Agricultural and Forest Meteorology 200 (2015) 57–65

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Clustering project management for drought regions determination: A case study in Serbia

Shahaboddin Shamshirband[a,i,*], Milan Gocić[b], Dalibor Petković[c], Hossein Javidnia[d], Siti Hafizah Ab Hamid[e], Zulkefli Mansor[f], Sultan Noman Qasem[g,h]

[a] Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia
[b] University of Nis, Faculty of Civil Engineering and Architecture, Aleksandra Medvedeva 14, 18000 Nis, Serbia
[c] University of Nis, Faculty of Mechanical Engineering, Deparment for Mechatronics and Control, Aleksandra Medvedeva 14, 18000 Nis, Serbia
[d] University of Guilan, Faculty of Engineering, Department of Computer Engineering, Iran
[e] Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
[f] Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Selangor, Bestari Jaya, Selangor, Malaysia
[g] Computer Science Department, College of Computer and Information Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia
[h] Department of Computer Science, Faculty of Applied Sciences, Taiz University, Yemen
[i] Department of Computer Science, Chalous Branch, Islamic Azad University (IAU), 46615-397 Chalous, Mazandaran, Iran

ARTICLE INFO

Article history:
Received 23 July 2014
Received in revised form 9 September 2014
Accepted 21 September 2014

Keywords:
Clustering
Drought
Fuzzy *c*-means
k-Medoids
Imperialist competitive algorithm
Serbia

ABSTRACT

Analyses of drought regions require long-term historical data to ensure reliable drought indices estimate. Therefore, various indices have been used to measure different drought characteristics depending on research objectives. The present study investigates the application of clustering methods on the standardized precipitation index (SPI) at the 12-month timescale values in Serbia to detect district drought clusters. The principal component analysis (PCA) was applied to capture the drought patterns with similar drought features, while the cluster analysis was used as one of the major data analysis technique. Thus, three clustering algorithms namely fuzzy *c*-means (FCM), *k*-medoids and imperialist competitive algorithm (ICA) were analyzed in this research work. These algorithms are implemented by means of practical approach to segment the drought regions (clusters). In this way, three different drought clusters were detected. Statistical results indicate that the *k*-medoids clustering method was more effective and efficient than the FCM and ICA. The ICA clustering technique had the worst classification capability. The obtained results confirm usefulness of clustering methods for drought regionalization.

1. Introduction

Drought is characterized by the lack of precipitation on whose occurrence varies in frequency, severity and duration (Wilhite and Glantz, 1985; Mishra and Singh, 2010), which plays an essential role in planning and management of water and agricultural resources. It is represented by using the drought indices. In this study, the standardized precipitation index (SPI) (McKee et al., 1993, 1995), one of the most common indices for detecting and monitoring drought, is used. SPI is recommended as a probabilistic drought index (Guttman, 1998; Hayes et al., 1999), which estimation is based on monthly precipitation data.

During the last decade, a large number of scientists worldwide have analyzed the drought patterns using the principal component analysis (PCA) and drought indices (Vicente-Serrano et al., 2004; Martins et al., 2012; Potop et al., 2014; Raziei et al., 2013). In Serbia, Tosic (2004) used PCA and spectral analysis to study the spatial and temporal variability of the winter and summer precipitation, while Stricevic and Djurovic (2013) applied the PCA with varimax rotation on precipitation data of Central Serbia to investigate spatial patterns of agricultural drought. Furthermore, Gocic and Trajkovic (2014a) applied SPI, S-mode PCA and agglomerative hierarchical cluster analysis to identify three different drought sub-regions.

On the other hand, clustering techniques can be used for segmentation drought areas. Clustering is the unsupervised classification of different data items into groups (clusters). A resulting partition should possess the following properties (Sheikh et al., 2008): (1) homogeneity within the clusters, i.e. data that belong to the same cluster should be as similar as possible, and (2)

* Corresponding author. Tel.: +6 0146266763; fax: +6 0146266763.
E-mail address: shamshirband1396@gmail.com (S. Shamshirband).

http://dx.doi.org/10.1016/j.agrformet.2014.09.020

heterogeneity between clusters, i.e. data that belong to the different cluster should be as different as possible.

In recent years, a variety of clustering approaches (Baraldi and Blonda, 1999; Berkhin, 2006; Filippone et al., 2008; Mei and Chen, 2010; Rokach, 2010; Duan and Huang, 2014; Velmurugan, 2014; Zhao et al., 2014) have been developed for applications in specific areas. Cluster analysis using *k*-means and fuzzy *c*-mean (FCM) methods was applied for the assessment of water quality for a large water distribution network (Chang et al., 2011). To identify the zone structure distribution of aquifer parameters, the kernel-based FCM clustering method was used (Ayvaz et al., 2007). Santos et al. (2010) applied PCA and *k*-means clustering to the SPI series to assess spatial and temporal patterns of droughts in Portugal. Sadri and Burn (2011) used FCM to form the clusters based on partial or fuzzy membership of each of the 36 unregulated flow monitoring sites located in the three Canadian provinces, while Zhang et al. (2012) categorized the Pearl River Basin into four homogenous regions using FCM. Yoo et al. (2012) employed a *k*-means clustering approach for regionalization of drought extremes based on drought attributes (e.g., duration and severity). To achieve spatial clustering, Ping et al. (2013) applied *k*th nearest-neighbor algorithm based on point process theory to the study of clustering extreme climatic events.

The three clustering approaches (FCM, *k*-medoids and imperialist competitive algorithm (ICA)) were applied to identify zones with similar drought patterns. In general, the goal of using cluster methods is to find an optimal grouping for which the observations or objects within each cluster are similar, but the clusters are dissimilar to each other. In the FCM algorithm, fuzzy membership values are assigned to each data point based on the relative distance of that point to the cluster centers. It groups the data points into a specific number of different clusters. *k*-Medoids approach produces a flat partition of the data set with *k* non-overlapped clusters that minimizes the total intra-cluster distance. In *k*-medoids, the center of a cluster is one of the real objects in the data set. ICA is an evolutionary algorithm that implements the socio-political process of imperialism of controlling many countries and using their sources once colonies are dominated by rules. If one empire loses its power, the others will compete to take its place.

The main contributions of this paper are as follows: (1) to use PCA and three clustering approaches (FCM, *k*-medoids and ICA) to determine the drought regions according to the SPI values and (2) to evaluate the values of clustering analysis.

2. Materials and methods

2.1. Clustering project management

Project management plays important role in improving the quality of projects (Ismail et al., 2012). Every single project requires complicated and extensive process as technologies used are from different disciplines and merging rapidly. An unambiguous and plain requirements act as a benchmark to have a successful project outputs. However, requirements are also changing as project moving on. The effect requirements will give impact to redundant functions and decision making. To overcome the rapid changing of requirements, the project must be managed within a clear defined framework focuses on the definition and delivery of the results. In this article, clustering project management approach enables to ensure the comparative analysis of the several clustering techniques. The three clustering approaches (FCM, *k*-medoids and imperialist competitive algorithm (ICA)) were applied to identify zones with similar drought patterns in Serbia. In general, the goal of using cluster methods is to find an optimal grouping for which the observations or objects within each cluster are similar, but the clusters are dissimilar to each other.

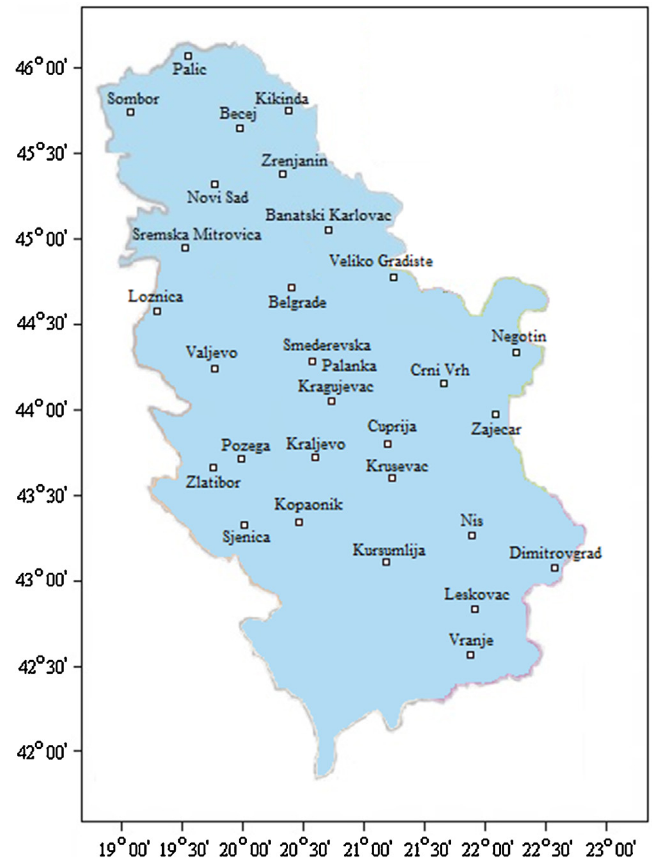

Fig. 1. Spatial distribution of the 29 synoptic stations in Serbia map.

Table 1
Geographical descriptions of the synoptic stations used in the study.

Station name	Longitude (E)	Latitude (N)	Elevation (m a.s.l.)
1. Banatski Karlovac	20°48′	45°03′	89
2. Becej	20°04′	45°37′	78
3. Belgrade	20°28′	44°48′	132
4. Crni Vrh	21°58′	44°08′	1027
5. Cuprija	21°22′	43°56′	123
6. Dimitrovgrad	22°45′	43°01′	450
7. Kikinda	20°28′	45°51′	81
8. Kopaonik	20°48′	43°17′	1711
9. Kragujevac	20°56′	44°02′	185
10. Kraljevo	20°42′	43°43′	215
11. Krusevac	21°21′	43°34′	166
12. Kursumlija	21°16′	43°08′	383
13. Leskovac	21°57′	42°59′	230
14. Loznica	19°14′	44°33′	121
15. Negotin	22°33′	44°14′	42
16. Nis	21°54′	43°20′	204
17. Novi Sad	19°51′	45°20′	86
18. Palic	19°46′	46°06′	102
19. Pozega	20°02′	43°50′	310
20. Sjenica	20°01′	43°16′	1038
21. Sombor	19°05′	45°47′	87
22. Smederevska Palanka	20°57′	44°22′	121
23. Sremska Mitrovica	19°38′	44°58′	82
24. Valjevo	19°55′	44°17′	176
25. Veliko Gradiste	21°31′	44°45′	80
26. Vranje	21°55′	42°33′	432
27. Zajecar	22°17′	43°53′	144
28. Zlatibor	19°43′	43°44′	1028
29. Zrenjanin	20°21′	45°24′	80