INTERNATIONAL JOURNAL OF PLANT, ANIMAL AND ENVIRONMENTAL SCIENCES

IJPAES

Volume-5, Issue-4, Oct-Dec-2015 Coden:IJPAJX-CAS-USA, Copyrights@2015 ISSN-2231-4490

Received: 16th Sept-2015 *Revised: 7th Oct -2015* *Accepted: 9th Oct-2015*

Research article

APPLICATION OF PCFC CLUSTERING ALGORITHM FOR ANALYSIS OF SURFACE WATER QUALITY IN GUNTUR CITY

Kamakshaiah.Kolli[1] and R. Seshadri[2*]

[1]SVU College of Engineering, SV University, Tirupati-India
[2]SVU Computer Center, SV University, Tirupati-India
*Corresponding author Address: ravalaseshadri@gmail.com

ABSTRACT: Surface water quality in rural areas usually has a great variation in India and is hard to characterize by classical statistic methods. In this paper, a PCFC clustering method is used to classify and assess rural surface water quality based on the monitoring data from 50 typical stations in 15 small cannals and 4 reservoirs in Guntur city. The results show that the 50 monitoring stations can be classified into 3 clusters in terms of water quality. The first cluster consists of 27 stations and most of their water quality indexes are nearly at or better than the national Grade II standards, while the second and third clusters respectively contain 13 and 10 stations, and their indexes of ammonia nitrogen and petroleum are at or worse than the national Grade V standards, and the index values in the third cluster generally execeed those in the second cluster. Thus, the overall quality or rural surface water in study are remains good, but there also exist some cannal sections contaminated with ammonia nitrogen and petroleum. Therefore, it is very necessary to establish water quality safety and risk assessment system for ensuring water supplies for production and daily life.

Key words: Rural surface water, Water quality assessment, Water quality index, PCFC clustering.

INTRODUCTION

Guntur city is a municipal corporation, which serves as the headquarters of mandal, revenue division and the district in Andhra Pradesh [1, 2]. It is the third most populous city in the state with a population of 743,354, and an urban agglomeration population of 1,028,667. As affected by seawater intrusion, its groundwater is severely salinized, and surface water plays a crucial role in maintaining people's normal life and production activities [3, 4]. With economic development and change of production and life style, rural water environmental problems are becoming more and more prominent, and water quality degradation has also caused more and more attention in recent years [5]. So it is essential to carry out water quality investigation and assessment for ensuring regional water safety and sustainable development [6, 7]. Cluster analysis is the task of assigning a set of objects into groups so that the objects in the same cluster are more similar to each other than to those in other clusters [8]. PCFC clustering can better describe degrees of each object belonging to different clusters by introducing the concept of membership function, so it can provide much more objective information for decision-making. Clustering, originally introduced by Jim Bezdck in 1981, is a kind of automatic clustering method without any intervention or supervision [9]. As it does not require much more auxiliary information and therefore is easy to be accepted. At present, the PCFC clustering technique has been widely used in many fields, including pattern recorgnition, image analysis and environmental management. In this paper, we try to apply the method of PCFC clustering to characterize and assess the rural surface water quality in Guntur city, so as to provide reference idea for environmental protection management.

MATERIALS AND METHODS

a) Water sample collection and analysis

Comprehensively considered the distribution of surface water bodies and their service functions, a total of 50 typical monitoring stations is identified from 15 small cannals and 4 reservoirs, and then their water samples are collected. The sampling and processing of water samples is in accordance with the requirements of "Technical SPECIFICATION Requirements for Monitoring of Surface Water and Waste Water.

Environmental Protection Agency, and water sample analysis is carried out according to the standard methods in reference. Amount all the water quality indexes, seven indexes including permanganate index, BOD5, ammonia nitrogen, petroleum, volatile phenol, total Hg and total Pb are chosen for the following water quality clustering and assessment.

b) PCFC algorithm

There are a lot of literatures about PCFC algorithm, here we only take water quality cluster analysis as an example to briefly introduce its basic theory, and more details can be found in the reference.

Let xijbe the measurement value of the jth water index of the ith monitoring station, then the water quality monitoring data set X can be expressed as the following matrix.

$$\begin{vmatrix} X_1^1, X_1^2, X_1^3, X_1^4 & \cdots & \cdots & \ldots,\ldots, X_1^{d-1}, X_1^d \\ \cdot\cdot & & & \cdot \\ \cdot\cdot & & & \cdot \\ \cdot & & & \ldots,\ldots, X_{n,}^{d-1} X_n^d \end{vmatrix} \| FCM \Rightarrow \begin{vmatrix} r_1^1 & \cdot & \cdots & .r_1^d \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ r_m^1 & & & .r_m^d \end{vmatrix}$$

Where n is the number of monitoring stations and p is the number of water quality indexes. And water quality clustering based on fuzzy c-means algorithm is a process of partitioning X into c ($2 \leq c \leq n$) clusters. Meanwhile, the c X p dimensional cluster center matrix, V, can be obtained.

$$V = \begin{vmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ v_{c1} & v_{c2} & \cdots & v_{cp} \end{vmatrix}$$

Where vik ($1 \leq i \leq c$, $1 \leq k \leq p$) is the centeroid value of the pth water quality index of the ith cluster. In fuzzy clustering, each object has a degree of belonging to clusters rather than belonging completely to just one cluster. Suppose uik ∈ [0,1] denotes the membership value of the ith station belonging to the kth cluster, then the objective function can be defined as:

$$F_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} u_{ij}^{z} \left\| X_i^{ref} - C_j^{mod} \right\|.^2$$

Where dik is the distance measure to characterize the similarity between the ith monitoring station and the kth which affects clustering results and controls the overlapping degree of each monitoring station belonging to different clusters; U is n x c dimensional membership degree matrix, which characterizes the degrees of monitoring stations belonging to different clusters. The basic process of PCFC algorithm is to seek the numerical solution when the objective function (3) gets minimum value. Accordingly, the elements, vkj and uik, of cluster centre matrix V and membership degree matrix U can be obtained by using iterative computation method. Usually, the selection of parameters c and m will directly affects the clustering results. In this paper, a method of multiple combinations of c and m, proposed by McBratney and Moore, is used to determine their optimal values. In this method, a derived function –[(ΔJ/Δm)c0.5] is first built, and then a set of curves for –[(ΔJ/Δm)c0.5] as a function of c and m are plotted. In a general way, the value of c, which is corresponded to the curve with smallest peak value, is the optimal number of clusters, and the value of m, which is corresponded to the curve's peak, is the best choice of fuzziness exponent.

RESULTS AND ANALYSIS

a) Statistical analysis for surface water quality monitoring data

Table-1: Descriptive statistics of monitoring data of the seven water quality indexes.

Item	Permanganate Index	BOD_5	Ammonia Nitrogen	Petroleum	Volatile Phenol	Hg	Pb
Max. (mg/L)	11.6	11.56	12.89	0.54	0.003	0.00002	0.0040
Min.(mg/L)	4.3	1.80	0.43	0.03	0.002	0.00001	0.0007
Mean.(mg/L)	4.8	3.78	1.47	0.08	0.001	0.00002	0.004
Standard deviation (mg/L)	2.21	2.42	2.67	0.16	0.002	0.00000	0.002
Variation coefficient (%)							

By comparing Table 1 and Table 2, we can divide the seven water quality parameters into 3 groups according to their differences between the mean values and the national standard limits. The first group includes volatile phenol, total Hg and total Pb and their average concentrations are below the upper limit values of Grade II, the second group contains permanganate index and BOD_5 and their mean values meet the third of national standards; and the third group is composed of ammonia nitrogen and petroleum and their mean values nearly reach the upper limit values of Grade IV. In addition, we also find that some of the index, such as ammonia nitrogen, petroleum and BOD5, have a relatively high coefficient of variation, even up to 198%, which means there may be exists serious degradation of water quality at some of the monitoring stations. Therefore, it is very necessary to make a classification of the monitoring stations for assessing water quality based on the monitoring data and thus to make better targeted response measures.

Table-2: National standard limits of water quality indexes (mg/L)

National Standard limits	Permanganate Index	BOD_5	Ammonia Nitrogen	Petroleum	Volatile Phenol	Hg	Pb
Grade I	≤2	≤3	≤0.15	≤0.05	≤0.002	≤0.00005	≤0.01
Grade II	≤4	≤3	≤0.5	≤0.05	≤0.002	≤0.00005	≤0.01
Grade III	≤6	≤4	≤1.0	≤0.05	≤0.005	≤0.0001	≤0.05
Grade IV	≤10	≤6	≤1.5	≤0.5	≤0.01	≤0.0001	≤0.05
Grade V	≤15	≤10	≤2.0	≤1.0	≤0.1	≤0.001	≤0.1

b) Fuzzy c-mean clustering for assessing surface water quality

In this paper, the water quality clustering is carried out by using matlab programming method with the FCM function of Matlab embedded. First, a 33x7 matrix X is substituted in the clustering program, meanwhile the cluster number c is specified as integers between 2 and 7, and the parameter of fuzzy exponent m is set as values ranging from 1.1 to 2.4 with a step size of 0.1 Besides, the maximum iterating times and the changing degree of the optimum solutions during iterating are used to control the terminating time, here the maximum iterations and the changing degree are set as 300 and 10-4 respectively. Once the program is running, a total of 84 kinds of fuzzy clustering results are given under combinations of c and m with different values. Fig.2 shows the computed results of derived function –[(ΔJ/Δm)c0.5] when c and m take different values. From Fig.2 we can find that among all the curves of –[(ΔJ/Δm)c0.5] against m, the curve with smallest peak value is when c equals 3, and at the same time the curve obtains its peak value when m equals 2.2 So the optimal c and m is 3 and 2.2 respectively.

Table 3 presents the optimal cluster centers when c = 3 and m = 2.2, and it reflects the average state of each water quality index of each cluster and is also used for assessing surface water quality and making management measures. As can be seen from Table 3, the values of water quality indexes except total Hg and total Pb are increasing from cluster A to cluster C. In cluster A, most of the water quality indexes are at or better than the national surface water Grade II standards, except that permanganate index is slightly worse than the national Grade II standard, which means the water quality of monitoring stations belonging to cluster A is good and no special measures are needed in the future environment management. While in cluster B and cluster C, the indexes of ammonia nitrogen and petroleum are at or worse than the national Grade V standards and the other indexes except volatile phenol, total Hg and total Pb are worse than the national Grade III standards, which means their water quality is poor and has been subject to different degrees of pollution, especially for the stations in cluster C, their water quality is seriously degraded and further measures are needed to treat pollution.

Table 3: Centroid values of the 7 water quality indexes corresponding to the fuzzy clusters (mg/L)

Cluster	Permanganate Index	BOD_5	Ammonia Nitrogen	Petroleum	Volatile Phenol	Hg	Pb
A	4.1	2.4	0.3	0.03	0.0015	0.0004	0.004
B	7.5	4.9	2.7	0.21	0.0028	0.00003	0.004
C	7.8	7.9	12.5	0.67	0.0044	0.00001	0.004

According to the principle of maximum membership degree, there is only one monitoring station, namely the station of S25, that can be classified into cluster C, and 5 stations of S04, S07, S12, S33 and S34 can be placed under cluster B, while the rest of 27 stations, accounting for 82% of the total, belong to cluster A. Thus it can be seen that most of monitoring stations meet the national grade II or III surface water standard, and overall quality of rural surface water in Guntur city remains good. But it is necessary to emphasize that some river sections have been contaminated with ammonia nitrogen and petroleum to varied extents and much more attention should be paid to them.

CONCLUSION

Surface water quality in rural areas usually has a great variation due to the differences of economic development and production style and is hard to characterize by classical statistic methods. The paper applies PCFC algorithm to cluster and assess the rural surface water quality based on the monitoring data from 50 typical monitoring stations in 15 cannals and 4 reservoirs. The results show that the optimal cluster number is 3 in terms of water quality in the study area, and the water quality indexes in the first cluster are nearly at or better than the national surface water Grade II standards, while in the second and third clusters the indexes of ammonia nitrogen and petroleum are at or worse than the national Grade V standards and the index valves in the third cluster generally exceed those in the second cluster. According to the principle of maximum membership degree, there are 23 monitoring stations can be classified into the second cluster and only 10 station can be put under the third cluster, which means most of the monitoring stations meet the national grade II or III surface water standard and the overall quality of rural surface water in Guntur city remains good, but there also exist some river sections contaminated with ammonia nitrogen and petroleum, and therefore it is very necessary to establish water quality safety and risk assessment system for ensuring water supplies for production and daily life.

REFFERENCES

[1] "Guntur History". National Informatics Centre. Retrieved 27 July 2014.
[2] "About Guntur". Vijayawada: VGTM Urban Development Authority. Retrieved 3 August 2014.
[3] “Governing body". Guntur Municipal Corporation. Retrieved 10 June 2014.
[4] "GMC Commissioner". Guntur Municipal Corporation. Retrieved 28 July 2014.
[5] "Municipal Corporation Profile" (PDF). Guntur Municipal Corporation. Retrieved 10 October 2014.
[6] "Guntur city population is 7, 43, 354 as per 2011 Census". The Hindu (Guntur). 26 May 2013. Retrieved 12 October 2014.
[7] "Demography". National Informatics Centre. Retrieved 30 April 2015.
[8] Inderjit S. Dhillon and Dharmendra S. Modha, 2001. "Concept Decompositions for Large Sparse Text Data using Clustering," Machine Learning, 42(1), pp. 143-175.
[9] Fionn. Murtagh, Jean-Luc Starck, and Michael W. Berry, 2000. "Overcoming the curse of dimensionality in clustering by means of the wavelet transform", The Computer Journal, 43, 107-120.

ISSN 2231-4490

International Journal of Plant, Animal and Environmental Sciences