

Regional drought frequency analysis using L-moments and adjusted charged system search

Amin Abdi, Yousef Hassanzadeh, Siamak Talatahari, Ahmad Fakheri-Fard and Rasoul Mirabbasi

ABSTRACT

The parameter estimation of statistical distributions is important for regional frequency analysis (RFA). The accuracy of different parts of RFA such as estimating the regional quantiles of the selected statistical distribution, determining the heterogeneity measure, and choosing the best distribution based on the Monte Carlo simulation, may be influenced by using the different values of regional parameters. To fulfill this aim, in the present study, a new model is developed for regional drought frequency analysis. This model utilizes the L-moments approach and the adjusted charged system search as an advanced meta-heuristic algorithm, in which some modifications on the equations of the algorithm are performed to improve its standard variant. The verification of the regional parameters estimated by the new methodology yields accurate results compared to other models. Furthermore, this study illustrates the usefulness of the robust discordancy measure against the classic one. For this purpose, different values of the subset factors (α) are utilized in the robust discordancy measure, and finally, the best value of subset factor is found equal to 0.8, which can accurately recognize discordant sites within the region.

Key words | adjusted charged system search, drought, L-moments approach, regional frequency analysis, robust discordancy measure

Amin Abdi (corresponding author)
Yousef Hassanzadeh
Siamak Talatahari
Department of Civil Engineering,
University of Tabriz,
Tabriz,
Iran
E-mail: amin.abdi@yahoo.com

Ahmad Fakheri-Fard
Department of Agricultural Engineering,
University of Tabriz,
Tabriz,
Iran

Rasoul Mirabbasi
Department of Water Engineering,
Shahrekord University,
Shahrekord,
Iran

INTRODUCTION

Drought is a normal part of any climate (Keshavarz *et al.* 2013). Generally, droughts take place when a region receives consistently below average precipitation (Hossein *et al.* 2012). Compared with other natural hazards, the damaging effects of droughts are non-structural, and much larger in terms of the spatial extent (Reddy & Ganguli 2012), where 22% of economic damages and 33% of the number of affected persons are related to droughts (Belayneh *et al.* 2014). Therefore, utilizing regional frequency analysis (RFA) seems necessary to recognize the properties of this phenomenon. RFA can circumvent the limitation of statistical estimates, such as the absence of lengthy records or too short records (Zhang *et al.* 2012). RFA leads to more accurate quantile estimations than at-site analysis (Santillán *et al.* 2014).

To perform RFA, the considered region should be homogeneous, otherwise, it should be subdivided into some homogeneous regions (Seckin *et al.* 2013). Among many homogeneity tests in the hydrologic literature, the L-moments (LMOM)-based tests proposed by Hosking & Wallis (1993, 1997) are known as the most powerful ones (Viglione *et al.* 2007). Thus, numerous studies have utilized the LMOM approach to the RFA of hydrologic phenomena such as rainfall, stream flow, drought, and other fields of water engineering (Modarres 2010; Chérif & Bargaoui 2013; Dikbas *et al.* 2013; Rahman *et al.* 2013; Rajsekhar *et al.* 2013; Zakaria & Shabri 2013; Dodangeh *et al.* 2014; Feng *et al.* 2014; Goyal & Gupta 2014; Sarhadi & Heydarizadeh 2014).

Detection of the possible outlying data in a region is an important matter for RFA (Saf 2010). Hosking & Wallis

(1993, 1997) recommended the discordancy measure based on the mean and covariance of the LMOM ratios using the Mahalanobis distance. This measure can be strongly affected by the discordant sites and is not robust for outlying data. To improve this problem, Neykov *et al.* (2007) proposed the robust discordancy measure. A few studies have then used this measure in hydrology. Saf (2010) analyzed the effect of discordancy detection measures on regional flood probability types and the accuracy of the estimation on regional analysis. Ilorme & Griffis (2013) utilized the robust discordancy measure to determine physically discordant sites and proposed a new methodology to identify the physical attributes that are the greatest demonstrator of extreme hydrological response. The mentioned studies showed the capability of robust discordancy measure against the classic one for outliers' identification; however, the importance of subset size for all data sets was not investigated. In this study, the different subset sizes are considered to evaluate the accuracy of the robust discordancy measure.

The LMOM method is also utilized for parameter estimation of statistical distribution for RFA, although there are no simple expressions and explicit solutions for fulfilling this aim. Therefore, several approximations or simplifications are suggested for solving such nonlinear and complex problems (Hassanzadeh *et al.* 2011; Reddy & Singh 2014). Meta-heuristic algorithms (MHA) are one useful tool to solve these problems and various MHA have been extensively used to solve the optimization problems of hydrologic phenomena (e.g., Rai *et al.* 2009; Hassanzadeh *et al.* 2011; Reddy & Singh 2014; Shin *et al.* 2014).

Recently, an MNA known as charged system search (CSS) was introduced by Kaveh & Talatahari (2010), and applied to different engineering optimization problems. CSS utilizes the Coulomb and Gauss's laws from electrostatics, and the Newtonian laws of mechanics. There are few studies on applying CSS in the field of water resources. Kaveh *et al.* (2012) developed four different models of composite open channels and then optimized them by CSS. The results of this method were compared to those of ant colony optimization and genetic algorithm (GA). Talatahari *et al.* (2013) applied CSS and particle swarm optimization (PSO) for identifying the parameters of the linear Muskingum model. In order to evaluate these algorithms, a numerical

example was utilized and the results were compared to those of other algorithms. Sheikholeslami *et al.* (2014) utilized CSS to optimal cost design of water distribution networks. They utilized some well-known benchmark instances as case studies. Comparison of the results of CSS with some other MHA demonstrated the high performance of this algorithm. In this study, an adjusted CSS (ACSS) algorithm is applied for estimating the regional parameters of Kappa distribution, and then the results are compared with those of the LMOM as the conventional approach.

Considering these points, the main contributions of this paper are:

- (1) presenting the robust discordancy measure with different values of subset factor and comparing with the classic discordancy measure;
- (2) applying the ACSS algorithm for the parameter estimation of statistical distributions; and
- (3) proposing a new model to give regional parameters for statistical distribution.

METHODOLOGY

ACSS

The ACSS algorithm as a meta-heuristic algorithm contains a number of charged particles (CPs), where each one treats as a charged sphere and can insert an electric force to the others. Thus, the resultant electrical force affect a CP results in its acceleration. ACSS uses the governing laws of motion from the Newtonian mechanics to determine the position of CPs. Application of these laws provides a good balance between the exploration and the exploitation of the algorithm (Kaveh & Talatahari 2010). The pseudo-code for the ACSS algorithm is summarized as follows.

Step 1: Initialization. The magnitude of the charge for each CP is defined as:

$$q_i = \frac{f_i - f_w}{f_b - f_w}, \quad i = 1, 2, \dots, N \quad (1)$$

where f_b and f_w are the best and the worst objective function values among all of the particles; f_i represents the fitness of

the agent i ; and N is the total number of CPs. The separation distance r_{ij} between any two CPs is defined as follows:

$$r_{ij} = \frac{\|\mathbf{X}_i - \mathbf{X}_j\|}{\|(\mathbf{X}_i - \mathbf{X}_j)/2 - \mathbf{X}_b\| + \varepsilon} \quad (2)$$

where \mathbf{X}_i and \mathbf{X}_j are the positions of the i th and j th CPs, respectively; \mathbf{X}_b is the position of the best current CP; and ε is a small positive number. The initial positions of CPs are determined randomly and the initial velocities of CPs are assumed to be zero.

Step 2: Charged memory creation. A number of the best CPs and the values of their corresponding objective functions are saved in the charged memory.

Step 3: Force determination. The probability of moving each CP towards the others is determined using the following function:

$$p_{ij} = \begin{cases} 1 & \frac{f_i - f_b}{f_j - f_i} > rand \text{ or } f_j > f_i \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Then, the resultant force vector for each CP is calculated as:

$$\mathbf{F}_j = q_j \sum_{i, i \neq j} \left(\frac{q_i}{a^3} r_{ij} \cdot i_1 + \frac{q_i}{r_{ij}^2} \cdot i_2 \right) p_{ij} (\mathbf{X}_i - \mathbf{X}_j) \quad (4)$$

$$j = 1, 2, \dots, N$$

$$\begin{cases} i_1 = 1, i_2 = 0 & \Leftrightarrow r_{ij} < a \\ i_1 = 0, i_2 = 1 & \Leftrightarrow r_{ij} \geq a \end{cases}$$

where \mathbf{F}_j is the resultant force acting on the j th CP; \mathbf{X}_i and \mathbf{X}_j are the positions of the i th and j th CPs, respectively.

Step 4: Force type identification. The kind of forces can be attractive or repelling, and this is determined by using the kind of force parameter, ar_{ij} , defined as:

$$ar_{ij} = \begin{cases} +1 & rand \leq k_t \\ -1 & rand > k_t \end{cases} \quad (5)$$

where ar_{ij} determines the type of the force, with +1 representing the attractive force and -1 denoting the repelling force, and k_t is the force type coefficient, which controls the effect of the kind of force. ar_{ij} should be multiplied to the resultant force.

Step 5: Solution construction. Each CP moves to the new position as:

$$\mathbf{X}_{j,new} = rand_{j1} \cdot k_a \cdot \mathbf{F}_j + rand_{j2} \cdot k_v \cdot \mathbf{V}_{j,old} + \mathbf{X}_{j,old} \quad (6)$$

$$\mathbf{V}_{j,new} = \frac{\mathbf{X}_{j,new} - \mathbf{X}_{j,old}}{\Delta t} \quad (7)$$

where k_a and k_v are the acceleration and the velocity coefficients, respectively; $rand_{j1}$ and $rand_{j2}$ are two random numbers; $\mathbf{V}_{j,old}$ and $\mathbf{V}_{j,new}$ are the old and new velocities of the j th CP, respectively; and Δt is the time step which is set to unity.

Step 6: Velocities correction. To prevent escaping CPs from the search area, the velocities of CPs in each dimension are restricted to a maximum velocity ($v_{j,max}$), as:

$$v_{i,j} = \begin{cases} v_{j,max} & \text{if } v_{i,j} > v_{j,max} \\ -v_{j,max} & \text{if } v_{i,j} < -v_{j,max} \end{cases} \quad (8)$$

The maximum velocity is defined as:

$$v_{j,max} = \gamma (x_{j,max} - x_{j,min}) \quad (9)$$

where γ is the fraction of the distance between the bounds in the range [0,1]. $x_{j,max}$ and $x_{j,min}$ represent the upper and lower bounds for each design variable, respectively.

Step 7: Solutions modification. The nearest limit and reflection methods is utilized for modifying the positions and velocities of violated CPs, respectively. Therefore, any variable of the violated CP is regenerated as follows:

$$\begin{cases} x_{i,j} = x_{j,min}, v_{i,j} = -v_{i,j} & \text{if } x_{i,j} < x_{j,min} \\ x_{i,j} = x_{j,max}, v_{i,j} = -v_{i,j} & \text{if } x_{i,j} > x_{j,max} \end{cases} \quad (10)$$

This step is utilized in the algorithm in order to direct the unfeasible solutions.

Step 8: Updating of charged memory. The best new vectors are included in the CM and the worst ones are excluded from the charged memory.

Step 9: Terminating criterion control. Steps 3–9 are repeated until a terminating criterion is satisfied.

It is notable that the steps 4, 6, and 7 are only applied in the ACSS algorithm.

LMOM approach

Hosking (1986, 1990) defined LMOM as linear combinations of probability weighted moments, which can be interpreted as measures of the location, scale, and shape of probability distributions. Based on LMOM, Hosking & Wallis (1993) proposed three useful statistics in RFA, the discordancy, heterogeneity, and goodness-of-fit measures. Detailed explanations and formula can be found in Hosking & Wallis (1997).

Discordancy measure

A discordancy measure is used to recognize discordant site(s) in a region to remove them. The classic discordancy measure (D_i) for the i th site in a region is defined as:

$$D_i = \frac{N}{3} (u_i - \bar{u})^T S^{-1} (u_i - \bar{u}) \quad (11)$$

where N is number of sites, $u_i = [t^{(i)}, t_3^{(i)}, t_4^{(i)}]^T$ is a vector of the LMOM ratios for the i th site, and its components are: t as the L-coefficient of variation (L-CV), t_3 as the L-coefficient of skewness (L-CS), and t_4 as the L-coefficient of kurtosis (L-CK), respectively; \bar{u} is the regional unweighted average and S is the matrix of sums of squares and cross-products. The critical discordancy measure is equal to 3 and the site becomes discordant when $D_i > 3$ (Hosking & Wallis 1993, 1997). Due to the outlier observations' influence on the classical estimates of \bar{u} and S , the number of discordant sites for a region can be smaller than the real number. Therefore, it is necessary to utilize the robust discordancy measure in RFA (Saf 2010).

There are many robust estimators in statistics, such as minimum covariance determinant (MCD), minimum volume ellipsoid, and M-estimators (Alameddine et al. 2010). The MCD estimator is a highly robust estimator of multivariate location and scatter (Rousseeuw & Van Driessen 1999; Hubert et al. 2005; Rousseeuw et al. 2006). Rousseeuw (1984, 1985) defined the MCD estimator, which can be computed efficiently with the FAST-MCD algorithm. Finding the h observations (out of N) whose classical covariance matrix has the lowest possible determinant, is the main objective of MCD. There is a relation between the subset size

(h) and subset factor (α) when the value of α varies from 0.5 to 1. If $\alpha = 0.5$, h equals $\lfloor (N + p + 1)/2 \rfloor$, and when $\alpha = 1$, h equals N (Pison et al. 2002; Neykov et al. 2007). For other values of α , the h can be calculated by linear interpolation.

Neykov et al. (2007) introduced the robust discordancy measure (RD_i) which uses the robust estimates (M and C), instead of classic ones (\bar{u} and S) in the Mahalanobis distance formula for RFA:

$$RD_i = RD(u_i) = \sqrt{(u_i - M)^T C^{-1} (u_i - M)} \quad (12)$$

where M and C are the average of h observation and its covariance matrix (Pison et al. 2002; Hubert et al. 2008).

Reweighted MCD estimators can be used to increase the finite-sample efficiency. Therefore, a weight w_i is defined based on the initial RD_i by:

$$w_i = \begin{cases} 1 & \text{if } RD_i \leq \sqrt{\chi_{p, 0.975}^2} \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

where χ^2 is a chi-square distribution with degrees of freedom by p for the corresponding probability 0.975. For $p = 3$, the critical discordancy value becomes $\sqrt{\chi_{3, 0.975}^2} = 3.06$. The reweighted MCD estimators are calculated as:

$$M_R = \frac{\left(\sum_{i=1}^N w_i u_i \right)}{\left(\sum_{i=1}^N w_i \right)} \quad (14)$$

$$C_R = \frac{\left(\sum_{i=1}^N w_i (u_i - M_R)(u_i - M_R)^T \right)}{\left(\sum_{i=1}^N w_i - 1 \right)} \quad (15)$$

The final robust discordancy measure is computed using these reweighted estimates (M_R , C_R).

In this study, the MATLAB library for robust analysis, LIBRA, introduced by Verboven & Hubert (2005), is used for the computation of the robust discordancy measure.

Heterogeneity measure

One of the main steps in the RFA contains identifying homogenous regions (Saf 2009). For this purpose, the

heterogeneity measure was proposed to compute the degree of heterogeneity in a region. The steps of the heterogeneity measure are illustrated as below (Hosking & Wallis 1993).

Step 1: Calculate the regional LMOM ratios.

Step 2: Compute the V variables.

Step 3: Fit the four-parameter Kappa (KAP) distribution.

Step 4: Simulate a large number of synthetic regions.

Step 5: Drive the H indices. Finally, heterogeneity measures are given as follows:

$$H_i = \frac{(V_i - \mu_{vi})}{\sigma_{vi}}, \quad i = 1, 2, 3 \quad (16)$$

where μ_{vi} and σ_{vi} are the mean and standard deviation of the simulated V variables.

Hosking & Wallis (1993) suggested that a region is 'acceptably homogeneous' if H_i is less than 1, 'possibly heterogeneous' if H_i is between 1 and 2, and 'definitely heterogeneous' if H_i is greater than 2. Also, they stated that the H_1 statistic compared to the statistics H_2 and H_3 has a discriminating power and is the most important heterogeneity measure (Rao & Hamed 2000).

Estimation of regional parameters

In order to determine the heterogeneity measure, it is necessary to estimate the regional parameters of KAP distribution for the sample data. In this study, three models are considered to fulfill this aim.

Model 1: In the first model, the parameters of the KAP distribution for each site are estimated using the LMOM and ACSS methods, separately. The cost function which is selected to be applied by the ACSS method in the i th site is calculated as (Hassanzadeh et al. 2011):

$$\text{Cost Function} = \text{Minimize} \left\{ \frac{\sum_{j=1}^{n_i} (S_{ij} - \hat{S}_{ij})^2}{\sum_{j=1}^{n_i} (S_{ij} - \bar{S}_i)^2} \right\} \quad (17)$$

where S_{ij} is the j th observed severity of the site i ; \hat{S}_{ij} is the j th computed severity of the site i ; and \bar{S}_i is the average of observed severity of the site i .

After estimating parameters for each site (θ_k^i), the at-site parameters are combined to give regional parameters (θ_k^R) as:

$$\theta_k^R = \frac{\sum_{i=1}^N (n_i \times \theta_k^i)}{\sum_{i=1}^N n_i}, \quad k = 1, 2, 3, 4 \quad (18)$$

Model 2: Based on this model, after combining the sites, the regional parameters of these sites are estimated using the proposed methods. For calculating the cost function, the combined data are used instead of at-site data.

Model 3: Finally, in the third model, all data of each site are used to estimate the regional parameters. In other words, the cost function is redefined in a way that the result will be matched with all sites with a small error, i.e.,

$$\text{Cost Function} = \text{Minimize} \left\{ \sum_{i=1}^N \left[\frac{\sum_{j=1}^{n_i} (S_{ij} - \hat{S}_{ij})^2}{\sum_{j=1}^{n_i} (S_{ij} - \bar{S}_i)^2} \right] \right\} \quad (19)$$

Due to the nature of the defined optimization problem, the regional parameters of KAP distribution can only be estimated by the ACSS method as a meta-heuristic algorithm.

Goodness-of-fit measure

Before conducting a hydrological frequency analysis, the best-fit distribution for interested variables must be selected by a goodness-of-fit measure. There are different goodness-of-fit measures, such as Z^{DIST} -based test (Hosking & Wallis 1993) and LMOM ratios-based test (Liou et al. 2008). The goodness-of-fit measure, Z^{DIST} , which is utilized in this study for selecting the best regional distribution, is defined as:

$$Z^{DIST} = \frac{(\tau_4^{DIST} - t_4^R + B_4)}{\sigma_4} \quad (20)$$

where τ_4^{DIST} is the L-CK of the fitted distribution to the data using the candidate distribution, $DIST$ can be any of three-parameter distributions, such as generalized Pareto (GPA), generalized extreme value (GEV), generalized logistic (GLO), lognormal (LN3) or Pearson type III (PE3). The

value of τ_4 for each probability distribution can be expressed by the following polynomial approximations (Wu et al. 2012):

$$\tau_4 = \sum_{k=0}^8 A_k \tau_3^k \quad (21)$$

The coefficient A_k was given by Hosking & Wallis (1997, Table A3).

Also, t_4^R is the regional L-CK and B_4 and σ_4 are the bias and the standard deviation of t_4^R , respectively, as:

$$B_4 = \sum_{m=1}^{N_{sim}} \frac{(t_4^{[m]} - t_4^R)}{N_{sim}} \quad (22)$$

$$\sigma_4 = \left[(N_{sim} - 1)^{-1} \left\{ \sum_{m=1}^{N_{sim}} (t_4^{[m]} - t_4^R)^2 - N_{sim} B_4^2 \right\} \right]^{0.5} \quad (23)$$

where $t_4^{[m]}$ is the regional L-CK values for the m th simulated region, and N_{sim} is the number of simulated regions with N sites. As mentioned in the heterogeneity measure, the KAP distribution is used in a similar way for the simulated region. For selected distribution, if $|Z^{DIST}| \leq 1.64$, then it can be suitable as a regional distribution (Hosking & Wallis 1997).

Estimation of the regional distribution

After choosing the suitable regional frequency distribution (s), its parameters and quantiles can be obtained. For a homogenous region with N sites, the quantile of nonexceedance probability F for each site ($\hat{Q}_i(F)$) is expressed as:

$$\hat{Q}_i(F) = \hat{\mu}_i \hat{q}(F) \quad (24)$$

where $\hat{\mu}_i$ is the scale factor (mean of data at site i), and $\hat{q}(F)$ is the regional growth curve. The dimensionless rescaled data $q_{ij} = Q_{ij}/\hat{\mu}_i$, with $i = 1, \dots, N$ and $j = 1, \dots, n_i$, are the basis for estimation $\hat{q}(F)$. Then, the regional growth curve is computed using these parameters and the nonexceedance probability F . Finally, Monte Carlo simulation is utilized to quantify the accuracy of the quantile estimates for various probabilities. Three different measures, such

as the regional average relative bias ($B^R(F)$), regional average absolute relative bias ($A^R(F)$), and regional average relative RMSE ($R^R(F)$) have been defined by Hosking & Wallis (1997) to fulfill this aim and here we utilized all of them in assessing the accuracy of quantile estimates by the following formulas:

$$B^R(F) = N^{-1} \sum_{i=1}^N \left\{ N_{sim}^{-1} \sum_{m=1}^{N_{sim}} \frac{\hat{Q}_i^{[m]}(F) - \hat{Q}_i(F)}{\hat{Q}_i(F)} \right\} \quad (25)$$

$$A^R(F) = N^{-1} \sum_{i=1}^N \left| N_{sim}^{-1} \sum_{m=1}^{N_{sim}} \frac{\hat{Q}_i^{[m]}(F) - \hat{Q}_i(F)}{\hat{Q}_i(F)} \right| \quad (26)$$

$$R^R(F) = N^{-1} \sum_{i=1}^N \left[N_{sim}^{-1} \sum_{m=1}^{N_{sim}} \left\{ \frac{\hat{Q}_i^{[m]}(F) - \hat{Q}_i(F)}{\hat{Q}_i(F)} \right\}^2 \right]^{0.5} \quad (27)$$

where $\hat{Q}_i^{[m]}(F)$ is the site- i quantile estimate for nonexceedance probability F at the m th simulated region.

Cluster analysis

The cluster analysis is used to assemble objects into a set of specific groups with a maximum similarity between the cluster members (Modarres 2010). Generally, the clustering algorithms can be classified into two types: hierarchical and partitional (Demirel et al. 2012). Among the different partitional algorithms, the commonest and most well-known method is the K-means method (Sönmez & Kömüçü 2011; Kar et al. 2012; Rahman et al. 2013).

Since the results can be affected by variables with different units, the data should be normalized with appropriate transformation functions before applying the K-means clustering method. Therefore, the catchment characteristics (variables/attributes of each site) should be rescaled as (Chen et al. 2011; Dikbas et al. 2012; Kar et al. 2012):

$$Y = \frac{(X - X_{min})}{(X_{max} - X_{min})} \quad (28)$$

where Y is the normalized value, and X_{min} and X_{max} are the minimum and maximum values for each variable in the data set (X), respectively.

To determine the optimal number of clusters (k), and to check the success of cluster assigning, the cluster validity indices are extensively used in a data set (Rousseeuw 1987; Rao & Srinivas 2006). In the present study, the silhouette width is used for this aim as:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (29)$$

where $a(i)$ is the average dissimilarity of the i th object with all other objects within the same cluster and $b(i)$ is the lowest average dissimilarity of the i th object with all other objects within any neighbor cluster. For the given k clusters, the average silhouette width ($\bar{S}(k)$) is the average of the silhouette widths over all objects (Kaufman & Rousseeuw 1990).

Adaption to droughts

Since droughts are regional in nature, it is logical to utilize RFA for drought assessment (Tallaksen et al. 1997). Various indices have been introduced to define drought characteristics, and commonly, the standardized precipitation index (SPI) is the most popular and widely used one for drought analysis (Raziei et al. 2009; Serinaldi et al. 2009; Santos et al. 2011; Mirabbasi et al. 2012; Chen et al. 2013). Among different time scales, the SPI values on a six-month time scale (SPI-6) is most useful for describing the shallow soil moisture available to agricultural crops (Reddy & Singh 2014; Abdi et al. 2016a, 2016b). After computing the SPI-6 values, the drought characteristics can be obtained from these values. In this study, we considered the drought severity (S_d) as one of the main characteristics of drought. The drought severity is defined based on the absolute value of cumulative SPI values within the duration of a drought event (Chen et al. 2013). The drought duration is defined as the number of consecutive intervals (months) where the SPI values are less than the threshold value of zero (Mirabbasi et al. 2012).

In order to adopt the LMOM approach to the drought severity, the details of the methodology can be summarized as follows:

- Consider a number of sites within the region.

- Assemble the monthly precipitation data of each site.
- Calculate the SPI values and extract the drought characteristics.
- Compute the discordancy and heterogeneity measures.
- Specify the homogeneous sub-regions by using the cluster analysis.
- Determine the best regional distribution for each sub-region.

Study area and data

The study area is the East-Azerbaijan province, northwest of Iran, which covers an area of over 45,491 km² (Figure 1). Monthly precipitation data for a period of 31 years from 1982 to 2013 are extracted for 60 gauging sites from the East-Azerbaijan Meteorological Organization. The recorded minimum and maximum air temperature is -26 and $+32$ °C, respectively, with a weighted average of 10.2 °C (Zarghami et al. 2011). The climate of the region is classified as semi-arid cold climate based on the Amberg climate classification. The attributes at all sites in the study are shown in Table 1. The summary statistics of regional drought severity calculated for all sites based on the LMOM are $(L-CV)_R = 0.6548$, $(L-CS)_R = 0.4807$, and $(L-CK)_R = 0.2440$.

RESULTS

Assumption of homogeneous region for all sites

First, the entire the study area is assumed as one homogeneous region. Hence, the classic discordancy (CD_i) measure and robust discordancy (RD_i) measure with five different values of α are computed and the discordant sites are shown in Table 2. As can be seen, sites 2, 43, 51, and 56 are the discordant sites according to the values of CD_i and RD_i . The difference between CD_i and RD_i with $\alpha = 0.75, 0.95$ is only at site 13, and with $\alpha = 0.80, 0.85, 0.90$, it is at sites 3 and 13.

Then, the regional parameters of the KAP distribution are estimated based on three proposed models. For model 1, the results of at-site parameters and the values of two main goodness-of-fit tests, the root mean square error (RMSE) and efficiency coefficient (EC), based on the

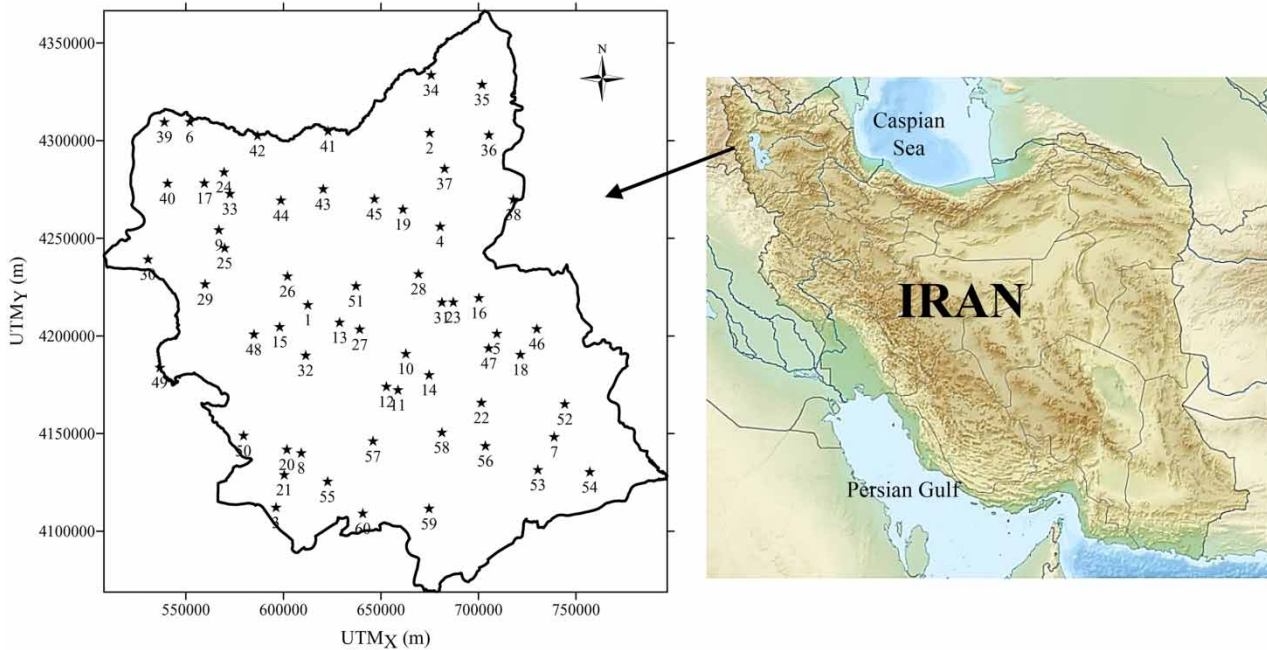


Figure 1 | Geographical location of the study area and sites.

Table 1 | Considered attributes in this study

Attribute	Statistic	Range
Annual precipitation	Average (mm)	199.9 to 508.3
	Maximum (mm)	306.0 to 1,101.7
	Minimum (mm)	70.1 to 287.3
	Coefficient of variation	0.2 to 0.6
Drought severity	Average	4.2 to 6.4
	Maximum	14.5 to 84.7
	Minimum	0.0 to 0.1
	Coefficient of variation	0.7 to 2.9
	L-CV	0.4 to 0.8
Geographical coordination	L-CS	0.1 to 0.8
	L-CK	-0.1 to 0.7
	Latitude (N)	37.1 to 39.1
	Longitude (E)	45.4 to 47.9

LMOM and ACSS methods, are presented in Tables 3 and 4 for the selected sites, respectively. Tables 3 and 4 indicate that the results of the ACSS method are better than the LMOM method for all sites, especially for sites 13, 51, and 60 where the EC value becomes smaller than 0.9 for LMOM, whereas the performances of ACSS are adequate

for these sites. Also, the values of ξ and α parameters of the LMOM results for sites 22, 42, 52, and 58 are significantly different in comparison with these parameters for the other sites. Therefore, the regional parameters can be extensively affected by the weighted ratio $(n_i / \sum_{i=1}^N n_i)$ of model 1. In contrast, the values of these parameters have some small variations for the results of the ACSS method.

The regional parameters of KAP distribution based on model 1 are given in Table 5. The results of the goodness-of-fit tests by using the regional parameters of model 1 for each site demonstrate that LMOM performs worse than ACSS. For 97% of sites, the results of EC are less than zero for LMOM while for ACSS only 8% of sites have $EC < 0$ and in return 65% and 30% of sites have EC greater than 0.5 and 0.8, respectively.

The regional parameters of KAP distribution for model 2 are given in Table 5. The results of the goodness-of-fit tests obtained by the regional parameters show that the performance of the LMOM and ACSS methods is close. However, the value of the sum of differences of the EC (SDEC) measure is equal to 0.4631, which shows the relative superiority of ACSS. In both methods, the value of EC for site 43 is less than zero. As shown in Table 2, site 43 is a discordant

Table 2 | The discordant sites for the study area

Discordancy measure		Robust					
		$\alpha = 0.75$	$\alpha = 0.80$	$\alpha = 0.85$	$\alpha = 0.90$	$\alpha = 0.95$	
Classic		2(3.45), 43(5.05), 51(3.24), 56(3.11) ^a	2(3.20), 13(3.67), 43(4.92), 51(3.88), 56(4.08)	2(3.06), 13(3.80), 43(4.83), 51(3.93), 56(3.85)	2(3.82), 3(3.40), 13(4.10), 43(4.78), 51(4.18), 56(3.82)	2(3.82), 3(3.40), 13(4.10), 43(4.78), 51(4.18), 56(3.82)	2(3.53), 13(3.12), 43(4.87), 51(3.13), 56(3.73)

^aSite number (D_i value).

site and has the maximum value of the discordancy measure among the others. Therefore, this model is less sensitive than the first one to the outliers. In addition, the results demonstrate that the percentage of sites which have EC values greater than 0.5 and 0.8, obtained by LMOM are equal to 97% and 72%, respectively. These values are 97% and 75% for ACSS, respectively.

Table 5 presents the regional parameters of the KAP distribution estimated from model 3. Based on these parameters, the value of the SDEC measure for this model is equal to 1.8301. The results of the ACSS method for this model are more accurate than those for the second model. The value of EC in model 3 for site 43 is equal to 0.3345, whereas less than zero was obtained for the previous model. Indeed, this indicates that the sensitivity of model 3 to the discordant sites is less than model 2. The values of EC obtained by the ACSS results for model 3 show that 98% and 87% of the sites have values greater than 0.5 and 0.8, respectively. Consequently, model 3 yields the global optimal parameters, while the optimum parameters of model 2 are local.

In order to demonstrate the accuracy of the proposed ACSS, some other MHA such as the standard CSS, PSO, and GA are also utilized to compare the results of the best model (model 3). The regional parameters and the mean goodness-of-fit tests obtained by the three algorithms are presented in Tables 5 and 6, respectively. Table 6 indicates the best value of the objective function obtained by the ACSS method. The mean values of EC and mean values of RMSE for the results of ACSS have maximum and minimum values compared to the other methods, respectively. In addition, the convergence rate histories for these algorithms are illustrated in Figure 2. This figure shows the superiority of the ACSS compared to the other algorithms in convergence speed and accuracy.

According to the results of various models and methods, model 3 with the ACSS method is selected for estimating the regional parameters of the KAP distribution. However, with regional parameters of the KAP distribution, sample sizes the same as at-site historical data and $N_{sim} = 1,000$, the heterogeneity and goodness-of-fit measures are calculated as shown in Table 7, which indicate the study area with 60 sites is 'definitely heterogeneous' before removal of the discordant sites according to the H_1 value. Also, based on the

Table 3 | The at-site parameters and goodness-of-fit tests obtained by the LMOM method for the selected sites

Site number	Parameters of KAP distribution				Goodness-of-fit tests	
	ξ	α	k	h	EC	RMSE
13	-0.3371	0.3914	-0.6539	2.0483	0.8579	1.0119
22	-15.8244	19.6832	0.9210	4.3378	0.9809	0.1882
42	-21.7919	23.9107	0.8254	5.4338	0.9853	0.1879
51	-0.3343	0.3368	-0.6962	2.2837	0.8293	1.1716
52	-41.4305	95.4353	2.1477	3.5142	0.9515	0.2266
58	-16.3784	25.8720	1.2713	3.6308	0.9894	0.1179
60	-3.9450	3.6644	0.2142	3.4678	0.8728	0.5099

Table 4 | The at-site parameters and goodness-of-fit tests obtained by the ACSS method for the selected sites

Site number	Parameters of KAP distribution				Goodness-of-fit tests	
	ξ	α	k	h	EC	RMSE
13	-0.3525	0.2005	-1.0000	3.1951	0.9975	0.1334
22	-4.7943	5.7087	0.5174	2.9539	0.9863	0.1595
42	-7.4693	7.8411	0.5289	3.6934	0.9885	0.1658
51	-0.7334	0.2095	-1.0000	4.9999	0.9988	0.0996
52	-3.9892	7.5960	1.0636	2.0900	0.9585	0.2096
58	-5.9095	8.9804	0.9055	2.6776	0.9897	0.1163
60	-8.9238	9.9841	0.6695	3.9585	0.9314	0.3744

Table 5 | The regional parameters of the KAP distribution based on the three models

Model number	Method	Parameters of the KAP distribution			
		ξ	α	K	h
1	LMOM	-3.3795	5.0803	0.1223	2.3682
	ACSS	-2.3047	2.9404	0.0998	2.338
2	LMOM	-1.0803	1.4655	-0.0731	2.0612
	ACSS	-1.0179	1.4671	-0.0613	1.9648
3	ACSS	-0.7107	1.4442	0.0256	1.6082
	CSS	-0.6209	1.3793	0.0109	1.5196
	PSO	-0.4925	1.2782	-0.0167	1.3974
	GA	-0.5197	1.2441	-0.0371	1.4939

Table 6 | The mean values of the at-site EC and RMSE values using the regional parameters based on model 3

Method	Best value of cost function	Mean of EC	Mean of RMSE
ACSS	7.6924	0.8718	0.5023
CSS	7.6958	0.8717	0.5028
PSO	7.7079	0.8715	0.5029
GA	7.7391	0.8710	0.5033

results of goodness-of-fit measure (Z^{DIST}), only the PE3 distribution is selected for the study area.

The LMOM ratio diagrams in terms of the L-CV, L-CS and L-CK of drought severity are shown in Figures 3 and 4, respectively. As expected, these figures demonstrate

that the entire region is heterogeneous. Data points are closely scattered in Figures 3 and 4. This is also supported by the results obtained from the regional heterogeneity measure. For example, site 43 has a maximum value based on classic and robust discordancy measures which is separated from the other sites. Also, sites 13 and 51 are separated from the other sites and the robust discordancy measure with different values of α confirm this concept, whereas according to the classic discordancy measure, only site 51

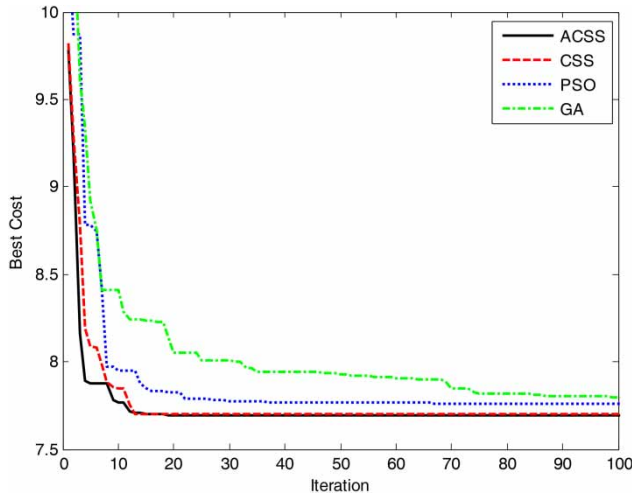


Figure 2 | Comparison of the convergence histories for model 3 obtained by different meta-heuristic methods.

is discordant. Consequently, it seems that the obtained number of discordant sites according to the classic discordancy measure is not exactly correct.

At last, after removing the discordant sites based on the classic and robust discordancy measures, the heterogeneity and goodness-of-fit measures are computed as demonstrated in Table 7. A value of $H_1 = 2.13$ in this table indicates that the study area is ‘definitely heterogeneous’. On the contrary, using the robust discordancy measure with different values of subset factor for recognizing and removing the discordant sites shows an improvement in heterogeneity measure and suggests that the whole region is ‘possibly heterogeneous’. Also, based

on the goodness-of-fit measure, the PE3 distribution is selected as the regional distribution for this region.

Clustering and identification of homogeneous regions

In order to identify the optimum number of clusters, the average silhouette width for some number of clusters are computed and illustrated in Table 8. According to this table, the maximum value of average silhouette width corresponds to the best number of clusters (k), which is equal to 3.

The values of the silhouette index for all sites related to $k = 3$ are presented in Figure 5. The silhouette index for a few sites in cluster 2 is less than zero, so the quality of clustering may not be good. Thus, the exact results can be achieved by using the heterogeneity measure for each sub-region. Figure 6 shows the location of the sites in the identified clusters.

The classic discordancy measure and robust discordancy measure with five different values of α are computed and the discordant sites for three clusters are presented in Table 9. The heterogeneity and goodness-of-fit measures are demonstrated in Table 10, as well. The results in Table 10 show that the third cluster is ‘acceptably homogeneous’, the first cluster is ‘possibly heterogeneous’, and the second one is ‘definitely heterogeneous’ according to the H_1 test.

In order to improve the heterogeneity measure, Table 9 demonstrates that the classic discordancy measure could not detect any discordant site for cluster 1. Also, only one discordant site is detected for cluster 2, whereas the different discordant sites are identified based on the

Table 7 | The heterogeneity and goodness-of-fit measures for the study area

		After removing discordant sites based on discordancy measure						
Measure	Before removing discordant sites	Classic	Robust					
			$\alpha = 0.75$	$\alpha = 0.80$	$\alpha = 0.85$	$\alpha = 0.90$	$\alpha = 0.95$	
Heterogeneity	H_1	3.3	2.13	1.51	1.29	1.29	1.29	1.51
	H_2	5.06	4.12	3.49	3.33	3.33	3.33	3.49
	H_3	6.68	5.58	4.66	4.53	4.53	4.53	4.66
Goodness-of-fit	GPA	-3.8	-4.71	-4.68	-4.62	-4.62	-4.62	-4.68
	GEV	-7.28	-8.54	-8.48	-8.36	-8.36	-8.36	-8.48
	GLO	-8.01	-9.73	-9.3	-9.18	-9.18	-9.18	-9.3
	LN3	-4.39	-5.64	-5.4	-5.38	-5.38	-5.38	-5.4
	PE3	0.51	0.53	-0.45	-0.31	-0.31	-0.31	-0.45

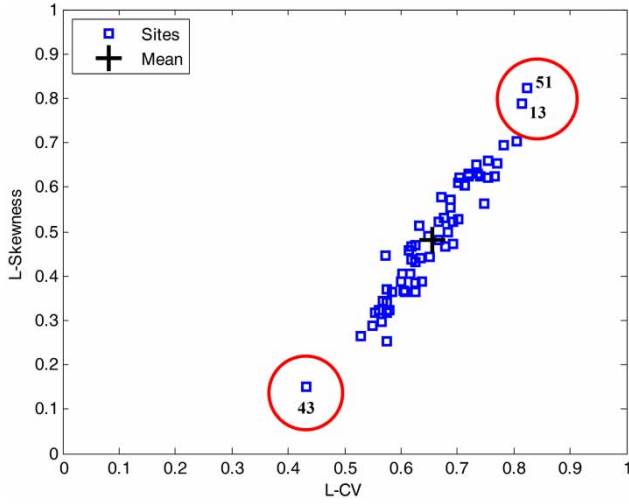


Figure 3 | The LMOM ratio diagram in terms of L-CV and L-CS.

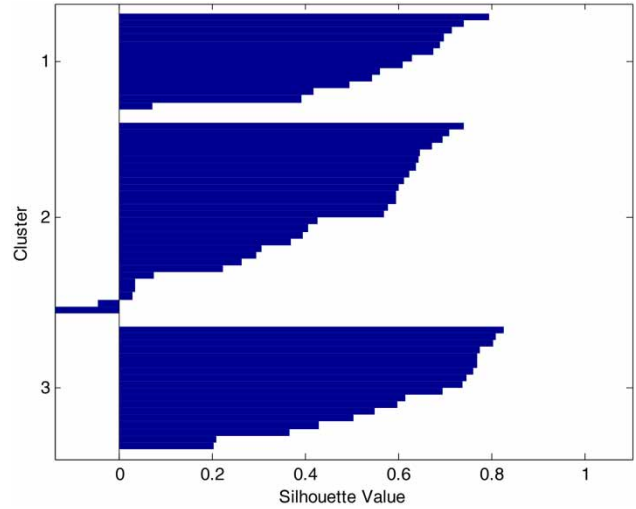


Figure 5 | The silhouette plot.

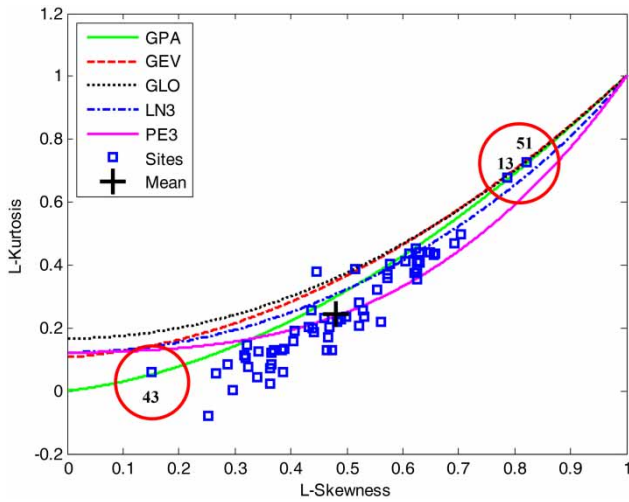


Figure 4 | The LMOM ratio diagram in terms of L-CS and L-CK.

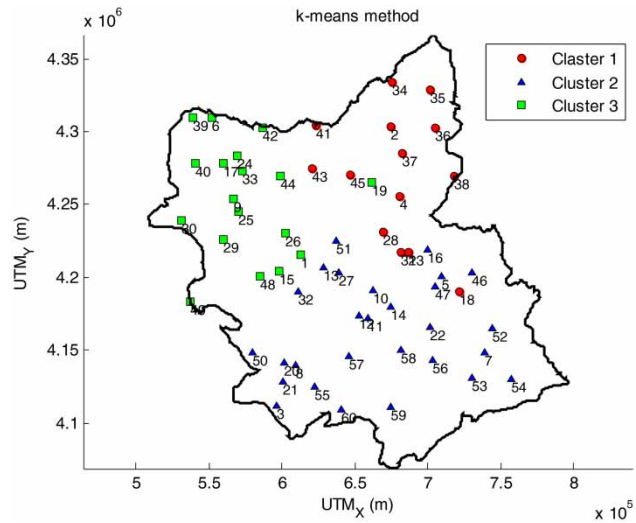


Figure 6 | The location of sites for three clusters.

Table 8 | The average silhouette width for a number of clusters

Number of clusters (<i>k</i>)	Average silhouette width
2	0.4315
3	0.5122
4	0.4907
5	0.4704
6	0.4815
7	0.5050

The values in bold indicate the optimum number of cluster and its average silhouette width.

Table 9 | The discordant sites for the all clusters

Cluster	Classic	Discordancy measure				
		Robust				
		$\alpha = 0.75$	$\alpha = 0.80$	$\alpha = 0.85$	$\alpha = 0.90$	$\alpha = 0.95$
1	-	23, 37, 43	37, 43	37, 43	43	43
2	56	13, 51, 56	13, 51, 53, 56	13, 51, 56	13, 51, 56	53, 56
3	-	42, 48	42, 48	42, 48	42, 48	42

Table 10 | The heterogeneity and goodness-of-fit measures for all clusters

Measure	Cluster		Before removing discordant sites	After removing discordant sites based on discordancy measure					
				Classic	Robust				
					$\alpha = 0.75$	$\alpha = 0.80$	$\alpha = 0.85$	$\alpha = 0.90$	$\alpha = 0.95$
Heterogeneity	H ₁	1	1.58	1.58	-0.93	-0.70	-0.70	0.29	0.29
		2	2.14	2.03	1.37	0.87	1.37	1.37	1.63
		3	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	H ₂	1	2.09	2.09	0.50	0.59	0.59	1.24	1.24
		2	3.44	3.55	2.75	2.19	2.75	2.75	3.13
		3	0.85	0.85	0.85	0.85	0.85	0.85	0.85
	H ₃	1	3.25	3.25	1.90	1.72	1.72	2.36	2.36
		2	5.10	5.31	3.93	3.49	3.93	3.93	5.09
		3	1.10	1.10	1.10	1.10	1.10	1.10	1.10
Goodness-of-fit	GPA	1	-2.48	-2.48	-2.52	-2.58	-2.58	-2.52	-2.52
		2	-2.88	-3.25	-3.66	-3.71	-3.66	-3.66	-3.13
		3	-1.93	-1.93	-1.93	-1.93	-1.93	-1.93	-1.93
	GEV	1	-5.15	-5.15	v4.81	-4.86	-4.86	-4.81	-4.81
		2	-5.07	-5.50	-6.00	-5.96	-6.00	-6.00	-5.17
		3	-3.84	-3.84	-3.84	-3.84	-3.84	-3.84	-3.84
	GLO	1	-5.88	-5.88	-5.46	-5.49	-5.49	-5.40	-5.40
		2	-5.44	-5.88	-6.44	-6.37	-6.44	-6.44	-5.49
		3	-4.26	-4.26	-4.26	-4.26	-4.26	-4.26	-4.26
	LN3	1	-3.57	-3.57	-3.50	-3.50	-3.50	-3.33	-3.33
		2	-2.87	-3.23	-3.83	-3.79	-3.83	-3.83	-3.02
		3	-2.33	-2.33	-2.33	-2.33	-2.33	-2.33	-2.33
	PE3	1	-0.87	-0.87	-1.25	-1.19	-1.19	-0.80	-0.80
		2	0.83	0.61	-0.15	-0.13	-0.15	-0.15	0.60
		3	0.23	0.23	0.23	0.23	0.23	0.23	0.23

robust discordancy measure with different values of α . Finally, after removing the discordant sites for clusters 1 and 2, the heterogeneity and goodness-of-fit measures are also shown in Table 10. In this condition, the value of H_1 demonstrates that cluster 2 is not ‘acceptably homogeneous’. On the contrary, after removing the discordant sites (based on robust discordancy measure with all values of α), cluster 1 is found to be ‘acceptably homogeneous’. Just $\alpha = 0.80$ yields ‘acceptably homogeneous’ for cluster 2 and the other values of α yield ‘possibly heterogeneous’.

The results of goodness-of-fit measure (Table 10) reveal that PE3 has the best distribution among the others. Also, according to Figure 7 which shows the LMOM ratio diagram in terms of the L-CS, L-CK for all clusters, the best regional distribution is PE3 for drought severity.

After the selection of PE3 distribution as a suitable regional frequency distribution, the regional parameters of the respective distribution are estimated using model 3

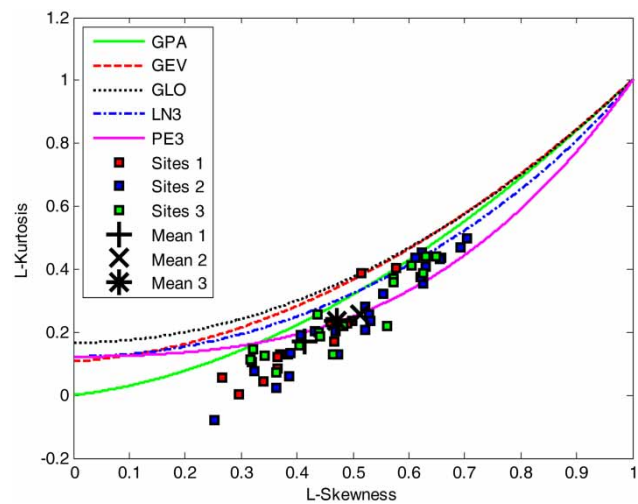


Figure 7 | The LMOM ratio diagram in terms of L-CS and L-CK for three clusters.

and by the ACSS method, then the regional quantiles of various return periods for all clusters are illustrated in Table 11.

Table 11 | Estimation of regional parameters and quantiles for the PE3 distribution

Cluster	Regional parameters			Regional quantiles				
	γ	α	β	T = 10	T = 25	T = 50	T = 100	T = 200
1	1.0000	2.7212	0.1482	9.92	15.88	21.52	28.01	35.26
2	1.0000	3.2201	0.1524	11.61	19.12	26.21	34.33	43.38
3	1.0000	3.1939	0.1461	12.04	19.86	27.28	35.84	45.41

The estimated regional growth curve values and the relative RMSE for each cluster are determined in Table 12. The results verify that by increasing the value of the return period, the accuracy of the regional growth curve decreases. In addition, Monte Carlo simulation is used to assess the accuracy of the estimates of the determined homogeneous regions, by generating the same sample sizes as those of the original sample. The number of random samples, N_{sim} , is set to 10,000 for selected quantiles corresponding to return periods of 10, 25, 50, 100, and 200 years. Table 13 shows the results of the $B^R(F)$, $A^R(F)$, and $R^R(F)$. According to Tables 12 and 13, the values of the RMSE of the estimated growth curve values are always less than the estimated

quantiles for the same return periods, which represent the uncertainty in the estimated scale factor. These results are verified by Saf (2009) and Hussain (2011).

CONCLUSIONS

In the standard CSS algorithm, search agents or CPs affect each other according to laws from electrostatics and mechanics. Here, a new ACSS is developed to improve the performance of CSS. The ACSS and CSS algorithms use the governing laws from electrical physics and Newtonian mechanics to determine the amount and the direction of a

Table 12 | The regional growth values and the relative RMSE values

Cluster	Regional growth curve					RMSE				
	T = 10	T = 25	T = 50	T = 100	T = 200	T = 10	T = 25	T = 50	T = 100	T = 200
1	2.01	3.22	4.37	5.69	7.16	0.1174	0.1582	0.1820	0.2021	0.2196
2	2.25	3.70	5.07	6.65	8.40	0.1464	0.1951	0.2252	0.2516	0.2752
3	2.17	3.57	4.91	6.45	8.17	0.1368	0.1677	0.1792	0.1860	0.1901

Table 13 | Summary of simulation results

Measure	Cluster	Return period (years)				
		T = 10	T = 25	T = 50	T = 100	T = 200
$B^R(F)$	1	-0.1280	-0.2256	-0.2604	-0.2791	-0.2892
	2	-0.1583	-0.2580	-0.2906	-0.3070	-0.3150
	3	-0.1453	-0.2601	-0.3026	-0.3271	-0.3420
$A^R(F)$	1	0.1435	0.2280	0.2604	0.2791	0.2892
	2	0.1634	0.2629	0.3014	0.3243	0.3408
	3	0.1540	0.2601	0.3026	0.3271	0.3420
$R^R(F)$	1	0.2056	0.2753	0.3052	0.3237	0.3358
	2	0.2306	0.3059	0.3384	0.3591	0.3730
	3	0.2311	0.3067	0.3405	0.3611	0.3739

CPs movement. The CSS and ACSS algorithms can distinguish the local search and the global search phases, and utilize suitable relationships in these phases resulting in a good balance between exploration and exploitation. However, applying some controlling limits and the constraints is the main difference between the ACSS and CSS algorithms.

On the other hand, a new model is introduced to estimate the regional parameters of the statistical distributions. Then, the ACSS method is utilized to estimate the accurate parameters and quantile estimates for a region. Also for the Monte Carlo simulation, this new model is employed to generate samples from the selected distribution. In addition, the robust discordancy measure with different subset factors is developed to detect and remove the discordant sites, which causes some improvements in the heterogeneity measure. The data used in this study are drought severity extracted for 60 sites in East-Azerbaijan, Iran. The relevant conclusions can be summarized as follows:

- The ACSS has the ability to improve the search and convergence speed compared to the standard CSS.
- The ACSS algorithm has stable convergence characteristics and good computational ability, and is an effective algorithm compared to other algorithms, such as the GA and PSO, for RFA.
- The ACSS can estimate the global optimum (or at least near it) of parameters without using nonlinear complex equations and any approximations.
- The new model is the best model among the others. This model is based on the participation of all the data of sites in a region for estimating the regional parameters and quantiles, while other models are based on the weighted at-site parameters or the combination of the sites.
- Identifying discordant sites using the robust discordancy measure is more accurate than the classic discordancy measure. Among the different values for the subset factor, the value of 0.8 is found suitable for robust discordancy measure and to detect the discordant sites of a heterogeneous region.
- The study area is divided into three sub-regions, south, northeast, and northwest by using the K-means clustering method and the Pearson type 3 distribution is found the best regional distribution for all sub-regions.

Finally, this method can be applied for RFA of other hydrological phenomena, such as rainfall and stream flow. Also, the multivariate RFA framework based on the multivariate LMOM approach is suggested for future application of this work. In addition, another goodness-of-fit measure, the LMOM ratios-based test, can be applied and compared with the Z^{DIST} -based test for selecting the best regional distribution.

REFERENCES

- Abdi, A., Hassanzadeh, Y., Talatahari, S., Fakheri-Fard, A. & Mirabbasi, R. 2016a Parameter estimation of copula functions using an optimization-based method. *Theor. Appl. Climatol.* doi: 10.1007/s00704-016-1757-2.
- Abdi, A., Hassanzadeh, Y., Talatahari, S., Fakheri-Fard, A. & Mirabbasi, R. 2016b Regional bivariate modeling of droughts using L-components and copulas. *Stoch. Environ. Res. Risk Assess.* doi: 10.1007/s00477-016-1222-x.
- Alameddine, I., Kenney, M. A., Gosnell, R. J. & Reckhow, K. H. 2010 Robust multivariate outlier detection methods for environmental data. *J. Environ. Eng.* **136** (11), 1299–1304.
- Belayneh, A., Adamowski, J., Khalil, B. & Ozga-Zielinski, B. 2014 Long-term SPI drought forecasting in the Awash river basin in Ethiopia using wavelet neural network. *J. Hydrol.* **508**, 418–429.
- Chen, L. H., Lin, G. F. & Hsu, C. W. 2011 Development of design hyetographs for ungauged sites using an approach combining PCA, SOM and Kriging methods. *Water Resour. Manage.* **25** (8), 1995–2013.
- Chen, L., Singh, V. P., Guo, S., Mishra, A. K. & Guo, J. 2013 Drought analysis using copulas. *J. Hydrol. Eng.* **18** (7), 797–808.
- Chérif, R. & Bargaoui, Z. 2013 Regionalisation of maximum annual runoff using hierarchical and trellis methods with topographic information. *Water Resour. Manage.* **27** (8), 2947–2963.
- Demirel, M. C., Booij, M. J. & Kahya, E. 2012 Validation of an ANN flow prediction model using a multistation cluster analysis. *J. Hydrol. Eng.* **17** (2), 262–271.
- Dikbas, F., Firat, M., Koc, A. C. & Gungor, M. 2012 Classification of precipitation series using fuzzy cluster method. *Int. J. Climatol.* **32** (10), 1596–1603.
- Dikbas, F., Firat, M., Koc, A. C. & Gungor, M. 2013 Defining homogeneous regions for streamflow processes in Turkey using a K-means clustering method. *Arab. J. Sci. Eng.* **38**, 1313–1319.
- Dodangeh, E., Soltani, S., Sarhadi, A. & Shiau, J. 2014 Application of L-moments and Bayesian inference for low-flow regionalization in Sefidroud basin, Iran. *Hydrol. Process.* **28** (4), 1663–1676.

- Feng, J., Yan, D., Li, C., Gao, Y. & Liu, J. 2014 Regional frequency analysis of extreme precipitation after drought events in the Heihe river basin, northwest China. *J. Hydrol. Eng.* **19** (6), 1101–1112.
- Goyal, M. K. & Gupta, V. 2014 Identification of homogeneous rainfall regimes in northeast region of India using fuzzy cluster analysis. *Water Resour. Manage.* **28** (13), 4491–4511.
- Hassanzadeh, Y., Abdi, A., Talatahari, S. & Singh, V. P. 2011 Meta-heuristic algorithms for hydrologic frequency analysis. *Water Resour. Manage.* **25** (7), 1855–1879.
- Hosking, J. R. M. 1986 *The theory of probability weighted moments. Research Report RC12210.* IBM Res. Div., Yorktown Heights, New York, USA.
- Hosking, J. R. M. 1990 L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Stat. Soc. Ser. B.* **52** (1), 105–124.
- Hosking, J. R. M. & Wallis, J. R. 1995 Some statistics useful in regional frequency analysis. *Water Resour. Res.* **29** (2), 271–281.
- Hosking, J. R. M. & Wallis, J. R. 1997 *Regional Frequency Analysis: An Approach Based on L-moments.* Cambridge University Press, New York, USA.
- Hosseini, S. Z., Shin, H. M. & Gyewoon, C. 2012 Evaluation of regional droughts using monthly gridded precipitation for Korea. *J. Hydroinform.* **14** (4), 1036–1050.
- Hubert, M., Rousseeuw, P. J. & Van Aelst, S. 2005 Multivariate outlier detection and robustness. In: *Handbook of Statistics, Vol. 24, Data Mining and Data Visualization* (C. R. Rao, E. J. Wegman & J. L. Solka, eds). Elsevier, The Netherlands, pp. 263–302.
- Hubert, M., Rousseeuw, P. J. & Van Aelst, S. 2008 High-breakdown robust multivariate methods. *Stat. Sci.* **23** (1), 92–119.
- Hussain, Z. 2011 Application of the regional flood frequency analysis to the upper and lower basins of the Indus River, Pakistan. *Water Resour. Manage.* **25** (11), 2797–2822.
- Ilorme, F. & Griffis, V. W. 2013 A novel procedure for delineation of hydrologically homogeneous regions and the classification of ungauged sites for design flood estimation. *J. Hydrol.* **492** (7), 151–162.
- Kar, A. K., Goel, N. K., Lohani, A. K. & Roy, G. P. 2012 Application of clustering techniques using prioritized variables in regional flood frequency analysis – case study of Mahanadi basin. *J. Hydrol. Eng.* **17** (1), 213–223.
- Kaufman, L. & Rousseeuw, P. J. 1990 *Finding Groups in Data: An Introduction to Cluster Analysis.* Wiley, New York, USA.
- Kaveh, A. & Talatahari, S. 2010 A novel heuristic optimization method: charged system search. *Acta Mech.* **213** (3–4), 267–289.
- Kaveh, A., Talatahari, S. & Farahmand Azar, B. 2012 Optimum design of composite open channels using charged system search algorithm. *IJST Trans. Civ. Eng.* **36** (C1), 67–77.
- Keshavarz, M., Karami, E. & Vanclay, F. 2013 The social experience of drought in rural Iran. *Land Use Policy* **30**, 120–129.
- Liou, J. J., Wu, Y. C. & Cheng, K. S. 2008 Establishing acceptance regions for L-moments-based goodness-of-fit tests by stochastic simulation. *J. Hydrol.* **355** (1–4), 49–62.
- Mirabbasi, R., Fakheri-Fard, A. & Dinpashoh, Y. 2012 Bivariate drought frequency analysis using the copula method. *Theor. Appl. Climatol.* **108**, 191–206.
- Modarres, R. 2010 Regional dry spells frequency analysis by L-moment and multivariate analysis. *Water Resour. Manage.* **24** (10), 2365–2380.
- Neykov, N. M., Neytchev, P. N., Van Gelder, P. H. A. J. M. & Todorov, V. K. 2007 Robust detection of discordant sites in regional frequency analysis. *Water Resour. Res.* **43**, W06417. doi:10.1029/2006WR005322.
- Pison, G., Van Aelst, S. & Willems, G. 2002 Small sample corrections for LTS and MCD. *Metrika* **55**, 111–123.
- Rahman, M. M., Sarkar, S., Najafi, M. R. & Rai, R. K. 2013 Regional extreme rainfall mapping for Bangladesh using L-moment technique. *J. Hydrol. Eng.* **18** (5), 603–615.
- Rai, R. K., Sarkar, S. & Singh, V. P. 2009 Evaluation of the adequacy of statistical distribution functions for deriving unit hydrograph. *Water Resour. Manage.* **23** (5), 899–929.
- Rajsekhar, D., Mishra, A. K. & Singh, V. P. 2013 Regionalization of drought characteristics using an entropy approach. *J. Hydrol. Eng.* **18** (7), 870–887.
- Rao, A. R. & Hamed, K. H. 2000 *Flood Frequency Analysis.* CRC Press LLC, Boca Raton, FL, USA.
- Rao, A. R. & Srinivas, V. V. 2006 Regionalization of watersheds by fuzzy cluster analysis. *J. Hydrol.* **318** (1–4), 57–79.
- Raziei, T., Saghafian, B., Paulo, A. A., Pereira, L. S. & Bordi, I. 2009 Spatial patterns and temporal variability of drought in western Iran. *Water Resour. Manage.* **23**, 439–455.
- Reddy, M. J. & Ganguli, P. 2012 Application of copulas for derivation of drought severity-duration-frequency curves. *Hydrol. Process.* **26** (11), 1672–1685.
- Reddy, M. J. & Singh, V. P. 2014 Multivariate modeling of droughts using copulas and meta-heuristic method. *Stoch. Environ. Res. Risk Assess.* **28** (3), 475–489.
- Rousseeuw, P. J. 1984 Least median of squares regression. *J. Am. Stat. Assoc.* **79**, 871–880.
- Rousseeuw, P. J. 1985 Multivariate estimation with high breakdown point. In: *Mathematical Statistics and Applications, Vol. B* (W. Grossmann, G. Pflug, I. Vincze & W. Wertz, eds). Springer, Dordrecht, The Netherlands, pp. 283–297.
- Rousseeuw, P. J. 1987 Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65.
- Rousseeuw, P. J. & Van Driessen, K. 1999 A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41** (3), 212–223.
- Rousseeuw, P. J., Debruyne, M., Engelen, S. & Hubert, M. 2006 Robustness and outlier detection in chemometrics. *Crit. Rev. Anal. Chem.* **36** (3–4), 221–242.
- Saf, B. 2009 Regional flood frequency analysis using L-moments for the west Mediterranean region of Turkey. *Water Resour. Manage.* **23** (3), 531–551.
- Saf, B. 2010 Assessment of the effects of discordant sites on regional flood frequency analysis. *J. Hydrol.* **380**, 362–375.

- Santillán, D., Mediero, L. & Garrote, L. 2014 Modelling uncertainty of flood quantile estimations at ungauged sites by Bayesian networks. *J. Hydroinform.* **16** (4), 822–838.
- Santos, J. F., Portela, M. M. & Pulido-Calvo, I. 2011 Regional frequency analysis of droughts in Portugal. *Water Resour. Manage.* **25**, 3537–3558.
- Sarhadi, A. & Heydarizadeh, M. 2014 Regional frequency analysis and spatial pattern characterization of dry spells in Iran. *Int. J. Climatol.* **34**, 835–848.
- Seckin, N., Cobaner, M., Yurtal, R. & Haktanir, T. 2013 Comparison of artificial neural network methods with L-moments for estimating flood flow at ungauged sites: the case of east Mediterranean river basin, Turkey. *Water Resour. Manage.* **27** (7), 2103–2124.
- Serinaldi, F., Bonaccorso, B., Cancelliere, A. & Grimaldi, S. 2009 Probabilistic characterization of drought properties through copulas. *Phys. Chem. Earth.* **34**, 596–605.
- Sheikholeslami, R., Kaveh, A., Tahershamsi, A. & Talatahari, S. 2014 Application of charged system search algorithm to water distribution networks optimization. *Int. J. Optim. Civ. Eng.* **4** (1), 41–58.
- Shin, J. Y., Heo, J. H., Jeong, C. & Lee, T. 2014 Meta-heuristic maximum likelihood parameter estimation of the mixture normal distribution for hydro-meteorological variables. *Stoch. Environ. Res. Risk Assess.* **28** (2), 347–358.
- Sönmez, İ. & Kömüscü, A. Ü. 2011 Reclassification of rainfall regions of Turkey by K-means methodology and their temporal variability in relation to North Atlantic Oscillation (NAO). *Theor. Appl. Climatol.* **106** (3–4), 499–510.
- Talatahari, S., Sheikholeslami, R., Farahmand Azar, B. & Daneshpajouh, H. 2013 Optimal parameter estimation for Muskingum model using a CSS-PSO method. *Adv. Mech. Eng.* doi: 10.1155/2013/480954.
- Tallaksen, L. M., Madsen, H. & Clausen, B. 1997 On the definition and modelling of streamflow drought duration and deficit volume. *Hydrol. Sci. J.* **42** (1), 15–33.
- Verboven, S. & Hubert, M. 2005 LIBRA: a MATLAB library for robust analysis. *Chemometrics Intell. Lab. Syst.* **75** (2), 127–136.
- Viglione, A., Laio, F. & Claps, P. 2007 A comparison of homogeneity tests for regional frequency analysis. *Water Resour. Res.* **43**, W03428. doi:10.1029/2006WR005095.
- Wu, Y. C., Liou, J. J., Su, Y. F. & Cheng, K. S. 2012 Establishing acceptance regions for L-moments based goodness-of-fit tests for the Pearson type III distribution. *Stoch. Environ. Res. Risk Assess.* **26**, 873–885.
- Zakaria, Z. A. & Shabri, A. 2013 Regional frequency analysis of extreme rainfalls using partial L moments method. *Theor. Appl. Climatol.* **113** (1–2), 83–94.
- Zarghami, M., Abdi, A., Babaeian, I., Hassanzadeh, Y. & Kanani, R. 2011 Impacts of climate change on runoffs in East Azerbaijan, Iran. *Glob. Planet. Change* **78** (3), 137–146.
- Zhang, Q., Xiao, M., Singh, V. P. & Li, J. 2012 Regionalization and spatial changing properties of droughts across the Pearl river basin, China. *J. Hydrol.* **472–473**, 355–366.

First received 21 November 2015; accepted in revised form 17 November 2016. Available online 17 December 2016