Evaluation of PNN model in groundwater quality classification in term of WQI

Asiyeh mousavi rad[1],Salman sharif azari[1],Mehdi amian[2], Parviz haghighatju[1].

1.Department of water engineering, Faculty of soil and water, University of Zabol

2.MSC. Educated in Biomedical Engineering School Of Electerical and Computer engineering, colleg of Engineering, college of Engineering, University of Tehran, Tehran, Iran.

Abstract

In Iran, groundwater resources are of high importance as one of the most important water supplies for agriculture, domestic, and industrial uses. Agriculture with a contribution of 95 percent and exploitation of more than 80 percent of groundwater resources plays a significant role in quantity and quality of aquifers. In this paper Probability Neural Network (PNN) is used for qualitative classification of groundwater resources in Fars province. The qualitative data of 2000 wells, qanats, and springs in Fars province were used for this study. The results showed that in most parts of the province, groundwater quality is proper. Also, PNN has an applied role in groundwater quality estimation.

Keywords: PNN, qanats, springs, wells.

Introduction

Water shortage is regarded as essential challenge and restriction to development and prosperity of countries in national, regional and social life levels and also to economic activities. In addition to water shortage, if the exploited waters are polluted, that nowadays is an essential problem, the problems associated to water as a vital factor for the human life will double [1]. Groundwater is one of most important natural resources in the globe; and a significant portion of consumed water particularly in Iran is supplied by groundwater [2].

There are several methods to evaluate the quality of water resources across the world. High diversity of parameters confirms the claim that specific parameters become important due to any specific application. Many physical and chemical indices presented in various papers and studies can be categorized in four categories: general Water Quality Indices (WQI), special consumption indices, design indices, statistical indices. Biologic indices can be categorized as a separate category [3, 4].

WQI determines the quality of water for drinking. By calculating this index, the quality of water for drinking will be defined. World Health Organization [5] has presented the criteria of drinking water for calculating WQI [5]. The methods that are used for qualitative water evaluation by indices include multivariable analyses and multivariable spatial analyses (6). Nevertheless, due to incomplete output results, these indices have shortcomings in application as the only groundwater quality evaluation and quality monitoring method. These shortcomings can be associated with the fact that the evaluation and monitoring of water quality include many ambiguities from the sampling stage through the analysis stage[7].The other parameter which is more important than parameters selection is to cast limitation on the selected parameters which is of course subjective so that may restrict the application[6]. Also, the limitation of the number of selected parameters can influence the application of indices considerably. So, it seems that the existence of a new system for classifying the quality water data is essential [7].Currently, artificial intelligence is used as a trustful tool to transfer the human intelligence to the understandable data for computers [8].

Artificial Neural Network (ANN) is a valuable method for groundwater quality classification. Sharif Azari *et al*. (2011) used PNN to estimate the qualitative class of groundwater based on WQI criterion. Their results showed that PNN has appropriate efficiency in estimating this index[9] . Farbod *et al*. (2011) worked on qualitative zoning of water by using the fuzzy model, PNN, and geographic information system. Their results indicated that the efficiency of the proposed method as an efficient tool in qualitative zoning of water plans [10]. Khalili *et al*. (2005) applied four techniques: artificial neural network, Support Vector Machines(SVM), weighted regression method, and communicative vector machines to predict the pollution penetration through groundwater. They showed that these intelligent methods can appropriately replace the pretty complicated and time consuming old mathematical models for simulating nitrate concentration in groundwater. In their study SVM with RBF kernel was used [11]. **Dahiya** *et al.* in 2007 investigated the groundwater quality and presented the desired index using 10 parameters[7] .Bashi *et al*. (2010) for the first time estimated the groundwater pollution sources by using PNN and probability vector machines . Water quality indices can be used in investigating water quality of one plain for several statistical periods as well as for

IJSER

comparison of water quality of different plains.
Lacking data and also numerous parameters required for calculating the indices, make the qualitative water resources investigations difficult. So, it makes sound to seek the models that explain qualitative situation of water with few parameters[12]. In this research, the performance of PNN model in estimating different classes of water quality in Fars province in term of WQI criterion is investigated.

Materials and Methods

Study area

Fars is a southern province in Iran with an area of about 133000 square kilometers. The climate is cold in the north of province. In the central parts in winter, it is rainy and temperate; and in summer is dry and warm. In the south and southeast, it is rainy and moderate in winter and is warm in the summer. The average rainfall is about 40 billion m^3. The total discharge of groundwater is about 7.11 billion m^3. Water penetration into the aquifers resulting from the returned agriculture and other waters is about 8.2 billion m^3. In recent years, Fars faced with widespread droughts that resulted in decline of aquifers levels. So, the experts believe that necessary forecasts should be taken place to prevent pollution of water resources and in particular groundwater resources so that the droughts do not bring environmental catastrophes. In Fars there are 93 principal plains that are monitored in order to identify the aquifer decline. Pollution of the aquifers is contributed to the urban and residential regions drainage that leaks into the ground either directly or after discharge into drainage reservoirs. The experts believe that accumulating the drainage and industries deposits in pits or their arrival to the wells causes portion of it to leak into the groundwater. So far any instance of groundwater resources pollution is not observed in Fars; however the necessary precaution should be done regarding the aquifer level decline. Groundwater in the province is exploited from alluvial and lime aquifers. Also, there are many restrictions for groundwater resources. Therefore, groundwater resources are more extracted so that around %74 of agricultural water in the province is supplied by groundwater resources. Overexploitation of these resources in the most crucial plains such as Darab, Neyriz, Arsanjan, Abadeh, Jahrom, etc.in spite of being banned regions now, is very concerning.

.

Fig1. location map of the study site

WQI index:

Qualitative index WQI presents the ranking method in which qualitative conditions of water are defined based on different qualitative parameters influences. This index identifies quality situation regarding

ISSN 2229-5518

drinking consumptions. After calculating this index, quality situation for drinking consumption will be determined. WHO presented the drinking water standards for calculating WQI [11]. In this method the weight of each qualitative parameter is considered as the reverse of the standard weight defined by WHO. Qualitative parameters required for calculating this index and also relative weights corresponding to each parameter are presented in Table 1.

Table 1. Required parameters for calculating WQI index and their relative weights

Parameter	WHO's standard	WHO's proposed weight	relative weight
Ph	8.5	3	0.103
TDS	500	5	0.179
CL	250	5	0.179
SO_4	250	5	0.179
Na	200	4	0.143
K	12	2	0.071
HCO_3	120	1	0.036
Ca	75	3	0.107
Mg	50	3	0.107
		$\sum w_i = 28$	$\sum W_i = 1$

As we see from Table 1, parameters: all soluble solids, chlorine, and sulfate have the largest weights in calculating WQI. This implies the importance of these parameters in water quality definition for drinking consumption. The calculation of WQI includes following steps. At the first step, relative weight for each parameter is derived using the following equation.

$$W_i = \frac{w_i}{\sum_{i=1}^{n} w_i} \qquad (1)$$

In which W_i is the weight proposed by WHO, and w_i is the relative weight of parameter, and n is the number of parameters. At the next step, the quality order of each parameter regarding its concentration (milligram/liter) and the standard value presented by WHO is calculated using following equation.

$$q_i = \frac{C_i}{S_i} \times 100 \qquad (2)$$

Where, q_i is *i*-th parameter order, C_i is measured concentration of parameter, and S_i is the standard concentration of the parameter. At last, the value of index WQI for the assumed point is calculated using the following equation.

$$WQI = \sum W_i q_i \qquad (3)$$

IJSER

The obtained value of the index is usually classified for drinking consumptions in five classes: excellent, good, bad, very bad, and unusable. This classification is presented in Table2.

Table 2. Classification of underground water resources based on WQI

Index range	Class
< 50	Excellent Quality
50 - 100.1	Good Quality
100 - 200.1	Bad Quality
200 - 300.1	Very bad Quality
> 300	Water unusable for drinking

$$\%VE=\frac{\sum_{i=1}^{N}\frac{obs_i-for_i}{obs_i}}{n}\times 10 \qquad (7$$

Results and Discussions:

The information used in this study is presented in Table1. These data is associated to 2000 wells in Fars province in 2011. Using this information and by

Probability Neural Network (PNN):

PNN is a special type of RBF neural network which is adjusted for an increase in training speed, and often is used in classification. Fig. 1 shows simple darken and moody RBF network which includes one input layer, one hidden layer, and one output layer. Input layer includes units connecting to input vector. Input nodes are fully connecting to output units with weighted connections. Hidden layer contains units connecting to input vector. Input nodes fully connect to output units with weighted connections.

To evaluate the validity of the model, correlation coefficient (R^2), accumulative error, residual value coefficient which its positive value indicates that the estimated values are larger than observed values, and Root Mean Square Error (RMSE). Their relations are as following.

$$RMSE=\sqrt{\frac{(obs_i-for_i)^2}{n}} \qquad (7$$

$$R^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}\times\frac{\sum(y_i-\overline{y})^2}{n-1}}} \qquad (8$$

$$CRM=\frac{(\sum_{i=1}^{n}obs_i-\sum_{i=1}^{n}for_i)}{\sum_{i=1}^{n}obs_i} \qquad (9$$

IJSER

applying the relations presented in the previous section, the WQI parameter for each well is calculated. Fig. 2 demonstrates quality zoning map of the wells of the province based on WQI criterion.

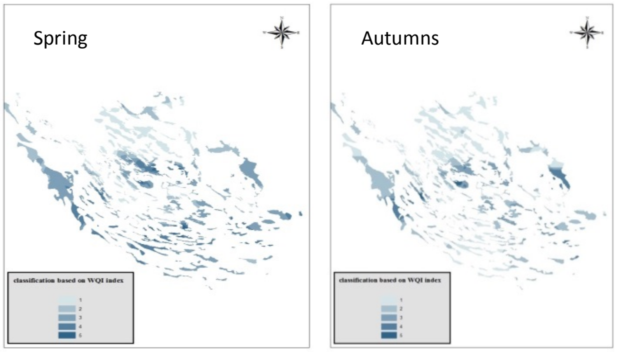


Fig2: Quality class changes in Fars province in spring and fall seasons

As Figure 2 demonstrates, quality class in spring declined so that water quality class reached to 5. Also, water quality declined in central plains. Central plains include Shiraz, Marvdasht, Kharameh, and Sarvestan which their qualities are lower than other plains. In most parts of the province in fall the quality is proper; but in spring mostly the northern plains have proper quality. Regarding the functional map, in the most regions where irrigated farming is done, the quality of water declined; however, in the dry farming regions a proper quality is observed. After deriving WQI index, the involved parameters in calculating the index are categorized in 10 different combinations as the inputs of PNN model. These combinations are listed in Table 3. Then, the sensitivity analysis is done by changing bandwidth in PNN. The combination with the least error and greater correlation coefficient as well as optimum bandwidth in PNN is defined. Table 4 shows RMSE of train and test stages.

The average values of RMSE in train and test stages in PNN model are 0.157904 and 0.2098 respectively.

Model	Optimal bandwidth	R^2	CRM	%VE

Table3.Combining input models

Input combination	Combination number	Input combinations	Combination number
TDS,CL,Na	6	CL,SO_4	1
TDS,SO_4,Na	7	TDS,CL,SO_4	2
Ca, Mg, PH	8	TDS,CL,SO_4 ,Na	3
Na, Mg, Ca	9	TDS,SO_4	4
TDS,Mg,Ca,CL, PH,SO_4	10	TDS,CL	5

IJSER

Table 4. The sensitivity analysis results for PNN model

No of combinations	1		2		3		4		5	
Stages	TS	TR	TS	TR	TS	TR	TS	TR	TS	TR
Optimal sigma	2	22	12	2	8	2	6	2	6	2
RMSE	0.2385	0.22254	0.2010	0.1074	0.2010	0.08655	0.1210	0.1085	0.1210	0.1046
No of combinations	6		7		8		9		10	
Stages	TS	TR	TS	TR	TS	TR	TS	TR	TS	TR
Optimal sigma	2	8	2	8	2	2	2	2	12	2
RMSE	0.1210	0.10784	0.1210	0.1074	0.5615	0.4610	0.2909	0.1887	0.1210	0.08495

Table 5 shows the results of applying the model.

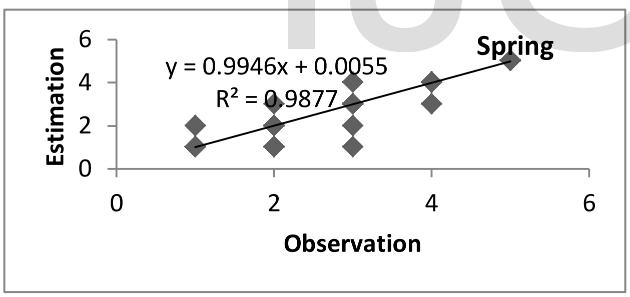
International Journal of Scientific & Engineering Research, Volume 4, Issue 9, September-2013 2482
ISSN 2229-5518

CL,SO_4	2	0.9671	0.0158	2.3116
TDS,CL,SO_4	12	0.9913	0.0025	0.7371
TDS,CL,SO_4,Na	8	0.9913	0.0025	0.7371
TDS,SO_4	6	0.9913	0.0025	0.7371
TDS,CL	6	0.9913	0.0025	0.7371
TDS, CL, Na	8	0.9913	0.0025	0.7321
TDS, SO_4, Na	8	0.9913	0.0025	0.7371
Ca, Mg, PH	2	0.8219	0.06333	8.5582
Na, Mg, Ca	2	0.9717	0.01166	1.8319
TDS, Mg, Ca, CL, PH, SO_4	12	0.9913	0.0025	0.7371

The best combination is the one with the least values of RMSE, cumulative error percentage, residual value coefficient, and also greater correlation coefficient. According to Table 5, the best combinations for PNN are TDS,CL,SO_4 - TDS,CL,SO_4,Na- TDS,SO_4 - TDS,CL-TDS,CL,NA- TDS,SO_4,Na- TDS,SO_4-CL,Ca,Mg,PH,SO_4,TDS; correlation coefficient, cumulative error percentage, residual value coefficient, and RMSE are 0.9913, 0.0025, 0.7371, 1210. The positive value of CRM implies that the estimated value is less than empirical value. It is also pretty ideal because of being close to zero. These values endorse the proper efficiency of PNN in classifying quality of groundwater.

After defining the optimum combination, the models with optimum bandwidth applied to data of fall. Because of its important role in calculating WQI and also the fewest number of parameters, combination 5 has been considered. Fig. 3 demonstrates the correlation between empirical and estimated data in both fall and spring for PNN model. The values of R^2 in both fall and spring are 0.9881 and 0.9771 respectively. The high correlation values imply the proper performance of PNN in estimating groundwater quality class.

IJSER



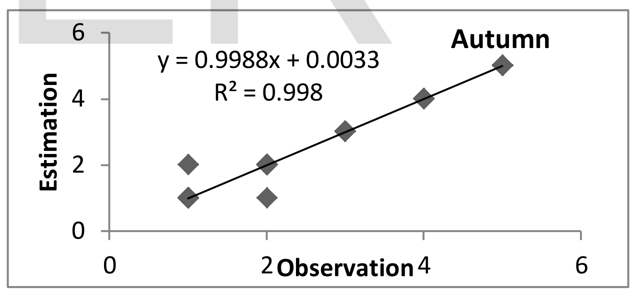


Fig. 3. Correlation diagram of observation and estimated data in spring and fall

Table 6 presents the percentage of quality classes, in both spring and fall, which derived by SPSS software.

Table 6: Classes of real data and the estimates in both autumn and spring

ISSN 2229-5518

Models \ Class	1	2	3	4	5
Experimental autumn	57.7	18.4	12.9	4.8	6.2
Autumn	57.5	18.5	12.9	4.8	6.2
Experimental Spring	52.2	22.5	12.7	7.1	5.5
Spring	52.5	22.4	12.5	7.1	5.5

According to Table 6, the highest percentage is for quality class 1, which implies that groundwater quality in most regions of Fars province is proper. Also, it indicates that the observed data are close to those estimated by PNN model.

We also worked on water quality zoning by using the predicted data by the model.

IJSER

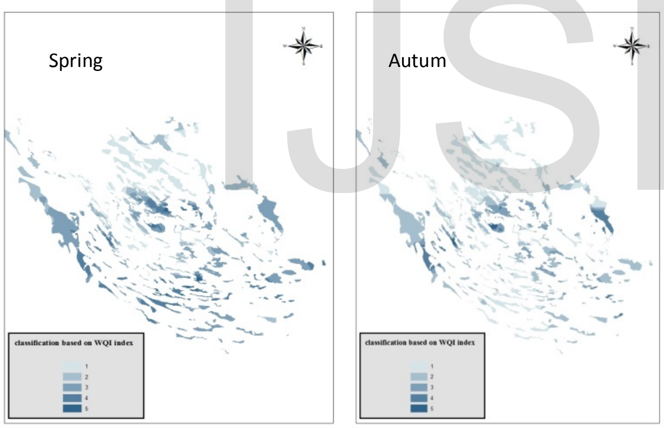


Figure 4: Changes in groundwater quality forecast model based on data fars Province (pnn)

Figure 4 shows the zoning of the plain. According to Figure 4, quality changes are close to those based on observed data which plotted in Figure 2. Also, it indicates that water quality in most plains of the province is proper. There is no considerable difference between Figures 2 and 4 that confirms the good performance of the model in estimating the index. The results show that PNN has a proper performance in estimating WQI.

Conclusions:

There are several methods for evaluating water quality situation. Among these methods, quality indices have numerous applications in investigating the quality of water resources. Calculating these indices in a specific region or in a specific period requires full quality information of parameters. WQI as one of these indices uses the information associated with nine quality parameters for calculating water quality classes. It is necessary to use the models including the fewest parameters for achieving to the most precise estimation [7]. One of the most famous and applicable methods in this ground is PNN which we used to estimate water quality class in the wells of Fars province. Parameters: chlorine and all solvable solids, because of the less error and also the importance in water quality class calculation based on WQI were considered as the inputs of the system. The results show that the system with the bandwidth as 6 has a proper performance in estimating the index. The estimated outcomes by the model are remarkably in accordance with the real outcomes and both outcomes imply the proper condition of groundwater resources quality in Fars province in term of WQI.

References

1. Ghasemiziarani, A. Faryadi, sh. Sheikhkazemi, sh.2006. zoning of pollution Karaj Dam watershed software with GIS, Environmental Engineering Conference - Tehran University - Faculty of Engineering.

2. Khodai,K. Shahsavari, A. Hatefi, R. Etebari,B.2005. Intrinsic aquifer vulnerability zoning Jovein against pollution using DRASTIC and GODS. Committee investigation Khorasan Regional Water Corporation.

3. Jafarzade,N. 2001. zoning, water quality and pollution sources karoon in the basin range studied basin reorganization plan. Proceedings of the Sixth International Conference on River Engineering, Ahvaz.

4. Nasrollahzadeh saravi,Z. Varedi,V.2001 survey Tajan river water quality using the curve of the quality index. Sixth International Conference River Engineering, Ahvaz, martyr Chamran University.

5.WHO. 2004 . Guidelines for drinking water quality: training pack. WHO, Geneva, Switzerland.

6.Stigter .T. Y, Ribeiro. L, Carvalho Dill , A. M. M. 2006. Application of a groundwater quality index as an assessment and communication tool in agro-environmental policies – Two Portuguese case studies. Journal of Hydrology (327): 578-591.

7.Dahiya, S. Singh, B. Gaur, S. Garg, V. K. and Kushwaha, H. S. 2007. Analysis of groundwater quality using fuzzy synthetic evaluation. *J. of Hazardous Materials* (1473), pp: 938-946.

8.Muhammetoglu, A. Yardimci ,A. 2006. A Fuzzy Logic Approach to Assess Groundwater Pollution Levels Below Agricultural Fields. Environmental Monitoring and Assessment 118:337-354.

9.Sharif azari, S. Araghi nezhad, SH. 2011. Performance evaluation of probabilistic neural network in estimating the Classification of groundwater quality based on WQI index. National Conference on Environmental Sensing and Modeling, Tehran University.

10.Farbod,A. Kerachian, R. Nikoo, M. 2011. zoning Water Quality Model Using Fuzzy Probabilistic Neural and Geographic Information System GIS: A Case Study, River Jajrud. The Sixth National Congress on Civil Engineering.

11.Khalil, A. Almasri, M. N. McKee, M. and Kaluarachchi, J. J. 2005, Applicability of statistical learning algorithms in groundwater quality modeling. *Water Resources Research* (41): 1-16.

12.Bashi-Azghadi, N. kerachian, R. Bazargan-lari, M. Solouki,K. 2010. Characterizing an unknown pollution source in groundwater resources systems using PSVM and PNN. *Expert Systems with Applications* 2010 (37):7154-7161.

13.Wasserman, P.1993. Advanced methods in neural computing. New York: Van Nostrand, Reinhold.

IJSER

IJSER