Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy *c*-means clustering

Cüneyt Güler

Geological Engineering Department, Çiftlikköy Kampüsü, Mersin Üniversitesi, Mersin, Turkey

Geoffrey D. Thyne

Department of Geology and Geological Engineering, Colorado School of Mines, Golden, Colorado, USA

Received 26 April 2004; revised 26 July 2004; accepted 21 September 2004; published 7 December 2004.

[1] In this paper, classification of a large hydrochemical data set (more than 600 water samples and 11 hydrochemical variables) from southeastern California by fuzzy *c*-means (FCM) and hierarchical cluster analysis (HCA) clustering techniques is performed and its application to hydrochemical facies delineation is discussed. Results from both FCM and HCA clustering produced cluster centers (prototypes) that can be used to identify the physical and chemical processes creating the variations in the water chemistries. There are several advantages to FCM, and it is concluded that FCM, as an exploratory data analysis technique, is potentially useful in establishing hydrochemical facies distribution and may provide a better tool than HCA for clustering large data sets when overlapping or continuous clusters exist. *INDEX TERMS:* 1806 Hydrology: Chemistry of fresh water; 1829 Hydrology: Groundwater hydrology; 1831 Hydrology: Groundwater quality; 1871 Hydrology: Surface water quality; *KEYWORDS:* clustering techniques, fuzzy logic, fuzzy *c*-means, hydrochemical facies, hydrochemistry inverse modeling

Citation: Güler, C., and G. D. Thyne (2004), Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy *c*-means clustering, *Water Resour. Res.*, *40*, W12503, doi:10.1029/2004WR003299.

1. Introduction

[2] Clustering techniques fall into two main categories. They are considered to be "hard or crisp" if an object (i.e., a water sample) belongs exclusively to a single cluster or "soft or fuzzy" if an object belongs to all clusters in varying degrees of membership. Whether they are hard or fuzzy, the main purpose of these techniques is to partition a data matrix with n samples and p variables into homogeneous c number of subsets by grouping closely related samples into tight clusters. In the resulting partition, samples within the same cluster are characterized by a high degree of similarity, while samples belonging to different clusters are characterized by a high degree of dissimilarity. These homogeneous clusters then can be represented by the reduced $c \times p$ dimensional matrix, which is composed of c number of typical samples representing each cluster and their respective parameter (p) values.

[3] In hydrochemical studies, classification serves the purpose of isolating a group of representative clusters (also known as water groups or hydrochemical facies) that reflect the processes generating the natural variability found in hydrochemical parameters. These representative clusters, which help to define the major chemical trends, can then be used to reconstruct the underlying processes or for exemplification [e.g., *Suk and Lee*, 1999; *Barbieri et al.*, 2001; *Güler et al.*, 2002; *Güler and Thyne*, 2004]. However, the underlying processes do not always produce discrete outcomes and are hard to characterize in complex hydrologic and hydrochemical systems. Often the physical and chemical properties of the natural system vary continuously (in both space and time), rather than abruptly. Because of this continuity, statistical clusters may not be well separated and instead may form a sequence of overlapping clusters [*Güler et al.*, 2002]. This limits the applicability of the hard (crisp) clustering techniques (e.g., hard *k*-means) for hydrochemical data classification, because the samples in the overlapping area show transitional character in chemistry; thus they are prone to misclassification. Therefore methods related to "fuzzy logic" may be best suited for that purpose since these methods can provide more information about the degree of membership of a water sample in a cluster than rigid classification techniques.

[4] Fuzzy logic, developed by *Zadeh* [1965], was designed to supplement the interpretation of linguistic or measured uncertainties for real-world random phenomena [*Chang et al.*, 2001]. Traditional Aristotelian logic (binary logic) imposes sharp boundaries; however, fuzzy logic has no sharp boundaries [*Fang and Chen*, 1990]. Fuzzy logic is basically a multivalued logic that allows intermediate values to be defined between conventional evaluations like yes/no, true/false, black/white, 0/1, etc. [*Zadeh*, 1965; *McNeill and Freiberger*, 1993]. Both fuzziness and probability describe uncertainty numerically; however, probability treats yes/no occurrences and is inherently a statistical method. Fuzziness deals with degrees and is a nonstatistical method.

[5] Although fuzzy clustering techniques are used in a wide range of clustering problems in many technical fields, they are rarely used in hydrochemical or hydrogeological studies. One of the most commonly used methods for solving fuzzy clustering problems is fuzzy *c*-means (also

0043-1397/04/2004WR003299$09.00

known as fuzzy *k*-means) [*Bezdek*, 1981]. The fuzzy *c*-means (FCM) algorithm has been successfully applied in various disciplines such as biology [*Zhang et al.*, 1995], chemistry [*Sun and Danzer*, 1996; *Linusson et al.*, 1998], climate [*McBratney and Moore*, 1985], food processing [*Hu et al.*, 1998], geology [*Burrough et al.*, 2000; *Rantitsch*, 2000], hydrochemistry [*Barbieri et al.*, 2001; *Güler et al.*, 2002], image analysis [*Bezdek et al.*, 1993; *Ahmed et al.*, 2002], and soil science [*McBratney and deGruijter*, 1992; *Odeh et al.*, 1992]. FCM has been modified and generalized in several ways in order to detect different cluster shapes and reduce the execution time of the algorithm for real-time classification purposes (i.e., image classification or grading fish products). As a result of this, numerous variations of the FCM algorithm are available in the recent literature such as conditional fuzzy *c*-means (CFCM) [*Pedrycz*, 1996], fuzzy symbolic *c*-means (FSCM) [*El-Sonbaty and Ismail*, 1998], fuzzy *c*-medians (FCMED) [*Kersten*, 1999], hierarchical unsupervised fuzzy clustering (HUFC) [*Geva*, 1999], fuzzy *k*-means with extragrades (FKME) [*Minasny and McBratney*, 2002], fuzzy *J*-means (FJM) [*Belacel et al.*, 2002], alternative fuzzy *c*-means (AFCM) [*Wu and Yang*, 2002], bias-corrected fuzzy *c*-means (BCFCM) [*Ahmed et al.*, 2002], and adaptive rough fuzzy leader clustering (ARFLC) [*Asharaf and Murty*, 2003]. All these fuzzy clustering algorithms rely on elements of *Zadeh*'s [1965] fuzzy set theory and many of them are based on the FCM algorithm, originally proposed by *Dunn* [1974] and extended by *Bezdek* [1981].

[6] In this study, the FCM clustering technique was applied to a large hydrochemical data set from the Indian Wells–Owens Valley area of southeastern California to delineate clusters of water samples with similar characteristics (hydrochemical facies) and to identify representative hydrochemical parameter values for each cluster. Then, the reduced data matrix ($c \times p$ matrix) was used to formulate (inverse) models to explain the hydrochemical variability observed in the water samples.

2. Study Area

[7] The study area is located in southeastern California and covers approximately 30,000 km^2 stretching from Owens Valley (OV) in the north to Indian Wells Valley (IWV) in the south (Figure 1). Both OV and IWV are bordered on the west by the Sierra Nevada with peaks rising from 1830 to 4400 m and on the east by several ranges with elevations ranging from 1500 to 2750 m. Topography of the valley floors is relatively flat and slopes to the south with elevations decreasing from 1100 m in the OV to nearly 663 m in the vicinity of IWV [*Güler*, 2002]. In a classic basin and range groundwater system [*Maxey*, 1968], water flows from high-elevation recharge areas in the mountains to topographically lower discharge areas in the valleys. Central portions of the valleys are generally occupied by playas, known in different localities as ''salt lakes,'' ''soda lakes,'' ''alkali marshes,'' ''dry lakes'' or ''borax lakes'' where the groundwater discharges by evapotranspiration. The most important and probably well known of these playas are Owens Lake, China Lake, and Searles Lake, which are located in Owens Valley, Indian Wells Valley, and Searles Valley, respectively (Figure 1). Because of the modern arid climate, surface water is scarce in the area and mostly occurs in manmade reservoirs (e.g., Haiwee Reservoir and Lake Isabella). The lithology of the area consists of a wide variety of rocks and unconsolidated deposits, which can be divided into five hydrogeologic units including igneous and metamorphic rocks, volcanic rocks, carbonate rocks, basin-fill deposits, and playa deposits. Detailed descriptions of these units and hydrogeology of the area are given by *Güler* [2002] and *Güler and Thyne* [2004].

3. Methodology

[8] The basic methodology is to begin with statistical clustering of the hydrochemical data followed by examining the spatial distribution of the clusters for spatial coherence and coincidence with major hydrologic and geological features. This process tests the choice of clustering methods and can be iterated to refine the statistical clusters so they are as geologically meaningful as possible. Then, the hydrochemical validity of the clusters is tested with inverse geochemical modeling, which uses mass balances to determine if the changes between the chemical compositions of the clusters can be explained as a result of water-rock interactions with the local mineralogy.

3.1. Database and Data Treatment

[9] Throughout the last 80 years, extensive hydrochemical data has been collected in the Indian Wells–Owens Valley area. The database used in this study is a subset of these data. Database construction procedures are discussed in detail by *Güler* [2002] and *Güler et al.* [2002], where the sources of the data are also presented. The entire database consists of chemical analyses of 152 spring, 153 surface, and 1063 well (groundwater) samples, including temporal series (samples collected over a period of time at the same location). Of the 39 hydrochemical variables (consisting of major ions, minor ions, trace elements, and isotope data) in the compiled database, the variables specific conductance (SC), pH, Ca, Mg, Na, K, Cl, SO_4, HCO_3, SiO_2, and F occur most often and were utilized in the FCM classification. These variables were also considered as being of primary significance in terms of delineating hydrochemical facies distribution. The remaining 28 variables are not used because they contain a high percentage (60–100%) of censored values (i.e., concentration values are reported as less than). In the case of multiple samples from the same location, the more recent and/or the more complete sample data were included in the FCM analysis. Analysis of temporal series, as well as statistical analysis of the entire database, suggested that relatively little change occurred in the water quality of samples with time [*Güler*, 2002; *Güler et al.*, 2002]. This indicates that spatial variability is the most important source of variation in the data, rather than the temporal factor. Therefore removal of duplicate (in some cases multiple) water samples from the same location is warranted. This reduced the total number of samples from 1368 to 691.

[10] Charge balance errors (CBE) are less than or equal to ±10% for the database, which is acceptable for the purpose of this study. Descriptive statistics of the data set have also shown that the values of variables varied by several orders of magnitude in the data set [*Güler et al.*, 2002]. In the FCM clustering, the results are strongly influenced by those variables with large variances, and therefore, a log-

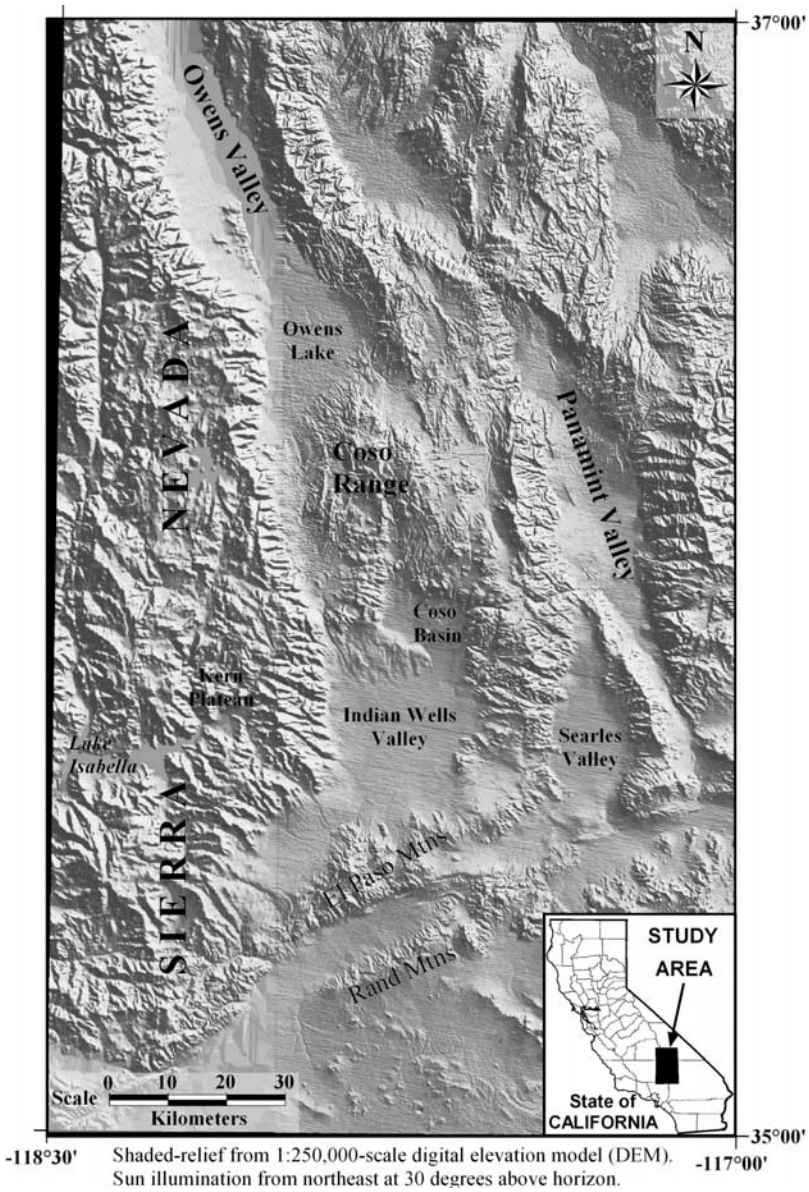

Figure 1. Location and important physiographic features of the Indian Wells–Owens Valley area, southeastern California.

transformed (base 10) and standardized data matrix was used as input data for the FCM clustering. Standardization scales the data to a range of approximately −3 to +3 standard deviations (σ), centered about a mean (μ) of zero, giving each variable equal weight in the analyses. The standardization procedure also eliminates the problem of comparison of variables measured in different units (e.g., specific conductance (μS cm^{-1}) and calcium (mg L^{-1})).

3.2. The Fuzzy *c*-Means Algorithm

[11] The following notation will be used throughout this paper. FCM clustering algorithm [*Bezdek*, 1981] is a multivariate data analysis technique and partitions a data set, $\boldsymbol{X} = \{x_1, \ldots, x_n\} \subset \Re^p$ (p-dimensional Euclidean space), into $c \in \{2, \ldots, n-1\}$ overlapping or fuzzy clusters, which are identified by their cluster centers (or prototypes), $\boldsymbol{v_i}$ ($i = 1, \ldots, c$). The partitioning of data into fuzzy clusters is achieved by minimizing the objective function:

$$J_{FCM}(M, C) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^{m} \parallel x_k - v_i \parallel^2 \qquad (1)$$

using an iterative procedure. In equation (1), $\boldsymbol{M}$ is the membership matrix, $\boldsymbol{C}$ is the cluster centers matrix, c is the number of clusters (classes or groups), n is the number of data points, and $\boldsymbol{u_{ik}}$ is the degree of membership of sample $\boldsymbol{k}$ in cluster $\boldsymbol{i}$. If the Euclidean distance (straight line distance between two points in p-dimensional space defined by p variables) between datum $\boldsymbol{x_k}$ and cluster center $\boldsymbol{v_i}$ is high, J_{FCM} is minimized. If the distance is small, the membership value approaches unity [*Höppner*, 2002]. The parameter $m \in (1, \infty)$ is a weighting exponent (also known as

fuzzification parameter) that controls the degree of the fuzziness of the resulting classification, which is the degree of overlap between clusters. With the minimum meaningful value of $m = 1$, the solution is a hard partition; that is, the result obtained is hard (or crisp). As m approaches infinity (∞) the solution approaches its highest degree of fuzziness [*Bezdek*, 1981]. The choice of $m = 2$ is widely accepted as a good choice of fuzzification parameter [*Hathaway and Bezdek*, 2001]. The matrix $\boldsymbol{M}$ is constrained to contain elements in the range [0,1] such that

$$\sum_{i=1}^{c} u_{ik} = 1, \qquad 1 \le k \le n \tag{2}$$

The second constraint on $\boldsymbol{M}$ matrix is that the coefficients for each cluster center must sum to less than the number of elements or

$$\sum_{k=1}^{n} u_{ik} < n, \qquad 1 \le i \le c \tag{3}$$

The objective function, J_{FCM}, is minimized by a two-step iteration. First, the $\boldsymbol{C}$ matrix is initialized with random values, and then the $\boldsymbol{M}$ matrix is estimated from the data set, $\boldsymbol{X}$, $m > 1$, and $\boldsymbol{C}$ where

$$u_{ik} = \left(\sum_{j=1}^{c} \left(\| x_k - v_i \| \,/\, \| x_k - v_j \| \right)^{2/(m-1)} \right)^{-1} \tag{4}$$

Then, the cluster centers (prototypes) are computed using the formula

$$v_i = \left(\sum_{k=1}^{n} \left(u_{ik}^{m} \right) X_k \right) \Bigg/ \left(\sum_{k=1}^{n} \left(u_{ik}^{m} \right) \right), \qquad 1 \le i \le c \tag{5}$$

[12] In FCM results, samples may not be a 100% member of a cluster or group; instead, the membership of samples is graded (partitioned) between groups. For example, a water sample may be mostly a member of a certain group, but it may be also a partial member of other groups. The assignment certainty of a sample to a specific cluster is described by its membership value (u) ranging between 0 and 1. The greater the certainty of a sample belonging to a cluster, the closer is its membership degree to 1. The membership values of each sample sum to 1. A more detailed discussion of FCM and examples is given by *Bezdek* [1981] and *Bezdek et al.* [1984, 1999]. The FCM algorithm for partitioning the water samples into representative clusters or hydrochemical facies can be summarized in the following steps:

[13] 1. Choose a value for the fuzzification parameter, m, with $m > 1$.

[14] 2. Choose a value for the stopping criterion, ε (e.g., $\varepsilon = 0.0001$ gives reasonable convergence).

[15] 3. Choose a distance measure in the variable-space (e.g., Euclidean distance).

[16] 4. Choose the number of classes or groups, c, with $c \in \{2,\ldots, n-1\}$.

[17] 5. Initialize $\boldsymbol{M} = \boldsymbol{M}^{(0)}$, e.g., with random memberships or with memberships from a hard k-means partition.

[18] 6. At iteration $it = 1, 2, 3$, recalculate $\boldsymbol{C} = \boldsymbol{C}^{(it)}$ using equation (5) and $\boldsymbol{M}^{(it-1)}$.

[19] 7. Recalculate $\boldsymbol{M} = \boldsymbol{M}^{(it)}$ using equation (4) and $\boldsymbol{C}^{(it)}$.

[20] 8. Compare $\boldsymbol{M}^{(it)}$ to $\boldsymbol{M}^{(it-1)}$ in a convenient matrix norm. If $\| \boldsymbol{M}^{(it)} - \boldsymbol{M}^{(it-1)} \| < \varepsilon$, then stop; otherwise return to step 6.

3.3. Cluster Validity Problem

[21] The fuzzy c-means algorithm partitions a data set into a predefined c number of clusters, whether or not the data set actually contains c clusters. For that reason, there is a need to establish a criterion for determination of the optimal number of clusters (also known as natural clusters) in the data, the so-called cluster validity problem [*Geva*, 1999; *Setnes*, 2000]. This problem is not unique to fuzzy clustering techniques but is shared by all the clustering methods. Cluster validity is a difficult problem that is crucial for the practical application of fuzzy clustering techniques [*Bezdek*, 1975]. The criteria used for discriminating between various cluster results are known as validity or performance measures [*Hammah and Curran*, 1999]. While some researchers have avoided the problem altogether by choosing the number of clusters intuitively or using their experience and familiarity with a particular data set [e.g., *McBratney and deGruijter*, 1992], others have used results from a hard clustering technique (e.g., hierarchical cluster analysis) to choose the optimal number of clusters [e.g., *Güler et al.*, 2002]. Obviously, employing such techniques to choose optimal number of clusters could bias the fuzzy clustering results.

[22] Several formal functions have been proposed for cluster validation that assess the goodness of a given partition considering criteria like the compactness of the clusters and the distance between the clusters [*Gunderson*, 1978; *Trauwert et al.*, 1991; *Bezdek and Pal*, 1998; *Kothari and Pitts*, 1999]. Among these, the most frequently used ones are partition coefficient [*Bezdek*, 1974], partition entropy [*Bezdek*, 1975], proportion exponent [*Windham*, 1981], Fukuyama and Sugeno index [*Fukuyama and Sugeno*, 1989], compactness separability [*Xie and Beni*, 1991], compose within and between scattering (CWB) [*Ramze et al.*, 1998], fuzziness performance index (FPI), and normalized classification entropy (NCE) [*Roubens*, 1982]. A drawback of these methods is the need for repetitive clustering of the data using different number of clusters to reach an optimal solution [*Setnes*, 2000].

[23] In this study, the optimal number of clusters was chosen by using the validation functions introduced by *Roubens* [1982]. FPI estimates the degree of fuzziness generated by a specified number of classes and given as

$$FPI = 1 - \frac{cF - 1}{c - 1}, \qquad \text{where } \mathrm{F} = \frac{1}{n} \sum_{i=1}^{c} \sum_{\mathrm{k}=1}^{\mathrm{n}} (\mathrm{u}_{ik})^2 \tag{6}$$

NCE estimates the degree of disorganization created by a specified number of classes and given as

$$NCE = \frac{H}{\log c}, \qquad \text{where } \mathrm{H} = \frac{1}{n} \sum_{i=1}^{c} \sum_{\mathrm{k}=1}^{\mathrm{n}} \mathrm{u}_{ik} \times \log(\mathrm{u}_{ik}) \tag{7}$$

The optimum number of classes is established on the basis of minimizing these two measures. A detailed explanation

and examples of these methods is given by *McBratney and Moore* [1985] and *Odeh et al.* [1992]. In this study, the optimal number of clusters determined by these two measures is compared to the results from a previous study [*Güler and Thyne*, 2004], which employed hierarchical cluster analysis (HCA) to cluster a data set (with 579 water samples and 11 variables) from the same area. In this study, a larger data set (691 samples with the same 11 variables) was used for fuzzy classification purposes. Therefore FCM classification results from the present study are expected to produce improved resolution compared to the previous statistical clustering analysis.

4. Results

4.1. Fuzzy Classification of Water Samples Into Hydrochemical Facies

[24] Fuzzy classification of a large hydrochemical data set (from the Indian Wells–Owens Valley area, southeastern California) into distinct hydrochemical groups was performed using the program FuzME (version 3) [*Minasny and McBratney*, 2002]. In this method of clustering, not only the number of clusters (c) is preselected but also the choice of distance measure, stopping criterion (ε), and fuzzification parameter (m) value need to be defined at the start of the analysis. For our study, Euclidean distance was chosen as the distance measure, which gives an equal weight to all variables since they are standardized prior to analysis. A value of 1×10^{-3} was used for the stopping criterion (ε). The FCM clustering algorithm stops as soon as the absolute value of differences of all pairs of elements in a successive pair of $\boldsymbol{M}$ matrices differs by less than 1×10^{-3}. However, there still remains the problem of selecting the optimal number of groups and fuzzification parameter value for a meaningful analysis result. For that reason, repetitive clustering of the data set with varying number of clusters and fuzzification parameter values was required. The limitation of this approach is the high computational effort and time required. For this study, the lognormalized and standardized data set is partitioned by means of FCM clustering into a range of c (between 3 and 10) and using a series of m values (i.e., 1.10, 1.15,. . ., 3.00) at an increment of 0.05. Using this procedure, a total of 312 runs were made and the data set was clustered using different combinations of c and m values.

[25] For the selection of the optimal number of groups in the data set, previously mentioned cluster validation functions (FPI and NCE) have been used. FuzME program also calculates the FPI and NCE values for each run and prints a separate summary file for the analysis. Figure 2 shows the plot of FPI and NCE values versus the number of groups for selected m values. A total of 39 plots resulted from the 312 FCM clustering runs, and only nine of them are given in Figure 2 for visualization purposes. Most of the plots showed similar patterns, and they are used in the selection of the optimal number of groups and fuzzification parameter. As we can see in Figure 2, in all plots both FPI and NCE are minimal at either $c = 5$ or $c = 4$. This means that the optimal number of c in the data set is either 5 or 4. Number of classes was chosen to be 5, because it has resulted in a hydrogeologically meaningful hydrochemical facies distribution (see below for further discussion). Choosing five groups also enables comparison of the results with the previously defined hydrochemical facies distribution using hierarchical cluster analysis (HCA) [*Güler and Thyne*, 2004]. From the trial runs it appeared that $m = 1.3$ resulted in membership values that were neither too fuzzy nor too hard for spatial mapping purposes, so this value was chosen. FPI values in Figure 2 also show that results become extremely fuzzy when m approaches to 1.7 (FPI = 1 is fuzzy and FPI = 0 nonfuzzy; in this study FPI = 0.25 was obtained for $c = 5$ and $m = 1.3$).

[26] The FCM analysis reduced the original 691×11 (row $\times$ column) data matrix to a 691×5 matrix of memberships $\boldsymbol{M}$ (Table 1) and to a 5×11 matrix of cluster centers (cluster prototypes) $\boldsymbol{C}$ (Table 2). Table 1 shows a portion of the membership matrix for the first 20 samples (the complete table is too large to present). The matrix of clusters centers in Table 2 shows the representative parameter values (for 11 physicochemical parameters) of the five cluster centers (or five water groups: groups 1, 2, 3, 4, and 5) derived by FCM objective function. These values reveal some trends between the clusters. TDS (sum of anions and cations) seems to be a major distinguishing factor with average concentrations increasing following the order: group-1 ($\sim$77 mg L^{-1}), group-2 ($\sim$318 mg L^{-1}), group-3 ($\sim$882 mg L^{-1}), group-4 ($\sim$8584 mg L^{-1}), and group-5 ($\sim$17,716 mg L^{-1}).

4.2. Spatial Distribution of the Hydrochemical Facies

[27] The FCM clustering algorithm is an effective method for exploring the structure of complex data sets where grouping of overlapping and vague elements is necessary. However, the numerical results from the FCM have little meaning unless the underlying physical or chemical processes can be related to naturally occurring processes and the spatial distribution of the clusters defined by FCM is meaningful. For that reason the membership values (u) for water samples in each water group were treated as a regionalized variable and separately interpolated on a regular grid by using the point kriging estimation method to define the hydrochemical facies distribution in the study area. Kriged cluster membership values for each water group are displayed as color-coded cluster value contours (see Figure 3a). To achieve a better cluster separation, only membership values above 0.5 were plotted on the map. In other words, $u = 0.5$ can be considered as a threshold membership value above which a given water sample is considered as belonging to the cluster.

4.3. Comparison of FCM With HCA

[28] Figures 3a and 3b show the results of the two clustering techniques on map view. In Figure 3b the HCA (hard) clusters overlain on the shaded-relief map (digital elevation model; DEM) for the study area. However, In Figure 3a the results of the FCM analysis are not overlain on the DEM to better highlight the differences. As we can see in Figure 3a, spatial overlay of the membership values for each cluster shows distinctive spatial patterns. The five groups are separated geographically, as well as physiographically with good correspondence between spatial locations and the FCM clusters. The samples that belong to the same group are located in close proximity to one another, suggesting the same processes and/or flow paths for those groups of samples. Both techniques (FCM and HCA) provide rapid and efficient methodology to cluster large

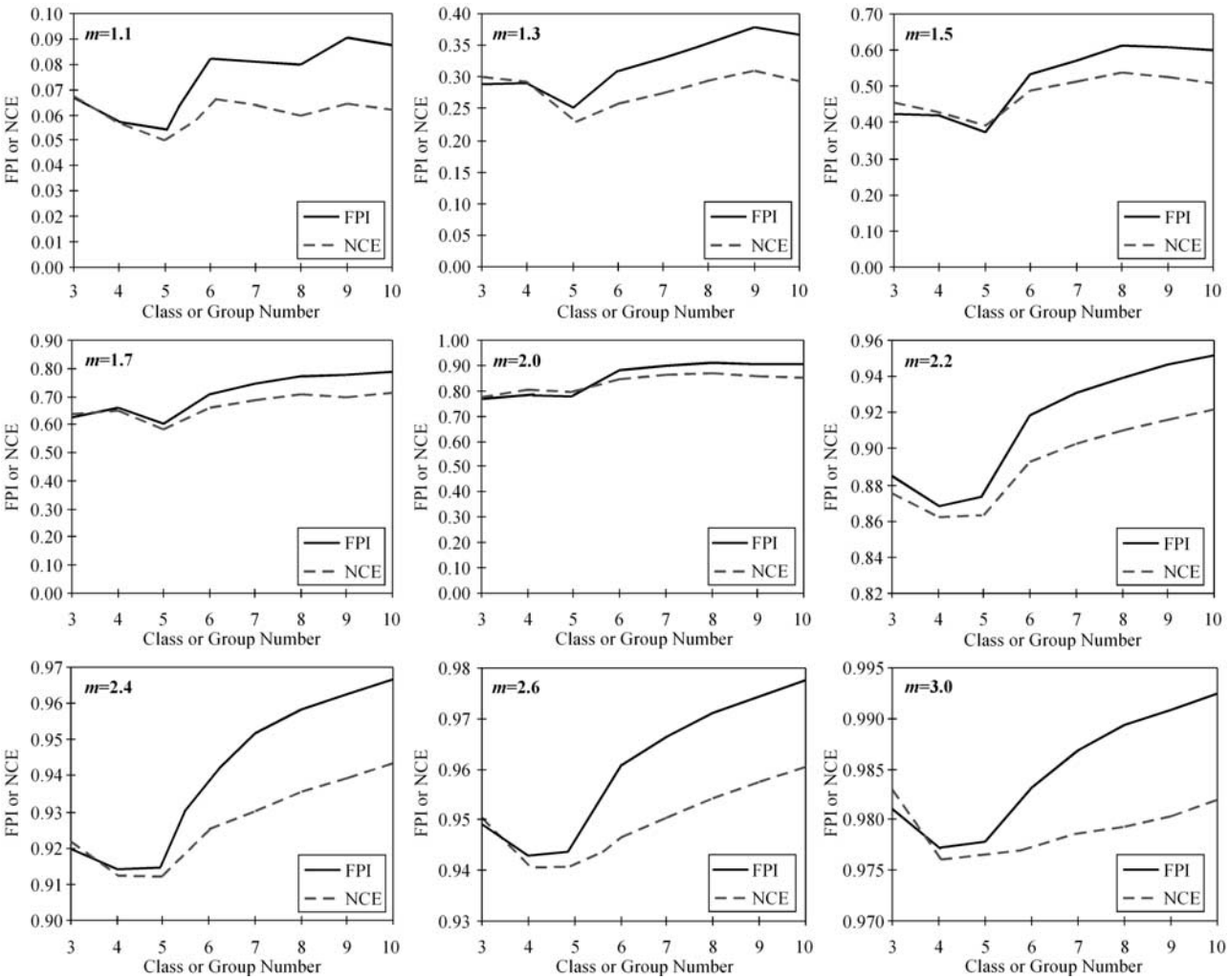

Figure 2. Plots of fuzziness performance index (FPI) and normalized classification entropy (NCE) values against the number of classes for the Indian Wells–Owens Valley data set for fuzzification parameter (m) values varying between 1.1 and 3.0. Minima of the FPI and NCE curves indicate the optimal number clusters existing in the data set.

amounts of data. Both can utilize chemical and physical measurements, although FCM can accommodate qualitative data in addition to the quantitative data used in this study. Both techniques produced five clusters, with most samples belonging to the same group regardless of method. The exceptions are the samples from Panamint Valley that were clustered into group-5 by the HCA analysis, but group-4 by the FCM technique and samples from southern Owens Valley where the classification was reversed (group-4 by HCA, but group-5 by FCM). The chemical significance of the clusters is discussed in detail below, but both techniques produce clusters that are geochemically valid as descriptions of the expected water-rock interactions along the regional flow paths [*Güler*, 2002]. In addition, the HCA technique requires the user to select the number of groups, while the FCM method employs an objective method to choose the number of groups. In addition to using an objective methodology to choose the number of clusters, FCM also produces clusters with gradational membership (soft partition) allowing plotting of this property as gray scales or color shades. The FCM technique also produces areas with no color (Figure 3a) that indicate transitional areas where membership is divided between more than two clusters (each having membership values below the 50% threshold ($u < 0.5$) and therefore they do not belong to any of the groups).

[29] There is a difference based on contouring technique in that kriging extrapolates cluster membership values (u) from the FCM analysis over the entire study area even to portions without any data (Figure 3a), while the hand contours simply outline samples with the same cluster values that were determined by HCA (Figure 3b). There are also occasional artifacts generated such as the group-1 triangular contours in the northwestern corner (Figure 3a). Nevertheless, the patterns are very similar including many of the smaller-scale details. The FCM clusters also show the degree of membership by shading and the transitional zones (white color), which is additional information not available from HCA.

[30] In order for the statistical clusters to be useful, they should create a representation of the hydrochemical system that is robust and can be tested by additional means such as spatial coherence within the statistical groups, spatial correspondence between the groups, and hydrologic controls

Table 1. First 20 Rows of the FCM Membership Matrix (*M*)

Sample Number	Class[a]	Membership Value (*u*)				
		1	2	3	4	5
SP1	3	0.0005	0.0459	(0.8921)	0.0597	0.0018
SP10	2	0.0082	(0.8281)	0.1608	0.0024	0.0005
SP100	5	0.0033	0.0241	0.0337	0.0532	(0.8858)
SP101B	3	0.0008	0.1670	(0.8261)	0.0053	0.0009
SP102F	2	0.0015	(0.9479)	0.0502	0.0003	0.0001
SP103	3	0.0003	0.2035	(0.7796)	0.0073	0.0092
SP104	2	0.1295	(0.8022)	0.0653	0.0021	0.0009
SP105	1	(0.9989)	0.0009	0.0002	0.0000	0.0000
SP106	5	0.0001	0.0039	0.0049	0.0016	(0.9896)
SP107	4	0.0059	0.0885	0.2871	(0.5199)	0.0987
SP108	2	0.0030	(0.8825)	0.1126	0.0016	0.0004
SP109	2	0.0004	(0.5845)	0.4131	0.0016	0.0004
SP11	3	0.0022	0.1459	(0.8134)	0.0373	0.0013
SP110	2	0.0021	(0.6936)	0.2991	0.0040	0.0012
SP111	2	0.0023	(0.9046)	0.0919	0.0008	0.0005
SP112	4	0.0002	0.0022	0.0130	(0.9726)	0.0120
SP113B	4	0.0001	0.0101	0.2354	(0.7063)	0.0481
SP114	3	0.0042	0.1599	(0.7989)	0.0340	0.0029
SP115	3	0.0004	0.1144	(0.8737)	0.0085	0.0031
SP116	2	0.3113	(0.6705)	0.0173	0.0005	0.0004

[a]Class memberships on the basis of which rows were selected are in parentheses.

such as topography and hydrochemical evolution that can be supported by inverse geochemical modeling [*Thyne et al.*, 2004].

[31] The FCM clusters display spatial distributions that are closely related to physiography. Group-1 samples are all located in the recharge areas of the high Sierra Nevada and plot above the 2000-m contour line. Group-2 samples are mostly located below the 2000-m contour line in the Sierra Nevada and other mountain ranges. These areas provide the majority of recharge to the basin-fill aquifers in the area. Group-3 samples are usually located on the basin floors, and are spatially between group-2 and group-4 samples. Group-4 samples are found in or very near the discharge areas (playas). Group-5 waters have the highest TDS concentrations and coincide with the playa or adjacent discharge areas (Figure 3a) and are characterized by lower Ca and Mg concentrations than group-4 waters (Table 2). The high degree of spatial and statistical coherence suggests that the changes between the hydrochemical facies define the normal hydrochemical evolution of water in the region.

[32] The TDS content of the waters is expected to increase as the water moves down the flow path and water-rock interaction takes place. In fact, the data show as the water moves eastward from the Sierran recharge areas to the discharge areas near and around the playas (Figure 3a), evolving from very dilute water (TDS = ~78 mg L^{-1}) to dilute water (TDS = ~319 mg L^{-1}) to brackish water (TDS = ~883 mg L^{-1}) to more saline water (TDS = ~8584 mg L^{-1}), and finally to brine (TDS = ~17,716 mg L^{-1}, with TDS locally as high as 342,186 mg L^{-1}). Thus the water groups, as determined by the FCM, appear to define hydrologic features such as discharge zones and the hydrochemical evolution between recharge zones and discharge zones.

4.4. Inverse Geochemical Modeling

[33] The validity of the statistical clusters, as representations of the hydrochemical evolution of groundwater, can be further evaluated with inverse geochemical models. Inverse modeling uses the mass balance approach to calculate all stoichiometric reactions that can produce the observed chemical changes between end-member waters [*Plummer and Back*, 1980]. A set of reactions needs to be defined before the inverse geochemical model can calculate the stoichiometric coefficients required to produce the observed set of compositions. This mass balance technique has been used to quantify reactions controlling water chemistry along flow paths and quantify mixing of end-member components in a flow system. For instance, *Garrels and MacKenzie*

Table 2. Main Features of the Cluster Centers (Representative Clusters) for Five Classes (Groups) Derived by FCM Objective Function and Mean Water Chemistry of Five Principal Water Groups Determined From Hierarchical Cluster Analysis (HCA)[a]

Parameters	N[b]	S. Cond.	pH	Ca	Mg	Na	K	Cl	SO_4	HCO_3	SiO_2	F
FCM Means Calculated Using Samples With $u \geq 0.50$												
Group-1	88	101	7.4	9	1.2	11	1.0	2.0	3.8	56	22	0.3
Group-2	236	515	8.0	33	7.7	66	3.6	44	40	184	32	0.9
Group-3	213	1,372	7.8	81	29.4	185	11	192	184	326	36	1.1
Group-4	94	10,591	7.5	185	125.5	2,817	65	4,153	883	585	61	3.2
Group-5	60	17,529	9.0	8	7.3	7,112	185	7262	1447	3296	36	11.2
FCM Means Calculated Using Samples With $u \geq 0.75$												
Group-1	81	90	7.3	9	1.2	8	1.0	1.5	2.9	52	22	0.3
Group-2	171	477	8.0	33	6.5	59	3.0	40	39	171	31	0.8
Group-3	145	1,222	7.8	79	30.9	152	10	138	186	322	35	1.0
Group-4	70	11,529	7.5	214	146.8	3153	68	4,696	1,109	484	61	1.6
Group-5	47	21,359	8.9	6	4.5	8,854	232	9,100	1,782	3,965	35	13.6
HCA Means Calculated Using All Samples												
Group-1	81	72	7.3	7	0.9	7	0.9	1.3	1.6	45	22	0.3
Group-2	229	542	8.0	37	8.5	67	3.6	41	51	195	33	0.9
Group-3	168	1,665	7.9	77	33.7	236	13	199	192	442	43	1.4
Group-4	80	7,495	7.9	143	79.9	1,574	52	1,900	892	660	41	2.2
Group-5	20	78,863	8.9	106	110.3	35,501	714	42,844	5,635	9,418	34	28.4

[a]FCM objective function means were calculated using samples having membership values ≥0.50 and 0.75. Group-1 and group-2 means were used in inverse geochemical calculations.

[b]Number of samples within respective FCM classes or principal HCA groups. Specific conductance (μS cm^{-1}), pH (standard units), mean concentrations (mg L^{-1}).

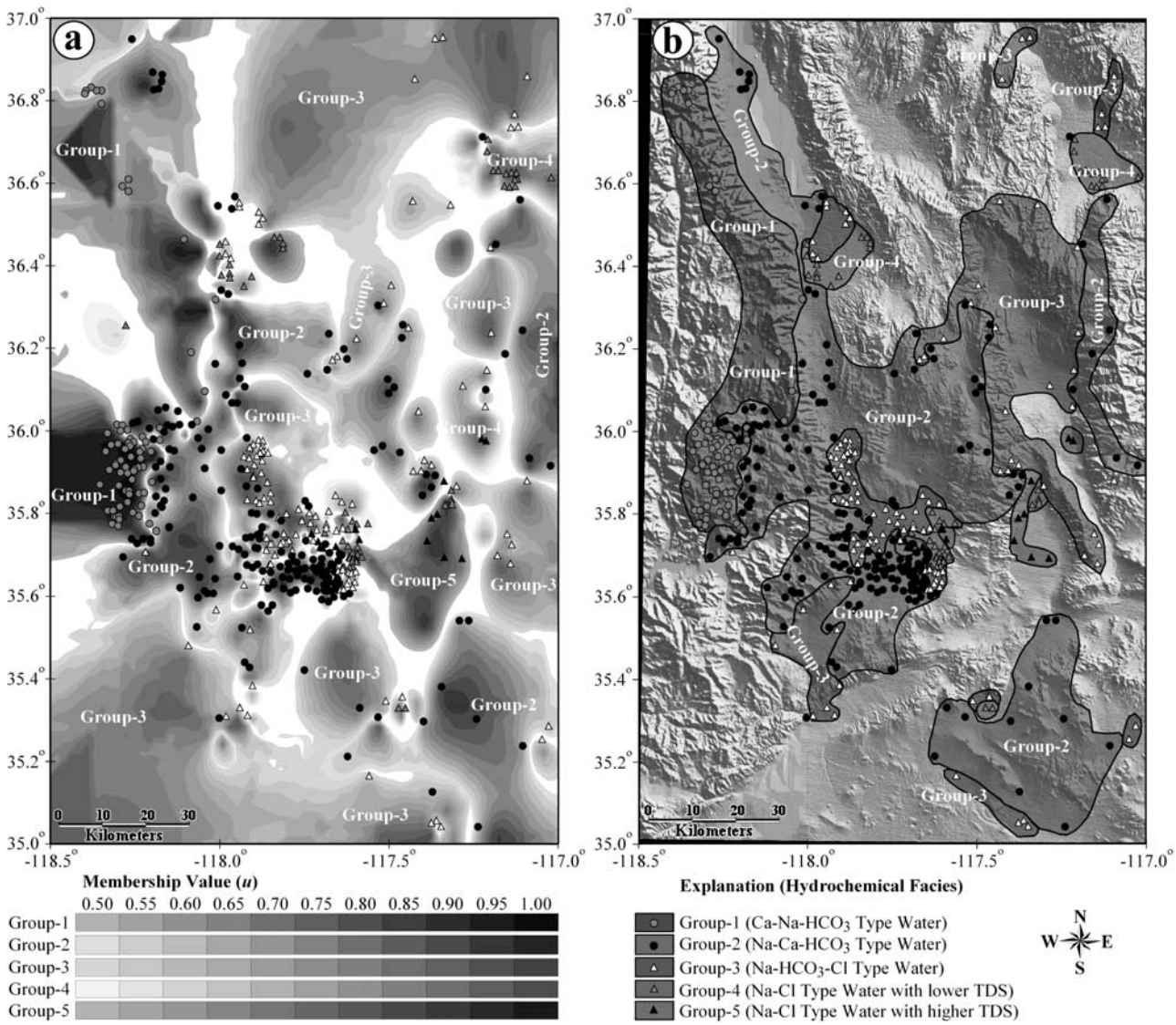

Figure 3. (a) Spatial distribution of the five hydrochemical facies determined by FCM objective function. Cluster memberships (u) for each facies range between 0.5 and 1 and separately interpolated on a regular grid by using point kriging algorithm. Each hydrochemical facies overlain on a single map and color-coded for better visualization purposes. Darker colors indicate strong membership to a particular facies, whereas lighter colors indicate a weak membership. (b) Spatial distribution of the statistically defined (by HCA) five principal water groups in the study area (modified from *Güler and Thyne* [2004]). In the HCA hard clustering method, samples are classified either as a group member or not (e.g., 1 or 0). Therefore, in the HCA method, samples close to group boundaries are more prone to misclassification than soft clustering methods such as FCM. See color version of this figure at back of this issue.

[1967] modeled Sierran spring water evolution for a small increase (80 ppm) in total dissolved solids (TDS) successfully. Inverse modeling is more difficult in systems that have similar amounts of several ions with multiple possible sources or large TDS changes typical in arid western watersheds [*Thomas et al.*, 1989].

[34] Three cases were evaluated, which include (1) the means of the HCA groups; (2) FCM cluster means where membership in the cluster was limited to samples with u = 0.50 or greater; and (3) FCM cluster means where membership in the cluster was limited to samples with u = 0.75 or greater membership values. Table 2 shows the means for the three cases. These mean values are used for inputs, together with the appropriate mineralogy [*Güler*, 2002], for the inverse geochemical models. All three cases produced viable inverse modeling solutions. Table 3 shows the results for the first two steps in the hydrochemical evolution, those of precipitation to group-1 samples and group-1 to -2 samples. While there are small differences, the basic reaction derived from inverse modeling for all three cases is the same:

Model 1

$$\text{precipitation} + CO_2\ \text{gas} + \text{plagioclase} + \text{biotite} + \text{K-feldspar} + \text{halite} + \text{gypsum} = \text{group-1 water} + \text{amorphous silica} + \text{kaolinite} + \text{calcite}$$

Table 3. Summary of Mass Transfer for Selected Inverse Geochemical Models Using Three Different Group Means Calculated Using FCM (for $u \geq 0.50$ and 0.75) and HCA[a]

	FCM ($u \geq 0.50$)		FCM ($u \geq 0.75$)		HCA	
Phases	Precipitation to Group-1	Group-1 to Group-2	Precipitation to Group-1	Group-1 to Group-2	Precipitation to Group-1	Group-1 to Group-2
Amorphous silica	−4.99E-04[b]	−3.08E-03	−3.44E-04	−2.74E-03	−2.49E-04	−3.54E-03
Biotite	1.36E-05	...	1.36E-05	...	1.15E-05	6.81E-05
Calcite	−5.03E-05	−7.16E-04	...	−6.14E-04	...	−8.74E-04
CO_2(g)	8.20E-04	3.13E-03	7.12E-04	2.77E-03	5.68E-04	3.68E-03
Forsterite	...	1.34E-04	...	1.08E-04	...	5.11E-05
K-feldspar	...	6.65E-05	...	5.12E-05	3.75E-06	...
Plagioclase	6.71E-04	2.40E-03	5.43E-04	2.17E-03	4.63E-04	2.85E-03
Kaolinite	−4.69E-04	−1.69E-03	−3.81E-04	−1.52E-03	−3.27E-04	−2.00E-03
Gypsum	2.91E-05	3.79E-04	1.97E-05	3.73E-04	3.86E-06	5.15E-04
Halite	4.03E-05	1.18E-03	2.91E-05	1.09E-03	2.44E-05	1.13E-03

[a]Model-1, precipitation to group-1 waters; model-2, group-1 to group-2 waters. Phases and thermodynamic data are from PHREEQC and accompanying databases [*Parkhurst and Appelo*, 1999]. Values in moles per kilogram H_2O (positive values indicate dissolution and negative values indicate precipitation). Ellipsis dots indicate phase not used in model.
[b]Read as -4.99×10^4.

Model 2

group-1 water + CO_2 gas + plagioclase + biotite + K-feldspar + forsterite + halite + gypsum = group-2 water + amorphous silica + kaolinite + calcite

[35] This compares well with other models for the region [*Feth et al.*, 1964; *Garrels and Mackenzie*, 1967; *Thomas et al.*, 1989] and indicates that all three approaches can define viable hydrochemical evolution models for the region. The FCM analysis also provides a more robust classification methodology. Water samples that are close to the center of a HCA cluster (samples having FCM membership values close to 1) are strong representatives of the cluster and are usually correctly classified by HCA. However, samples that are close to the boundaries of clusters (samples having FCM membership values close to 0.50 or lower) are weaker representatives and are prone to misclassification. However, even though FCM produces much better mapping information (gray or color scale) and is more robust than the HCA, it cannot entirely eliminate the problem of sample misclassification. The samples most prone to misclassification are those in white colored areas in Figure 3a.

5. Discussion

[36] The changes in water chemistry in the study area are complex with a transitional character. The statistical associations present the information in a compact format as the first step to determine if samples can be grouped into distinct populations (hydrochemical facies) and help define a preliminary hydrochemical evolution for the study area. The associations (clusters) can be tested for spatial coherence, geological relevance, and hydrochemical validity to establish cause-and-effect relationships. In this case, the hydrochemical facies showed strong spatial coherence and correlated well with the major hydrologic features, an unlikely result if the statistical associations did not reflect underlying physical and chemical processes.

[37] The mean values of the clusters also served as input to inverse geochemical models to further verify that the statistical clusters were significant in the hydrochemical context. The use of mean values from clusters for inverse model inputs enhances the power of the approach, which works best in simple systems with a few dominant water-rock reactions. Cluster analysis creates statistical groups from multiple parameters, reducing the variability, which can be viewed as information noise, while still extracting the essential aspects of each group. This input is particularly useful in handling large data sets and defining the important components of hydrochemical evolution.

[38] The fuzzy clusters are fundamentally similar to the HCA clusters but offer the additional advantages of a more objective criteria for choosing the number of clusters and gradation of cluster (degree of membership) that are impossible with hard partitioning. This information can be presented as contours using a variety of standard techniques such as kriging. The ability to set the threshold for membership can also allow multiple realizations that can be evaluated for the best match to criteria such as physiographic and hydrologic features, internal spatial coherence, and hydrochemical validity.

6. Summary and Concluding Remarks

[39] In this study, the FCM clustering technique was used to create a spatially and geochemically coherent classification of water chemistry data from a regional groundwater system. Using the FCM technique, five hydrochemical facies were delineated and cluster centers representing each hydrochemical facies were determined. Plotting the kriged cluster membership value contours on a map demonstrated the existence of five spatially continuous, well-defined clusters of water samples. Mean values of each variable from the hydrochemical facies were used for further evaluation of the system with inverse geochemical models that describe the major processes creating variability between samples. Hydrochemical facies defined by FCM can be also used to study the groundwater flow systems and as a tool to verify aquifer connectivity in the same fashion as hydrochemical facies defined by nonfuzzy methods.

[40] Our results show that a fuzzy classification gives improved spatial definition and less error sensitive clusters than a classical hard classification. Water samples close to class boundaries are especially prone to misclassification

when using conventional hard clustering methods such as HCA. In contrast, fuzzy clustering allows each sample to belong to several clusters minimizing misclassification. The membership values (u) of the clusters, which are a statistical composite of all the clustered parameters, can be displayed as contour maps. These contour maps can be overlain with other graphical information such as topography, lithology, etc., to produce enhanced visualization and analysis.

[41] Our approach used the Euclidean distance and a fuzziness coefficient (m) of 1.3 for optimal clustering. The five groups produced by the FCM analysis were hydrochemically meaningful. The results suggest that fuzzy classification may better reflect the continuous nature of the processes underlying the generation of water chemistry variability in the study area. Furthermore, fuzzy clustering methods can handle large data matrices (typical of regional-scale studies) [*Höppner*, 2002; *Kolen and Hutcheson*, 2002] and provide a rapid, systematic, and objective means for reducing large amounts of hydrochemical data into more useful information. We conclude that the robust classification scheme for partitioning water chemistry samples into homogeneous groups produced by FCM can be an important tool for the successful characterization of regional-scale hydrogeologic systems.

[42] **Acknowledgment.** The authors thank the Ministry of National Education of Turkey for the senior author's financial support.

References

Ahmed, M. N., S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty (2002), A modified fuzzy *c*-means algorithm for bias field estimation and segmentation of MRI data, *IEEE Trans. Med. Imaging*, *21*, 193–199.

Asharaf, S., and M. N. Murty (2003), An adaptive rough fuzzy single pass algorithm for clustering large data sets, *Pattern Recognit.*, *36*, 3015–3018.

Barbieri, P., G. Adami, A. Favretto, A. Lutman, W. Avoscan, and E. Reisenhofer (2001), Robust cluster analysis for detecting physicochemical typologies of freshwater from wells of the plain of Friuli (northeastern Italy), *Anal. Chim. Acta*, *440*, 161–170.

Belacel, N., P. Hansen, and N. Mladenovic (2002), Fuzzy J-means: A new heuristic for fuzzy clustering, *Pattern Recognit.*, *35*, 2193–2200.

Bezdek, J. C. (1974), Cluster validity with fuzzy sets, *J. Cybern.*, *3*, 58–73.

Bezdek, J. C. (1975), Mathematical models for systematics and taxonomy, in *Proceedings of the Eighth International Conference on Numerical Taxonomy*, edited by G. Estabrook, pp. 143–166, W. H. Freeman, New York.

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

Bezdek, J. C., and N. R. Pal (1998), Some new indexes of cluster validity, *IEEE Trans. Syst. Man Cybern.*, *28*, 301–315.

Bezdek, J. C., R. Ehrlich, and W. Full (1984), FCM: The fuzzy *c*-means clustering algorithm, *Comput. Geosci.*, *10*, 191–203.

Bezdek, J. C., L. O. Hall, and L. P. Clarke (1993), Review of MR image segmentation techniques using pattern recognition, *Med. Phys.*, *20*, 1033–1048.

Bezdek, J. C., J. M. Keller, R. Krishnapuram, and N. R. Pal (1999), *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Kluwer Acad., Norwell, Mass.

Burrough, P. A., P. F. M. van Gaans, and R. A. McMillan (2000), High-resolution landform classification using fuzzy *k*-means, *Fuzzy Sets Syst.*, *113*, 37–52.

Chang, N.-B., H. W. Chen, and S. K. Ning (2001), Identification of river water quality using the fuzzy synthetic evaluation approach, *J. Environ. Manage.*, *63*, 293–305.

Dunn, J. C. (1974), A fuzzy relative ISODATA process and its use in detecting compact well-separated clusters, *J. Cybern.*, *3*, 32–57.

El-Sonbaty, Y., and M. A. Ismail (1998), Fuzzy clustering for symbolic data, *IEEE Trans. Fuzzy Syst.*, *6*, 195–204.

Fang, J. H., and H. C. Chen (1990), Uncertainties are better handled by fuzzy arithmetic, *AAPG Bull.*, *74*, 1228–1233.

Feth, J. H., C. E. Roberson, and W. L. Polzer (1964), Sources of mineral constituents in water from granitic rocks, Sierra Nevada, California and Nevada, *U.S. Geol. Surv. Water Supply Pap., 1535-I*.

Fukuyama, Y., and M. Sugeno (1989), A new method of choosing the number of clusters for the fuzzy *c*-means method, in *Proceedings of the Fifth Fuzzy Systems Symposium*, pp. 247–250, Japan.

Garrels, R. M., and F. T. MacKenzie (1967), Origin of the chemical compositions of some springs and lakes, in *Equilibrium Concepts in Natural Waters*, pp. 222–242, Am. Cancer Soc., Washington, D. C.

Geva, A. B. (1999), Hierarchical unsupervised fuzzy clustering, *IEEE Trans. Fuzzy Syst.*, *7*, 723–733.

Gunderson, R. (1978), Application of fuzzy ISODATA algorithms to star-tracker pointing systems, paper presented at the 7th Triannual World IFAC Congress, Helsinki, Finland.

Güler, C. (2002), Hydrogeochemical evaluation of the groundwater resources of Indian Wells–Owens Valley area, southeastern California, Ph.D. dissertation, 262 pp., Colo. School of Mines, Golden, Colo.

Güler, C., and G. D. Thyne (2004), Hydrologic and geologic factors controlling surface and groundwater chemistry in Indian Wells–Owens Valley area, southeastern California, USA, *J. Hydrol.*, *285*, 177–198.

Güler, C., G. Thyne, J. E. McCray, and A. K. Turner (2002), Evaluation of graphical and multivariate statistical methods for classification of water chemistry data, *Hydrogeol. J.*, *10*, 455–474.

Hammah, R. E., and J. H. Curran (1999), On distance measures for the fuzzy *k*-means algorithm for joint data, *Rock Mech. Rock Eng.*, *32*, 1–27.

Hathaway, R. J., and J. C. Bezdek (2001), Fuzzy *c*-means clustering of incomplete data, *IEEE Trans. Syst. Man Cybern. B*, *31*, 735–744.

Höppner, F. (2002), Speeding up fuzzy *c*-means: Using a hierarchical data organization to control the precision of membership calculation, *Fuzzy Sets Syst.*, *128*, 365–378.

Hu, B.-G., R. G. Gosine, L. X. Cao, and C. W. de Silva (1998), Application of a fuzzy classification technique in computer grading of fish products, *IEEE Trans. Fuzzy Syst.*, *6*, 144–152.

Kersten, P. R. (1999), Fuzzy order statistics and their application to fuzzy clustering, *IEEE Trans. Fuzzy Syst.*, *7*, 708–712.

Kolen, J. F., and T. Hutcheson (2002), Reducing the time complexity of the fuzzy *c*-means algorithm, *IEEE Trans. Fuzzy Syst.*, *10*, 263–267.

Kothari, R., and D. Pitts (1999), On finding the number of classes, *Pattern Recognit. Lett.*, *20*, 405–416.

Linusson, A., S. Wold, and B. Nordén (1998), Fuzzy clustering of 627 alcohols, guided by a strategy for cluster analysis of chemical compounds for combinatorial chemistry, *Chemom. Intell. Lab. Syst.*, *44*, 213–227.

Maxey, G. B. (1968), Hydrogeology of desert basins, *Ground Water*, *6*, 10–22.

McBratney, A. B., and J. J. deGruijter (1992), A continuum approach to soil classification by modified fuzzy *k*-means with extragrades, *J. Soil Sci.*, *43*, 159–175.

McBratney, A. B., and A. W. Moore (1985), Application of fuzzy sets to climatic classification, *Agric. For. Meteorol.*, *35*, 165–185.

McNeill, D., and P. Freiberger (1993), *Fuzzy Logic: The Revolutionary Computer Technology That Is Changing Our World*, Simon and Schuster, New York.

Minasny, B., and A. B. McBratney (2002), FuzME version 3, Aust. Cent. for Precis. Agric., Univ. of Sydney, Sydney, N. S. W., Australia. (Available at http://www.usyd.edu.au/su/agric/acpa.)

Odeh, I. O. A., A. B. McBratney, and D. J. Chittleborough (1992), Soil pattern recognition with fuzzy *c*-means: Application to classification and soil-landform interrelationship, *Soil Sci. Soc. Am. J.*, *56*, 505–516.

Parkhurst, D. L., and C. A. J. Appelo (1999), User's guide to PHREEQC (Version 2): A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations, *U.S. Geol. Surv. Water Resour. Invest. Rep., 99-4259*.

Pedrycz, W. (1996), Conditional fuzzy *c*-means, *Pattern Recognit. Lett.*, *17*, 625–631.

Plummer, L. N., and W. W. Back (1980), The mass balance approach: Application to interpreting the chemical evolution of hydrological systems, *Am. J. Sci.*, *280*, 130–142.

Ramze, M., R. B. P. F. Lelieveldt, and J. H. C. Reiber (1998), A new cluster validity index for fuzzy *c* mean, *Pattern Recognit. Lett.*, *19*, 237–246.

Rantitsch, G. (2000), Application of fuzzy clusters to quantify lithological background concentrations in stream-sediment geochemistry, *J. Geochem. Explor.*, *71*, 73–82.

Roubens, M. (1982), Fuzzy clustering algorithms and their cluster validity, *Eur. J. Oper. Res.*, *10*, 294–301.

Setnes, M. (2000), Supervised fuzzy clustering for rule extraction, *IEEE Trans. Fuzzy Syst.*, *8*, 416–424.

Suk, H., and K.-K. Lee (1999), Characterization of a ground water hydrochemical system through multivariate analysis: Clustering into ground water zones, *Ground Water*, *37*, 358–366.

Sun, L.-X., and K. Danzer (1996), Fuzzy cluster analysis by simulated annealing, *J. Chemom.*, *10*, 325–342.

Thomas, J. M., A. H. Welch, and A. M. Preissler (1989), Geochemical evolution of ground water in Smith Creek Valley: A hydrologically closed basin in central Nevada, U.S.A., *Appl. Geochem.*, *4*, 493–510.

Thyne, G. D., C. Güler, and E. Poeter (2004), Sequential analysis of hydrochemical data for watershed characterization, *Ground Water*, *42*, 711–723.

Trauwert, E., L. Kaufman, and P. Rousseeuw (1991), Fuzzy clustering algorithms based on the maximum likelihood principle, *Fuzzy Sets Syst.*, *42*, 213–227.

Windham, M. P. (1981), Cluster validity for fuzzy clustering algorithms, *Fuzzy Sets Syst.*, *5*, 177–185.

Wu, K.-L., and M.-S. Yang (2002), Alternative *c*-means clustering algorithms, *Pattern Recognit.*, *35*, 2267–2278.

Xie, L. X., and G. Beni (1991), A validity measure for fuzzy clustering, *IEEE Trans. Pattern. Anal. Machine Intell.*, *13*, 841–847.

Zadeh, L. A. (1965), Fuzzy sets, *Inf. Control*, *8*, 338–353.

Zhang, C.-T., K.-C. Chou, and G. M. Maggiora (1995), Predicting protein structural classes from amino acid composition: Application of fuzzy clustering, *Protein Eng.*, *8*, 425–435.

C. Güler, Geological Engineering Department, Çiftlikköy Kampüsü, Mersin Üniversitesi, Mersin 33342, Turkey. (cguler@mersin.edu.tr)

G. D. Thyne, Department of Geology and Geological Engineering, Colorado School of Mines, 1500 Illinois Street, Golden, CO 80401, USA.

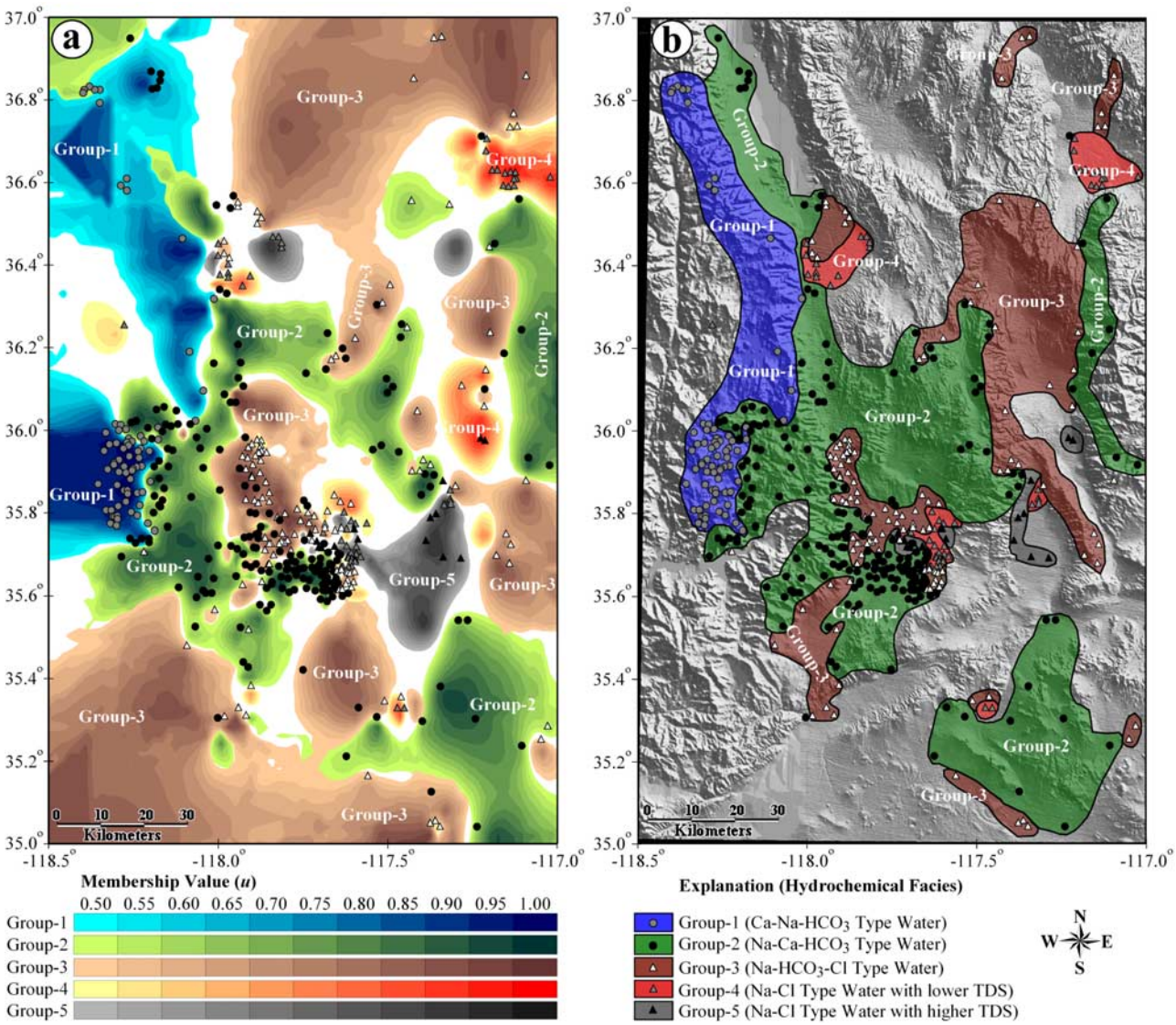

Figure 3. (a) Spatial distribution of the five hydrochemical facies determined by FCM objective function. Cluster memberships (u) for each facies range between 0.5 and 1 and separately interpolated on a regular grid by using point kriging algorithm. Each hydrochemical facies overlain on a single map and color-coded for better visualization purposes. Darker colors indicate strong membership to a particular facies, whereas lighter colors indicate a weak membership. (b) Spatial distribution of the statistically defined (by HCA) five principal water groups in the study area (modified from *Güler and Thyne* [2004]). In the HCA hard clustering method, samples are classified either as a group member or not (e.g., 1 or 0). Therefore, in the HCA method, samples close to group boundaries are more prone to misclassification then soft clustering methods such as FCM.