Extracting physical homogeneous regions out of irrigation networks using fuzzy clustering method: a case study for the Ghazvin canal irrigation network

M. J. Monem and S. M. Hashemy

ABSTRACT

Improving the current operation and maintenance activities is one of the main steps in achieving higher performance of irrigation networks. Improving the irrigation network management, influenced by different spatial and temporal parameters, is confronted with special difficulties. One of the controversial issues often faced by decision-makers is how to cope with the spatial diversity of irrigation systems. Homogeneous area detection out of the irrigation networks could improve the current management of networks. The idea behind this research is to present a quantitative benchmark for exploring the homogeneous areas with similar physical attributes out of the network region. Five physical attributes, such as length, capacity, number of intakes, number of conveyance structures and the covered irrigated area for each canal reach, are used for spatial clustering. Two fuzzy clustering algorithms, namely FCM and GK, are applied to the Ghazvin irrigation network. Using a clustering validity index, *SC*, shows that the GK algorithm is the more appropriate tool for clustering of the considered dataset. According to the results the optimal number of clusters for the Ghazvin irrigation project is derived as nine clusters and the irrigated district is classified into nine homogeneous areas. Physical homogeneous regions provide a context for better and easier decision-making.

Key words | fuzzy clustering, irrigation network, regionalization

M. J. Monem (corresponding author)
S. M. Hashemy
Department of Hydraulic Infra-Structures,
Faculty of Agriculture,
University of Tarbiat Modares,
Aleahmad Ave.,
Chamran Crossing,
Tehran,
Iran
E-mail: *Monem_mj@modares.ac.ir*

NOTATION

X	dataset
N	number of data
c	number of clusters
c_{max}	maximum number of clusters
x_k	measured variables
v_i	cluster center
V	vector of cluster centers
V_c	cluster center of cth cluster
U	partition matrix
$\mu_{i,k}$	membership values of x_k to ith cluster
$D_{i,kA}$	squared inner-product distance norm
A	distance norm
J	objective function
l	number of iterations
A_i	adaptive distance norm

ABBREVIATIONS

GIN	Ghazvin Irrigation Network
O&M	Operation and Maintenance
FCM	fuzzy c-means
GK	Gustafson–Kessel
CE	classification entropy
S	separation index
XB	Xie & Beni
SC	partition index

doi: 10.2166/hydro.2010.058

INTRODUCTION

Irrigated agriculture contributes about 40% of the global food production. This comes mainly from about 287 million hectares of irrigated lands (FAO 2009). An FAO analysis of 93 developing countries indicates that within the 32 year period from 1998 to 2030, agricultural production should be increased by 81% in irrigation systems. Therefore, the major portion of additional food production should come from irrigated land, three-quarters of which is located in developing countries (Playan & Mateos 2006).

Many irrigation schemes in developing countries suffer from poor management, both in its technical and social dimensions. This often leads to unsustainable practices with decaying infrastructure and a reluctance of users to contribute to the maintenance activities, which causes poor performance of irrigation schemes (Fernandez *et al.* 2003).

Achieving the desirable performance of irrigation networks is impossibe without altering the guidelines or any research methods into workable procedures, techniques and tools to solve different managerial problems. Practical guidelines and feasible research methods are reference tools to assist policy-makers, planners, technical experts and farmers involved in irrigation management programs.

Beside external factors, there have been some internal spatial and temporal parameters which influence irrigation canal performance (Steiner & Walter 1993). Irrigation network performance is improved, provided that variant spatial and temporal factors are considered. One of the concerns of irrigation managers is how to cope with spatial diversity and temporal variability of parameters that strongly affect the operation and maintenance (O&M) activities (Sarwara *et al.* 2001). The spatial diversity is created by different reasons such as physical differences between canals and structures in different areas of the irrigation networks; operators with different knowledge and experiences, and various types of management (governmental, cooperative and private). Physical differences in irrigation networks are created by the following factors: capacity of canals and structures, type and number of structures, length of canals, distances to source of water, etc. Resource variations and changing O&M policies over time could be accounted as the reason for temporal variation in canal irrigation networks. According to the aforementioned reasons, an irrigation network assessment requires the total consideration of spatial and temporal variation of the irrigation district. The process of collecting and analyzing detailed data from all around the irrigation network is expensive and time-consuming.

Unfortunately, no quantitative approaches have been introduced to assess how O&M activities should be oriented with respect to spatial and temporal parameters. Many surveys carried out by the FAO show that O&M activities are not well mastered and consequently physical deterioration of the irrigation infrastructure and poor services have increased (Renault & Facon 2007). Therefore, better managerial approaches are required to facilitate O&M activities and to achieve higher performance of irrigation systems (Montazar *et al.* 2010). Regionalization of irrigation networks is one of the approaches which could be used to facilitate activities such as evaluation of the irrigation network, modernization, rehabilitation and, especially, for O&M activities.

On the other hand, many records are collected in regular inspections and annual surveys in irrigation districts. In many cases, these statistics are just used in the form of charts and tables in reports. Usually, comparing these statistics gives hints as to how to make decisions for O&M activities. However, the key point is that comparing data for extended systems like irrigation networks without applying powerful tools is beyond human abilities. Consequently, important decisions are often made based not on the information but rather on a decision-maker's intuition, simply because the decision-maker does not have the tools available to extract the valuable knowledge embedded in the data. Thus, using assured and capable approaches, like data mining approaches that can be applied to a wide variety of data, are indispensible for the regionalization of irrigation networks.

Malano & Gao used a fuzzy clustering technique in a specific period of time for the Goulburn and Shi-jin irrigation networks in Australia and China, respectively. Clustering is applied on temporal data, as a managerial tool, to determine the turning point of the O&M activities of the project during the considered period of time (Malano & Gao 1992). Temporal clustering helps managers and decision-makers to discover the turning points in the considered duration of time and, after that, to change or reform policies and management methods if this is necessary.

The purpose of this paper is to identify homogeneous parts of irrigation networks based on the physical features, using a fuzzy clustering tool. Considering the mutual relationship between O&M activities and the physical characteristics of irrigation networks, physical attributes of canal networks are considered in clustering (Alshaikh 1995). The clustering technique enables managers to limit the spatial boundaries of decision-making from large scale, irrigated areas to limited homogeneous regions. The major reason that clustering has attracted a great deal of attention in different research areas in recent years is due to the wide availability of huge amounts of data and the imminent need for turning such data into useful patterns, information and knowledge (Han & Kamber 2006).

According to the best knowledge of the authors, this is the first time that a fuzzy clustering approach is used to extract similar physical homogeneous regions out of an irrigation network.

METHODS

Fuzzy cluster analysis

The process of grouping a set of objects into classes with similar attributes is called clustering (Han & Kamber 2006). The main potential of clustering is to detect the underlying structure in the dataset. Cluster analysis, also called the unsupervised pattern recognition method, is very popular because of its ability to classify sets of unlabeled data (Alberto *et al.* 2001). Unlabeled data refers to situations where datasets exist without any prior information for their analysis (Koskela 2004). Two advantages of cluster analysis over manual grouping are in applying a specified objective function consistently in the clustering approach to form the groups, which avoids possible inconsistency due to human error, and grouping the dataset in a short time (Srinivasa & Duckstein 2004). Different applications of clustering analysis in water science research have been reported (Alberto *et al.* 2001; Koskela 2004; Doan *et al.* 2005; Liou & Lo 2005). Clustering algorithms can be classified into hard and fuzzy classes. In the hard clustering approach each object does or does not belong to a cluster, while in fuzzy clustering an object allows allocation to several clusters with different membership values. In many real situations, fuzzy clustering is better than hard clustering, because objects on the boundaries between several classes are not forced to fully belong to one of the classes. However, the objects are assigned membership degrees in the range of 0 to 1, indicating their partial memberships to different classes. In this paper two fuzzy clustering algorithms, named fuzzy c-means (FCM) and Gustafson–Kessel (GK) are used because of their wide and successful applications in several fields.

Fuzzy c-means (FCM) algorithm

The fuzzy c-means clustering algorithm is based on the minimization of an objective function which is introduced as Equation (1) (Bezdek 1981):

$$J(X;U,V) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{i,k})^m \|x_k - v_i\|_A^2 \tag{1}$$

where $X = \{x_k | k = 1,2,\ldots,N\}$, $U = [\mu_{i,k}]$, $X_k = [x_{k1}, x_{k2}, \ldots, x_{km}]^T$ and cluster centers, v_i, can be calculated by Equation (2):

$$v_i^{(l)} = \frac{\sum_{k=1}^{N}(\mu_{i,k}^{(l-1)})^m X_k}{\sum_{k=1}^{N}(\mu_{i,k}^{(l-1)})^m},\ 1 \le i \le c \tag{2}$$

Membership values of x_k to the ith cluster are computed by Equation (3):

$$\mu_{i,k}^{(l)} = \frac{1}{\sum_{j=1}^{c}\left(\frac{D_{i,kA}}{D_{j,kA}}\right)^{\frac{2}{m-1}}}. \tag{3}$$

The squared inner-product distance norm is defined as

$$D_{i,kA}^2 = \|x_k - v_i\|_A^2 = (x_k - v_i)^T A (x_k - v_i), \tag{4}$$

The FCM algorithm is applied using the standard Euclidean distance norm in which $A = I$. Thus, it can only detect clusters with the same shape and orientation. The most popular method for minimization of the c-means objective function is the simple Picard iteration through the loop defined by Equations (2) and (3), to obtain the cluster centers, which produces the minimal objective function for a fixed group number c in each iteration (Liou *et al.* 2003). This process is related to different group numbers (c) to find the optimal number of clusters.

The Gustafson–Kessel (GK) algorithm

Gustafson & Kessel employed an adaptive distance norm instead of a Euclidian distance norm in order to detect clusters of different geometrical shapes through a dataset (Gustafson & Kessel 1979). Each cluster has its own norm-inducing matrix A_i, which yields the following inner-product norm:

$$D^2_{i,kA} = (x_k - v_i)^T A_i (x_k - v_i), \quad 1 \le i \le c, \quad 1 \le k \le N. \tag{5}$$

The extra parameter used in the objective function of GK clustering is $A = (A_1, A_2, \ldots, A_c)$. The matrices A_i allow each cluster to adapt the distance norm to the local topological structure of the data. The objective function of the GK algorithm is defined by Equation (6) (Babuska *et al.* 2002):

$$J(X; U, V, A) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{i,k})^m D^2_{i,kA_i} \tag{6}$$

For the sake of simplicity, a synthetic dataset in R^2 and clustering results are presented in Figures 1 and 2. The dots represent the data points, the 'o' markers are the centers of clusters and the membership values are shown by curves. As could be seen in Figure 2(a), the FCM is enabled to extract the four clusters with spherical shapes while the GK, Figure 2(b), has identified four clusters with different shapes out of the dataset.

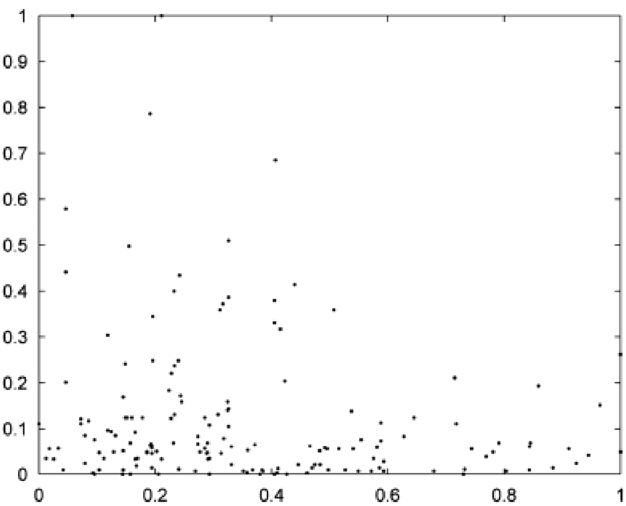

Figure 1 | Synthetic data in R^2.

Fuzzy clustering validity indices

One of the controversial issues in using clustering methods is choosing an optimal number of clusters (Weatherill & Burton 2008). In this paper four cluster validity indices named *PC*, *CE*, *S* and *XB* are applied for finding the optimal *c*. Cluster validity refers to the situation where a given fuzzy partition fits the data in a best possible fashion (Wu & Yang 2005). The clustering algorithm should be run through from 2 to c_{max} and after computing the validity indices, the optimal number of clusters is found. The value of c_{max} can be chosen according to the user's knowledge of the dataset; however, as this is not always possible, a rule of thumb that many investigators use is $c_{max} \le \sqrt{N}$ (Kim *et al.* 2004). The applied fuzzy clustering validity indices in this research are presented in Table 1. As is mentioned in the explanation column of this table, the first index on their maximum values and other indices on their minimum values determine the optimal number of clusters.

Fuzzy clustering approaches could be distinguished using the partition index (*SC*). The *SC* index (Equation (7)) is useful when comparison of two different fuzzy clustering approaches is done with equal numbers of clusters. A lower value of *SC* indicates a better fuzzy clustering method (Bensaid *et al.* 1996):

$$SC(c) = \sum_{i=1}^{c} \frac{\sum_{k=1}^{N} (\mu_{i,k})^m \|x_k - v_i\|^2}{\sum_{k=1}^{N} (\mu_{i,k}) \sum_{j=1}^{c} \|v_j - v_i\|^2} \tag{7}$$

Ghazvin irregation network

The case study in this research is the Ghazvin Irrigation Network (GIN) located in the northwest of Iran, depicted in Figure 3. This command area lies between 35°24′N to 36°48′N latitude and 48°45′ to 50°51′E longitude. The network covers an area of 60 000 ha and its water is supplied from the Taleghan Dam and 102 integrated water wells scattered over the network area. The network comprises 94 km of main canal, 220 km of secondary canals (12 branches), 330 km of lateral channels III (158 branches) and 550 km of subsidiary channels IV, with 30 000 branches and related outlets (Montazar & Riazi 2008).

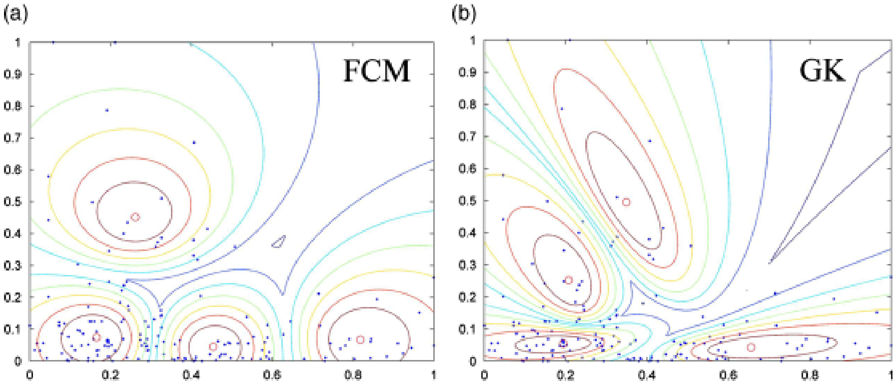

Figure 2 | The result of clustering by (a) FCM and (b) GK algorithm.

In this study, our concentration is focused on the 12 secondary canals. These canals are divided into 162 reaches having constant capacity and bounded between two control structures. For all of the reaches five physical attributes were considered. These attributes are: length, capacity, number of intakes, number of conveyance structures and the irrigated area covered. The variational ranges of these attributes are given in Table 2. The average delivery efficiency of this irrigation network from canal head to root zone is 30–36% (Montazer & Riazi 2008). This low rate of efficiency confirmed the lack of appropriate O&M activities in this irrigation network. To improve the O&M activities and reduction of the misapplication of money and water resources, the improvement of current O&M activities is needed.

RESULTS AND DISCUSSIONS

Finding the optimal number of clusters is done by determination of the cluster validity indices values for $c = 2$–12 numbers of clusters. The values of fuzzy clustering validity indices are organized for both FCM and GK clustering algorithms in Table 3. The optimum value for each index is given in bold in Table 3. According to the *S* and *XB* indices, the optimal number of clusters is nine clusters for both fuzzy clustering methods. The *CE* index has determined that 12 clusters is the optimal number of clusters, while the value of the index has marginal improvement compared to nine clusters. Therefore the general consensus of whole indices is choosing nine clusters as an optimal number of clusters for both methods.

Table 1 | The applied fuzzy clustering methods in this paper

Validity index	Formula	Explanation	Reference
Classification entropy (*CE*)	$CE(c) = -\frac{1}{N}\sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{i,k}\,ln(\mu_{i,k})$	The optimal number of cluster is at the maximum value of the CE	Bezdek (1981)
Separation index (*S*)	$S(c) = \frac{\sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{i,k})^2\Vert x_k - v_i\Vert^2}{N\min_{i,j}\Vert v_j - v_i\Vert^2}$	The optimal number of cluster is at the minimum value of S.	Bensaid *et al.* (1996)
Xie and Beni's index (*XB*)	$XB(c) = \frac{\sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{i,k})^m\Vert x_k - v_i\Vert^2}{N\min_{i,j}\Vert x_k - v_i\Vert^2}$	The minimum value of *XB* shows the optimal number of clusters.	Xie & Beni (1991)

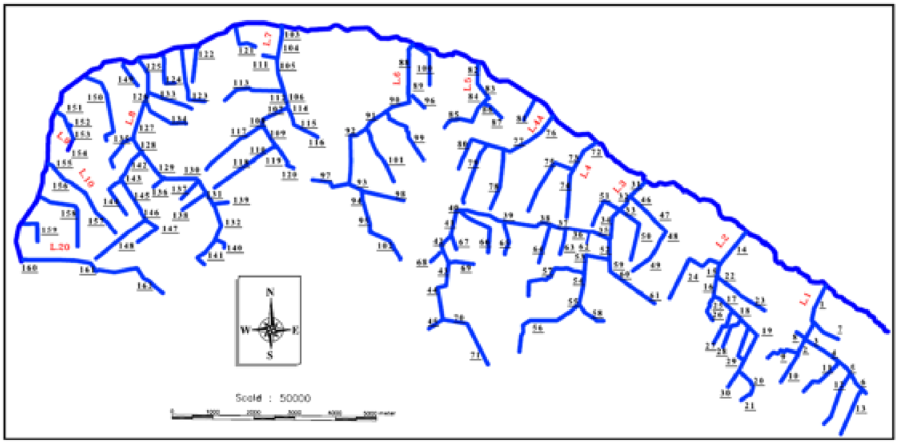

Figure 3 | Secondary canals of the Ghazvin irrigation network, Iran.

Table 2 | Attributes considered in the study

Attribute	Range
Length (m)	120–6190.9
Capacity (cm)	0.1–0.73
Number of intakes	1–13
Number of conveyance structures	0–56
Covered irrigated area (ha)	0.31–4912.84

By comparing the values of the *SC* index, it could be found that, except for the case of $c = 2$, the GK clustering method is a more suitable approach to clustering the dataset of the Ghazvin irrigation network. The non-spherical distribution of the dataset of this study and employing an adaptive distance norm, in order to detect clusters of different geometrical shapes in the dataset, could justify the better results of GK in comparison with FCM. Thus, the 162 canal reaches of the Ghazvin irrigation network are grouped into nine clusters and the result of the GK fuzzy clustering method is presented.

The membership value conditions in different clusters are given in Table 4. As can be identified, 37 reaches have membership values of more than 70%, 36 reaches have membership values between 50–70% and 20 reaches have membership values in the range of 40–50%. The rest of the reaches have membership values less than 40%. Therefore, these 93 reaches which have membership values more than 40% are considered in the analysis. This ranking of

Table 3 | The values of the fuzzy clustering validity indices

	FCM algorithm			GK algorithm			SC index	
Clusters	CE	S	XB	CE	S	XB	FCM-SC	GK-SC
2	0.4	0.0143	2.5886	0.5098	0.0273	2.507	2.3148	4.429
3	0.78	0.0219	1.5463	0.735	0.0125	1.4847	2.2583	1.615
4	0.98	0.0204	1.4778	1.022	0.0151	1.3021	2.0931	1.756
5	1.11	0.0142	1.2356	1.0943	0.0118	1.1903	1.5511	1.317
6	1.28	0.0166	0.9812	1.2231	0.0095	1	1.583	1.058
7	1.42	0.0159	0.9726	1.3418	0.0096	1.0093	1.6021	1.064
8	1.56	0.0161	0.8941	1.4393	0.0097	0.9188	1.6777	1.14
9	1.62	**0.014**	**0.8677**	1.4379	**0.0068**	**0.7747**	**1.4936**	**0.801**
10	1.76	0.0148	0.8797	1.4696	0.0074	0.9623	1.5038	0.825
11	1.78	0.0143	0.9373	1.4671	0.0078	0.9572	1.4998	0.82
12	**1.81**	0.0141	0.8738	**1.4699**	0.0079	0.955	1.5102	0.82

Table 4 | Canal reaches with membership values more than 40%

Membership values (%)	C1	C2	C3	C4	C5	C6	C7	C8	C9	Total
70–100	9	6	4	5	5	0	3	5	0	37
50–70	4	0	10	4	3	5	2	4	4	36
40–50	3	0	0	0	0	4	1	0	12	20
Total	16	6	14	9	8	9	6	9	16	93

membership values allows managers to take decisions about canal reaches by different ranges of certainty. The results indicate that the canal reaches are spread among all clusters. Canal reaches in the same cluster have similar physical attributes, so it provides the capability to make the same decision for these objects.

For the canal reaches with membership degrees lower than 40%, the knowledge and experiences of managers, authorities and operators accompanied by the clustering results could be utilized to make managerial decisions. Table 5 gives the range of physical attribute variation in all clusters. Similar features of each cluster are presented in Table 6. All of the five considered physical features contribute to the clustering of the dataset and there is no dominant feature in the clustering results. Cluster C6 has the most similar members and only the irrigated area attribute is not similar between canal reaches. After C6, the most similar canal reaches are placed in clusters C1, C7 and C9 which have members with three similar features.

The 93 canal reaches which have membership values more than 40% create the Ghazvin canal network regionalization map presented in Figure 4. Some interesting results are visible in the depicted regionalization map, which are discussed here.

The desired condition for managers is that the canal reaches in the same cluster are situated next to each other in one or more than one colonies. When several similar reaches are situated close together, the numbers of O&M inspections and subsequently the operational and maintenance expenses will be decreased. This desired state happens for three clusters, which are marked A1–A3 in Figure 4. Clusters 1, 6 and 7 create the physical homogeneous regions that spread in the desired O&M condition, in the form of colonies, throughout the irrigation network. These regions are located in the downstream, middle part and upstream of the network, respectively. This indicates that the managers of the network could easily arrange these

Table 6 | Similar features of every cluster

Conveyance structures	Intakes	Irrigated area	Capacity	Length	Cluster
Similar	Similar		Similar		c1
Similar		Similar			c2
	Similar			Similar	c3
		Similar			c4
				Similar	c5
Similar	Similar		Similar	Similar	c6
		Similar	Similar	Similar	c7
Similar	Similar				c8
Similar		Similar	Similar		c9

Table 5 | The variation range of physical attributes in every cluster

Conveyance structures		Intakes		Irrigated area (ha)		Capacity (m^3)		Length (m)			
Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Number of objects	Cluster
39	25	5	4	581.16	212.74	0.27	0.22	5728	1900	15	**C1**
12	9	4	1	491.84	404.62	3.71	1	6190	1060	9	**C2**
30	0	2	1	2357.6	94.421	2.7	0.1	3200	2500	6	**C3**
12	2	5	1	67.859	57.038	4.3	1.33	110	600	9	**C4**
56	29	11	5	944.09	290.07	1.63	0.3	4861	3128	8	**C5**
6	2	2	1	857.43	480.31	1.25	1.1	2129	1950	9	**C6**
34	7	13	4	39.539	35.445	0.18	0.1	6191	5500	12	**C7**
16	15	2	1	2736	118.93	7.35	2.4	2790	470	6	**C8**
12	7	4	1	57.949	54.354	0.99	0.74	2375	120	16	**C9**

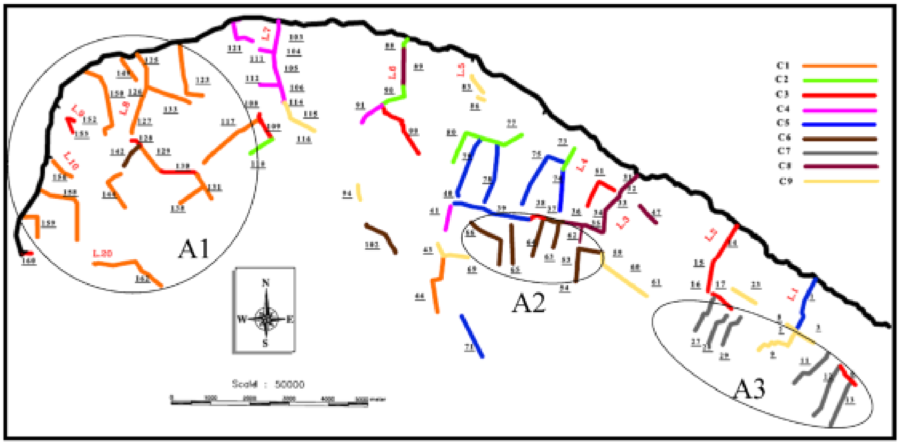

Figure 4 | The Ghazvin irrigation network regionalization map using fuzzy GK method.

O&M activities through local offices in each region, as can be seen in Figure 5, to handle similar facilities and to deal with similar situations. These local offices could provide services to other homogeneous colonies located next to them as well. For instance, the middle part office (office 2) could service clusters C5 and C8, and the downstream office (office 1) could serve cluster C4 as well.

CONCLUSIONS

According to the best knowledge of the authors, this is the first time that a fuzzy clustering method has been used as a qualitative benchmark for the regionalization of irrigation networks. The FCM and GK fuzzy clustering algorithms are employed to group the irrigation canal reaches by using five physical attributes. According to the results of fuzzy clustering validity indices, the canal network district is regionalized into nine clusters using the GK fuzzy clustering algorithm. Every cluster is representative of a physical homogeneous region of canals with similar physical attributes, which could help managers to prioritize the existing facilities for managing the O&M activities. Clustering reduces the dimension of assessments from a large extended irrigation district to a limited number of homogeneous regions.

On the condition that the experiences of operators and managers are the pillars of decision-making for ongoing O&M activities, this approach may provide complementary clues for considering the physical features of canal reaches in making decisions. The results are practical to some extent,

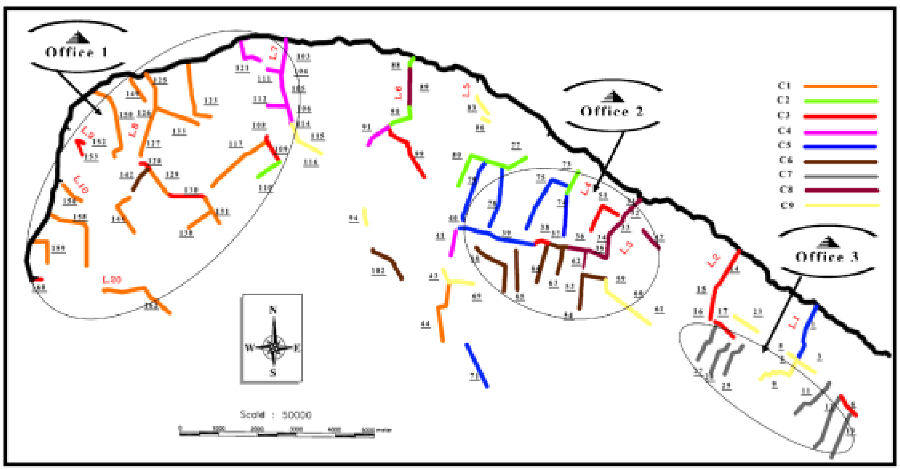

Figure 5 | The district of each O&M local offices.

especially in some cases where similar objects are located close together in the form of a colony.

It is now well understood that projects like modernization and rehabilitation should combine hardware, software and human-ware. Since human-ware is a key factor for success in these projects, other parts should prepare suitable circumstances to improve the ability of this factor.

The clustering technique alone, like other data-mining approaches, is generally not sufficient to solve the problem. The clustering results are useful, if human expertise is then applied for the interpretation of the results.

ACKNOWLEDGMENTS

The authors wish to thank the Iranian Water Resource Management Company and the Ghazvin Irrigation Network Authorities for their corporation in data gathering from the Ghazvin Irrigation Network, Iran. Also special appreciation is due to Dr. Feil Balázs for his indirect valuable help in learning data-mining methods by young interested students.

REFERENCES

Alberto, W. D., Pilar, D. M., Valeria, A., Fabiani, P. S., Cecilia H. A. & Angeles, B. M. D. L. 2001 Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. a case study: Suquia river basin. *Wat. Res.* **35**(12), 2881–2894.

Alshaikh, A. 1995 Optimal design of irrigation canal network under uncertainty using response surface method. *Wat. Int.* **20**(3), 155–162.

Babuska, R., Van der Veen, P. J. & Kaymak, U. 2002 Improved covariance estimation for Gustafson–Kessel clustering. In: *Proc. 2002 IEEE International Conference on Fuzzy Systems, Hawaii* vol 2, pp. 1081–1085. IEEE Publishing.

Bensaid, A. M., Hall, L. O., Bezdek, J. C., Clarke, L. P., Silbiger, M. L., Arrington, J. A. & Murtagh, R. F. 1996 Validity-guided (re)clustering with applications to image segmentation. *IEEE Trans. Fuzzy Syst.* **4**(2), 112–123.

Bezdek, J. C. 1981 *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum, New York.

Doan, C. D., Liong, S. Y. & Karunasinghe, D. S. K. 2005 Derivation of effective and efficient data set with subtractive clustering method and genetic algorithm. *J. Hydroinf.* **7**(4), 219–233.

FAO 2009 *FAO Statistics*. Available at: http://faostat.fao.org/site/377/DesktopDefault.aspx?PageID=377#ancor

Fernandez, S., Garces-Restrepo, C. & Vidal, A. 2003 Improving the water service in irrigation: a series of case studies on irrigation modernization. In: *Proc. 2nd International Conference on Irrigation and Drainage Phoenix, AZ*. USCID, Phoenix, AZ pp. 71–86.

Gustafson, D. E. & Kessel, W. C. 1979 Fuzzy clustering with fuzzy covariance matrix. In: *Proc. IEEE CDC*, San Diego, CA, pp. 761–766. IEEE, San Diego, CA.

Han, J. & Kamber, M. 2006 *Data Mining, Concepts and Techniques*. Morgan Kaufman, San Francisco.

Kim, D. W., Lee, K. H. & Lee, D. 2004 On cluster validity index for estimation of the optimal number of fuzzy clusters. *J. Pattern Recog, Soc.* **37**(10), 2009–2025.

Koskela, J. 2004 *Pattern Recognition in Water Resources Management: A Literature Review and An Application to Long-term Inflow Forecasting*. Master of Science thesis, Department of Civil and Environmental Engineering, Helsinki University of Technology.

Liou, S. M., Lo, S. L. & Hu, C. H. 2003 Application of two-stage fuzzy set theory to river quality evaluation in Taiwan. *Wat. Res.* **37**(6), 1406–1416.

Liou, Y. T. & Lo, S. L. 2005 A fuzzy index model for trophic status evaluation of reservoir waters. *Wat. Res.* **39**(7), 1415–1423.

Malano, H. M. & Gao, G. 1992 Ranking and classification of irrigation system performance using fuzzy set theory: case studies in Australia and China. *Irrig. Drainage Syst.* **6**(2), 129–148.

Montazer, A. & Riazi, H. 2008 Optimization of water allocation in Qazvin irrigation command area. *J. Appl. Irrig. Sci.* **43**(21), 29–134.

Montazar, A., Riazi, H. & Behbahani, S. M. 2010 Conjunctive water use planning in an irrigation command area. *J. Water Resour. Manage.* **24**(3), 577–596.

Playan, E. & Mateos, L. 2006 Modernization and optimization of irrigation systems to increase water productivity. *J. Agric. Wat. Mmgmnt.* **80**(1–3), 100–116.

Renault, D. & Facon, T. 2007 *Modernizing Irrigation Management – The MASSCOTE Approach*. FAO irrigation and drainage paper. Available at: http://www.fao.org/docrep/010/a1114e/a1114e00.htm.

Sarwara, A., Bastiaanssenb, W. G. M. & Feddesc, R. A. 2001 Irrigation water distribution and long-term effects on crop and environment. *J. Agric. Wat. Mngmnt.* **50**(2), 125–140.

Srinivasa, R. K. & Duckstein, L. 2004 Integrated application of cluster and multi-criterion analysis for ranking water resources planning strategies: a case study in Spain. *J. Hydroinf.* **6**, 295–307.

Steiner, R. A. & Walter, M. F. 1993 The effect of allocation schedules on the performance of irrigation systems with different levels of spatial diversity and temporal variability. *J. Agric. Wat. Mngmnt.* **23**(3), 213–224.

Weatherill, G. & Burton, P. W. 2008 Delineation of shallow seismic source zones using K-means cluster analysis with application to the Aegean region. *Geophys. J.* **176**(2), 565–588.

Wu, K. L. & Yang, M. S. 2005 A cluster validity index for fuzzy clustering. *Pattern Recog. Lett.* **26**(9), 1275–1291.

Xie, X. L. & Beni, G. A. 1991 Validity measure for fuzzy clustering. *IEEE Trans.* **13**(8), 841–846.

First received 4 April 2010; accepted in revised form 19 July 2010. Available online 13 December 2010