for AMDI-SA, MSPI, MSSI, $\mathrm{MSPI}^{wh}$ and $\mathrm{MSSI}^{wh}$. The classification is according to US drought monitor (USDM) program's objective criteria (Svoboda et al., 2002). Drought classes are described in Table 1. A drought event occurs any time when the index reaches a severity less than or equal to -0.5. The event ends when the index becomes more than this cut-off value. Each drought event has a duration defined by its beginning and ending.

Since the drought index in Eq.(8) is based on $K_C$, in cases where the theoretical $K_C$ is significantly different from the empirical one, according to the univariate Kolmogorov-Smirnov test with 5% significance level, Least Squares method has been applied between the empirical and theoretical $K_C$ function to re-estimate the parameter of the copula, and select the one with the smallest value of the maximum difference. The Kolmogorov-Smirnov test has been used to check the goodness-of-fit. If this test is not passed, then the empirical $K_C$ is used in Eq.(8).

**2.6 Interpretation of AMDI-SA**

Without loss of generality, the threshold level for drought severity is assumed to be AMDI-SA=-0.5 corresponding to $K_C(t) = 0.3$ (the 30th percentile). Fig. 1 illustrates the isoline of Kc=0.3 and its corresponding drought domain, i.e., all the points located under this isoline, in $F_{\mathrm{MSPI}^{wh}} - F_{\mathrm{MSSI}^{wh}}$ plane. Also included, are the empirical copula isoline $C(u,v) = 0.3$ and the L-shaped isoline of the co-monotonicity copula $M(u,v) = \min\{u,v\} = 0.3$. The co-monotonicity copula is considered, since it represents the riskiest dependence structure. Namely, it has a conservative approach which identifies critical condition if at least one of its components is in a critical situation.

In Fig.1, the L-shaped isoline of the co-monotonicity copula is placed under the copula isoline and above the $K_C$ isoline. Regardless of the choice of copula, this is true due to $W \leq C \leq M$ and $t = K_M(t) \leq K_C(t)$.

What does it mean in terms of drought index? If the copula isoline is used to identify the drought domain (as done in Hao and AghaKouchak, 2013 with MSDI), then are considered droughts also conditions where both the two variables ($MSPI^{wh}$ and $MSSI^{wh}$) do not indicate drought (i.e., points located between the copula isoline and L-shaped isoline of the co-monotonicity copula). This criterion seems to overestimate the drought conditions. If the co-monotonicity copula is used to identify the drought domain, then drought condition in one variable means drought condition of the (multivariate) index. Again this criterion seems too precautionary in identifying the drought conditions. If the $K_C$ isoline is used to identify the drought domain (as proposed here), the drought condition in one variable (e.g., precipitation through $MSPI^{wh}$) does not imply drought condition of the (multivariate) index. In other words, the drought severity detected by the multivariate index is in between the severity of the input indices. The difference between the drought domain associated to the co-monotonicity copula isoline and the one with the $K_C$ isoline, is represented by the two light-grey areas. Notice that there is a region (Z in Fig.1), within which the $K_C$ isoline does not indicate the drought even if both variables are in drought conditions. This could represent a weakness of choosing the $K_C$ isoline. However, this area is extremely small and thus can be easily neglected. In conclusion, the drought condition identified by the AMDI-SA ($K_C$ isoline) seems more prudent in identifying drought conditions with respect to the use of copula isolines.

## 3. Study Area

Urmia lake basin is an endorheic basin, located between 37°4′ to 38°17′ latitude and 45°13′ to 46° longitude in northwestern Iran (Fig. 2). Three provinces share the Lake Urmia basin: East Azerbaijan (19000 $km^2$), West Azerbaijan (21500 $km^2$), and Kurdistan (5000 $km^2$) (Yekom, 2005). Major use of water is for the agriculture sector, which is mainly supported by dryland farming with low efficiency (Hesami and Amini, 2016). The climate of Urmia lake basin is harsh and continental, affected mainly by the mountains surrounding the lake (Ghaheri and Baghal-Vayjooee,1999). Considerable seasonal fluctuations in air temperature occur in this semi-arid region. The temperature in the region ranges between 0 and 23°C in winter and up to 39°C in summer (IRIMO, 2009). The annual average precipitation is about 500 mm falling mostly between November and April, while summer months are typically dry.

Monthly precipitation data for the basin are available at daily 0.25° × 0.25° resolution from PERSIANN-CDRdataset (Ashouri et al., 2015). It is retrieved for the period of 1983–2010. The accuracy of PERSIANN precipitation data for Iran and Urmia lake basin has been assessed by Moazami et al., 2013; Bodagh-Jamli, 2015; Ghajarnia et al., 2015; Katiraie-Boroujerdy et al., 2013.

Soil moisture data are derived using ERA-Interim-Land surface fluxes and near-surface meteorology to force the land surface model HTESSEL for the period 1983–2010 (Balsamo et al. 2015). It is considered that most of the vegetation roots are within the first 3 layers of soil in HTESSEL model (0-7cm, 7-28cm and 28-100cm). The weighted average of the water content over these three layers has been calculated to obtain a single value for soil moisture for each 0.25° × 0.25° pixel. To investigate the spatial distribution of droughts, 79 pixels covering the basin (not only the lake) are considered and processed.

## 4. Results

It is necessary to assess the stationarity of timeseries before any other analysis (Adeloye and Montaseri, 2002). Non-parametric Mann–Kendall procedure (Hamed, 2014) was used to test the presence of trends in $MSPI^{wh}$ and $MSSI^{wh}$ time series using a significance level of 0.05. The results of the test indicated the absence of trends for all the data points of Urmia lake basin. The Ljung–Box test (Wang et al., 2012) has been used to assess the presence of autocorrelation within $MSPI^{wh}$ and $MSSI^{wh}$ time series at a significance level of 0.05. The test statistic was placed always in region of acceptance for all data points within Urmia lake basin, indicative of no statistically significant autocorrelation.

In this study, MSPI and MSSI calculation is based on the scores of the first Principal Component (PC1) of the 12 modified variables for selected time windows (w=1,2,…,12) representing seasonal variations throughout the year. To show the variability justified by the first principal component (PC1), the scree plots of the modified SPI/SSI variables have been illustrated for the whole basin as an average in Fig. 3.

To show the statistical dependence between $MSPI^{wh}$ and $MSSI^{wh}$, the scatter plot of the couples, at point A in Fig. 2, is given in Fig. 4. Soil moisture conditions respond to precipitation anomalies on a relatively short time lag or even no significant lag. The Kendall's tau association measure between $MSPI^{wh}$ and $MSSI^{wh}$ for the whole basin is in the range of 0.18-0.36. The positive association depicted by positive values of Kendall's tau in the basin may be due to the direct impact of precipitation on soil moisture in the basin.

Input variables of copula function needs to be statistically white. To assess the effectiveness of the adopted whitening procedure, values of the auto-correlation coefficient of monthly MSPI, MSSI, $MSPI^{wh}$ and $MSSI^{wh}$ time series at point B (in Fig. 2) for lags 1-24 are given in Fig. 5.

AMDI-SA reflects the combined effect of soil moisture and precipitation. Fig. 6 reports the time series of MSPI, MSSI (upper panel), and $MSPI^{wh}$ and $MSSI^{wh}$ (lower panel) along with AMDI-SA at point B in Fig. 2. For brevity, only values for the period 1998-2002 are presented, corresponding to a significant drop of agricultural production. As shown in the figure, AMDI-SA mostly lie between $MSPI^{wh}$ and $MSSI^{wh}$.

In Fig. 7, spatial patterns of drought using MSPI, MSSI, $MSPI^{wh}$, $MSSI^{wh}$ and AMDI-SA are given for March 1999, during which a drought period occurred over the basin. MSSI and MSPI generally show more severe drought condition in northern part of the basin, while

AMDI-SA shows more severe drought condition in the southeast of the basin. However, values of whitened series of MSSI and MSPI show the same spatial pattern as AMDI-SA.

Worthwhile to mention that sample points (A, B and C) in Fig. 2 are selected randomly to cover the whole basin, close to the lake and in mountainous area. The same result can also be achieved by reporting any other data points in the basin.

**5. Discussion**

According to Fig. 3, the justified variabilities in the modified SPI and modified SSI by the first principal component (PC1) are 72% and 89%, respectively. Such high values show capability of PCA technique to integrate the great part of within-year variability existing in modified SPI and modified SSI time series into one series. Since soil moisture shows less variability than precipitation, PC1 is able to justify greater percentage of variability of modified SSI throughout the year.

As shown in Fig. 5, the auto-correlation in MSSI is greater than MSPI for all lags. This may be due to relatively stronger memory of soil moisture. However, in whitened MSPI and MSSI series, the auto-correlation is not significant indicating the effectiveness of the adopted ARMA whitening procedure.

According to Fig. 6, since AMDI-SA reflects the combined effect of soil moisture and precipitation, it mostly lies between $MSPI^{wh}$ and $MSSI^{wh}$. In contrast to $MSPI^{wh}$, $MSSI^{wh}$ and AMDI-SA, MSPI and MSSI are auto-correlated time series and cannot capture rapid changes in wetness condition. As showed in Fig. 7, drought severity class detected by AMDI-SA is less severe than both or one of input indices, $MSPI^{wh}$ and $MSSI^{wh}$. Moreover, severe drought detections of MSPI and MSSI may be the effect of auto-correlation.

As stated before, in case of significant difference between empirical and theoretical values of the Kendall distribution function, empirical values have been used to calculate AMDI-SA. An example of such case occurs at point C (in Fig. 2). First, a Gumbel copula is selected due to its lowest value of Akaike Information Criterion (AIC). However, since there is a significant difference between theoretical and empirical values of $K_C$, a Least Squares fitting on $K_C$ function has been done for copulas with closed-form of $K_C$. In Fig. 8, there is the comparison between theoretical and empirical $K_C$, and the Least Squares estimates are indicated with the "LS" suffix. Since, neither the LS estimates are close enough to the empirical Kendall function, empirical values are used to calculate AMDI-SA.

According to AIC, Gumbel and Gaussian copulas are the selected families, in 41 (52%) and 38 (48%) cases, respectively. In 9 cases, where the Gumbel family was selected, the theoretical Kendall function was not close enough to the empirical one. Finally, in 32 cases the Gumbel copula, in 38 cases the Gaussian model, and in 9 cases the empirical copula have been selected (Fig. 9).

The Gumbel copula, compared to the Gaussian copula, has more probability concentrated in the tails, especially in the right one, i.e. higher values of one variable is more likely to be followed by higher than normal values of the other. The ability of Gumbel copula to justify the dependence between $MSPI^{wh}$ and $MSSI^{wh}$ may imply saturated condition of soil moisture for months with upper-normal precipitation. In other words, for upper-normal precipitation condition, soil moisture will not be depleted, rapidly. It may be due to poor vegetation cover in these locations.

## 6. Conclusions

A single distribution function may not fit the global precipitation/soil moisture data and hence, the original parametric SPI/SSI may not be applicable. In this study, the proposed approach does not require the assumption of a parametric distribution function for describing drought-related variables. It is also worth pointing out that unlike parametric indices, the suggested nonparametric framework does not require parameter estimation and goodness-of-fit evaluation. However, due to the use of a nonparametric framework to derive SPI/SSI, the proposed methodology requires long-term observations to derive the joint distribution of precipitation and soil moisture, and a short record of observations could lead to biases in the indices values. On the other hand, given that satellite-based hydro-climate data records are emerging, the authors expect that, in the near future, more research will be devoted to investigating spatial patterns of climate extremes using space-borne observations specially soil moisture data.

Drought mitigation and response plans often rely on different indicator variables and drought triggers. No single index can represent all aspects of drought so it is best to use a multi-index approach for operational drought monitoring. However, many drought indicators are not directly statistically comparable (Steinemann and Cavalcanti, 2006). Moreover, limited statistical models are currently available for linking or merging different drought-related variables into one composite map.

AMDI-SA, a multivariate composite agro-meteorological drought index, is suggested, which joints together two drought indices, namely Multivariate Standardized Precipitation Index and the Multivariate Standardized Soil moisture Index through the copula concept and the Kendall function.

The properties of AMDI-SA can be summarized as follows: (a) AMDI-SA is a properly defined normal-distributed drought index gaining information from both precipitation and soil moisture; (b) using appropriate whitening procedure and normalization, AMDI-SA accounts for auto-correlation and seasonality; (c) Due to its probabilistic nature, it can be used for drought risk assessment tool and aid to decision-makers in drought mitigation and response plans. (d) Typically, precipitation detects the drought earlier and soil moisture better describes the persistence events (Farahmand et al., 2015). AMDI-SA generally captures the drought onset similar to the precipitation and drought persistence similar to the soil moisture, combines the properties of both. Though, soil moisture levels may remain high even long after precipitation.

However, since it uses whitened time series of MSPI and MSSI, shows even more quick reflection to drought onset and more fluctuation than both MSPI and MSSI.

The proposed framework for creating AMDI-SA is rather general, and other indices can be integrated together to form such a composite drought index which could be considered a strength of the approach. Such a methodology can potentially improve drought monitoring if each of the selected drought-related variables can capture certain aspects of droughts. Efforts are underway to extend the AMDI-SA concept by integrating more drought indicators such as evapotranspiration. The distinct advantages of the AMDI-SA include its assessment of drought from the aggregate perspective of meteorological and agricultural water shortages, and its direct mathematical formulation, which can be rapidly applied to new observational data in a straightforward manner.

It seems that AMDI-SA is not meant to replace the currently available indices. Rather, it uses additional information that can potentially improve drought modeling. Finally, it should be noted that the best choice for a set of drought indicators to be combined may vary, depending on the problem at hand.

## 6. References

Adeloye, A. J., & Montaseri, M. (2002). Preliminary streamflow data analyses prior to water resources planning study. *Hydrological Sciences Journal*, *47*(5), 679-692.

AghaKouchak, A. (2014). A baseline probabilistic drought forecasting framework using standardized soil moisture index: application to the 2012 United States drought. *Hydrol. Earth Syst. Sci.*, *18*(7), 2485-2492. doi: 10.5194/hess-18-2485-2014

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Prat, O. P. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, *96*(1), 69-83.

Azmi, M., Rüdiger, C., Walker, J. P. (2016). A data fusion-based drought index. *Water Resour. Res., 52*(3), 2222-2239. doi: 10.1002/2015WR0178

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H.,Vitart, F. (2015). ERA-Interim/Land: a global land surface reanalysis data set. *Hydrol. Earth Syst. Sci., 19*(1), 389-407. doi: 10.5194/hess-19-389-2015

Bazrafshan, J., Hejabi, S., & Rahimi, J. (2014). Drought monitoring using the multivariate standardized precipitation index (MSPI). *Water Resources Management, 28*(4), 1045-1060.

Behrangi, A., Loikith, P., Fetzer, E., Nguyen, H., & Granger, S. (2015). Utilizing Humidity and Temperature Data to Advance Monitoring and Prediction of Meteorological Drought. *Climate, 3*(4), 999.

Bodagh-Jamli, J. (2015). Validation of Satellite-Based PERSIANN Rainfall Estimates Using Surface-Based APHRODITE Data over Iran. *Earth Sciences, 4*(5), 11. doi: 10.11648/j.earth.20150405.11

Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: forecasting and control, 1976. *ISBN: 0-8162-1104-3*.

Dracup, J. A., Lee, K. S., & Paulson, E. G. (1980). On the definition of droughts. *Water Resources Research, 16*(2), 297-302. doi: 10.1029/WR016i002p00297

Entekhabi, D., Rodriguez-Iturbe, I., &Castelli, F. (1996). Mutual interaction of soil moisture state and atmospheric processes. *Journal of Hydrology, 184*(1–2), 3-17. doi: http://dx.doi.org/10.1016/0022-1694(95)02965-6

Farahmand, A., & AghaKouchak, A. (2015). A generalized framework for deriving nonparametric standardized drought indicators. *Advances in Water Resources, 76*, 140-145.doi:http://dx.doi.org/10.1016/j.advwatres.2014.11.012

Farahmand, A., AghaKouchak, A., & Teixeira, J. (2015). A vantage from space can detect earlier drought onset: An approach using relative humidity. *Scientific reports*, *5*. doi:http://dx.doi.org/10.1038/srep08553

Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, *44*(2), 199-213. doi: http://dxgoo.doi.org/10.1016/j.insmatheco.2007.10.005

Ghaheri, M., Baghal-Vayjooee, M. H., &Naziri, J. (1999). Lake Urmia, Iran: A summary review. *International Journal of Salt Lake Research*, *8*(1), 19-22. doi: 10.1007/bf02442134

Ghajarnia, N., Liaghat, A., & DaneshkarArasteh, P. (2015). Comparison and evaluation of high resolution precipitation estimation products in Urmia Basin-Iran. *Atmospheric Research*, *158–159*, 50-65. doi: http://dx.doi.org/10.1016/j.atmosres.2015.02.010

Hamed, K. H. (2014). Significance of statistical tests and persistence in hydrologic processes. In *Handbook of Engineering Hydrology: Modeling, Climate Change, and Variability* (pp. 517-535). CRC Press.

Hao, Z., & AghaKouchak, A. (2013). Multivariate Standardized Drought Index: A parametric multi-index model. *Advances in Water Resources*, *57*(0), 12-18. doi: http://dx.doi.org/10.1016/j.advwatres.2013.03.009

Hao, Z., & AghaKouchak, A. (2014). A nonparametric multivariate multi-index drought monitoring framework. *Journal of Hydrometeorology*, *15*(1), 89-101. doi: http://doi.org/10.1175/JHM-D-12-0160.1

Hao, Z., Hao, F., Singh, V. P., Xia, Y., Ouyang, W., & Shen, X. (2016). A theoretical drought classification method for the multivariate drought index based on distribution properties of standardized drought indices. *Advances in Water Resources*, *92*, 240-247. doi: http://dx.doi.org/10.1016/j.advwatres.2016.04.010

Hao, Z., & Singh, V. P. (2015). Drought characterization from a multivariate perspective: A review. *Journal of Hydrology, 527*, 668-678. doi: http://dx.doi.org/10.1016/j.jhydrol.2015.05.031

Heim, 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. doi: 10.1175/1520-0477(2002)083<1149:AROTDI>2.3.CO;2

Hesami, A., & Amini, A. (2016). Changes in irrigated land and agricultural water use in the Lake Urmia basin. *Lake and Reservoir Management,32*(3), 288-296. doi: 10.1080/10402381.2016.1211202

Hipel K. W., & McLeod A. E. (1996). *Time series modeling of water resources and environmental systems*, Elsevier.

IRIMO (2009). Iranian Meteorological Office, Data Processing Center, Tehran, Iran.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*: CRC Press.

Kao, S.-C., & Govindaraju, R. S. (2010). A copula-based joint deficit index for droughts. *Journal of Hydrology, 380*(1–2), 121-134. doi: http://dx.doi.org/10.1016/j.jhydrol.2009.10.029

Katiraie-Boroujerdy, P.-S., Nasrollahi, N., Hsu, K.-l., & Sorooshian, S. (2013). Evaluation of satellite-based precipitation estimation over Iran. *Journal of Arid Environments, 97*, 205-219. doi: http://dx.doi.org/10.1016/j.jaridenv.2013.05.013

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).

Lall, U., Devineni, N., & Kaheil, Y. (2016). An Empirical, Nonparametric Simulator for Multivariate Random Variables with Differing Marginal Densities and Nonlinear Dependence with Hydroclimatic Applications. *Risk Analysis, 36*(1), 57-73. doi: 10.1111/risa.12432

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303. doi: 10.1093/biomet/65.2.297

McKee, T. B., Doesken, N. J., & Kleist, J. (1993). *The relationship of drought frequency and duration to time scales.* Paper presented at the Proceedings of the 8th Conference on Applied Climatology.

Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. *Journal of Hydrology, 391*(1–2), 202-216. doi: http://dx.doi.org/10.1016/j.jhydrol.2010.07.012

Moazami, S., Golian, S., Kavianpour, M. R., & Hong, Y. (2013). Comparison of PERSIANN and V7 TRMM Multi-satellite Precipitation Analysis (TMPA) products with rain gauge data over Iran. *International Journal of Remote Sensing, 34*(22), 8156-8171. doi: 10.1080/01431161.2013.833360

Nelsen, R. B. (2013). *An introduction to copulas* (Vol. 139): Springer Science & Business Media.

Nelsen R.B., Quesada-Molina J.J., Rodríguez-Lallena J.A., Úbeda-Flores M. (2003). Kendall distribution functions, *Statistics & Probability Letters*, 65(3), 263-268. doi:10.1016/j.spl.2003.08.002

Rajsekhar, D., Singh, V.P., & Mishra, A. K. (2014). Multivariate drought index: An information theory based approach for integrated drought assessment. *Journal of Hydrology, 526*, 164-182.doi:10.1016/j.jhydrol.2014.11.031.

Soláková, T., De Michele, C., & Vezzoli, R. (2014). Comparison between Parametric and Nonparametric Approaches for the Calculation of Two Drought Indices: SPI and SSI. Journal of Hydrologic Engineering, 19(9), 04014010-1.

Salvadori, G., & De Michele, C. (2010). Multivariate multi-parameter extreme value models and return periods: A copula approach. *Water Resour. Res.*, 46, W10501, doi:10.1029/2009WR009040.

Salvadori, G., De Michele, C., Kottegoda, N., & Rosso, R. (2007). *Extremes in Nature: An Approach Using Copulas*, Springer.

Sklar, M. (1959). *Fonctions de répartition à n dimensions etleursmarges*: Université Paris 8.

Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F., & Stahl, K. (2015). Candidate distributions for climatological drought indices (SPI and SPEI). *International Journal of Climatology*, *35*(13), 4027-4040.

Steinemann, A. C., & Cavalcanti, L. F. (2006). Developing multiple indicators and triggers for drought plans. *Journal of Water Resources Planning and Management*, *132*(3), 164-174.

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., ... & Miskus, D. (2002). The drought monitor. *Bulletin of the American Meteorological Society*, *83*(8), 1181-1190.

Wang, H., Gao, X., Qian, L., & Yu, S. (2012). Uncertainty analysis of hydrological processes based on ARMA-GARCH model. *Science China Technological Sciences*, *55*(8), 1-11.

Waseem, M., Ajmal, M., & Kim, T.-W. (2015). Development of a new composite drought index for multivariate drought assessment. *Journal of Hydrology*, *527*, 30-37.

Wilks, D. S. (2011). Statistical methods in the atmospheric sciences.

Yekom, C. C. (2005). Environmental impacts (qualitative and quantitative) of Water Resources Development Projects in Urmia Lake Basin (in persian).

Zargar, A., Sadiq, R., Naser, B., & Khan, F. I. (2011). A review of drought indices. *Environmental Reviews*, *19*(NA), 333-349. doi: 10.1139/a11-013

Table 1. Drought classification for normal distributed indices, according to U.S. Drought Monitor (USDM) classification for drought severity.

| **Category** | **Description** | **Index Value** |
|---|---|---|
| **D4** | Exceptional drought | ≤-2.0 |
| **D3** | Extreme drought | (-2.0, -1.6] |
| **D2** | Severe drought | (-1.6, -1.3] |
| **D1** | Moderate drought | (-1.3, -0.8] |
| **D0** | Abnormally dry | (-0.8, -0.5] |
| **N** | Normal | (-0.5, 0.5) |
| **W0** | Abnormally wet | [0.5, 0.8) |
| **W1** | Moderately wet | [0.8, 1.3) |
| **W2** | Severely wet | [1.3, 1.6) |
| **W3** | Extremely wet | [1.6, 2.0) |
| **W4** | Exceptionally wet | ≥2.0 |



Figure 1. Example to describe the properties of AMDI-SA.
Figure 2. Urmia lake basin in the northwest of Iran and discretization in some pixels (displayed as circles on the left). Results of this study reported in the randomly selected sample points A, B and C in the basin.
Figure 3. Scree plots of the SPI and SSI time series for the time scales set of 1–12 months.
Figure 4. Scatter plot showing pairs of $MSPI^{wh}$ and $MSSI^{wh}$ data (for point A in Fig. 2) with Kendall's tau= 0.26.
Figure 5. Influence of the whitening process on auto-correlation; auto-correlation versus different time lags of (a) MSPI and $MSPI^{wh}$, (b) MSSI and $MSSI^{wh}$.
Figure 6. (a) MSPI and MSSI, (b) $MSPI^{wh}$ and $MSSI^{wh}$ in comparison with AMDI-SA during 1998-2002 for a grid cell in Urmia lake basin (for point B in Fig. 2).
Figure 7. Spatial variation of drought severity classes based on MSPI, MSSI, $MSPI^{wh}$, $MSSI^{wh}$ and AMDI-SA for each pixel of Urmia lake basin in March 1999.
Figure 8. Different Kendall distribution functions ($K_C$) for a grid cell in Urmia lake basin (for point C in Fig. 2). t stands for the level of probability.
Figure 9. Selected copula families to model the dependence between $MSPI^{wh}$ and $MSSI^{wh}$ for each pixel of Urmia lake basin.