

Prediction of groundwater quality in the east of Thailand with discriminant analysis method

Jatupat Mekparyup¹ & Kidakan Saithanu^{2*}

^{1,2*}Department of Mathematics, Faculty of Science, Burapha University, Chonburi, Thailand

*[Email: ksaithan@buu.ac.th]

Received 05 January 2017; revised 29 March 2017

The 316 samples of groundwater wells in the east of Thailand were analyzed for consumption purpose. Concentrations of 21 various chemical parameters were considered to predict and then evaluate the quality of groundwater with discriminant analysis method and correct classification rate (CCR), respectively. The study results found stepwise discriminant analysis determined 5 significant parameters, Manganese (Mn), Copper (Cu), Iron (Fe), Fluorine (F) and Carbonate (CO₃), which discriminated the groundwater quality into 2 groups; nonstandard and standard quality groups. Derived discriminant model was capable correctly classified the group of groundwater quality with a rather high value of average CCR (78.80%) for the validation data set.

[Keywords:Groundwater, Groundwater Quality, Discriminant Analysis]

Introduction

Groundwater quality depends not only on natural factors such as aquifer lithology, groundwater velocity, quality of recharge waters and interaction with other types of water or aquifers, but also on human activities, which can alter these fragile systems, either by polluting them or by modifying the hydrological cycle, to such an extent that the use of groundwater becomes restricted¹. The topography of eastern Thailand is made up of various rivers, plain and hilly areas and are the chief sources of natural water. Industrial centers and urban coastal provinces, i.e. Chonburi and Rayong are located in this region. Hence, groundwater is popularly objective to reserve water for both of water source shortage and industrial areas. Many researchers had examined the groundwater quality with the standard for consumption by applying different statistical procedures. Most papers often used correlation and multiple regression analysis^{2,3,4,5,6}. There were a few studies which worked on the prediction of groundwater quality in Thailand by using multivariate analysis like discriminant analysis. Present study consists the stepwise discriminant method to fit the proper predictive model for predicting the quality of groundwater in the east of Thailand. Assumptions of discriminant analysis were also then tested whether they were satisfied. Average CCR of validation data set was finally determined for accuracy of the derived model.

Materials and Methods

The 316 groundwater wells in Chonburi and Rayong were taken as the representative wells of the provinces in the east of Thailand for industrial and rural areas. The 21 parameters, Depth of well (Depth), Electrical Conductivity (EC), PH, Calcium (Ca), Magnesium (Mg), Sodium (Na), Potassium (K), Iron (Fe), Manganese (Mn), Copper (Cu), Zinc (Zn), Sulfate (SO₄), Chlorine (Cl), Carbonate (CO₃), Hydrogen Carbonate (HCO₃), Fluorine (F), Nitrogen dioxide (NO₂), Nitrate (NO₃), Total Hardness as CaCO₃ (TH), Non-carbonate Hardness as CaCO₃ (NH) and **Total Dissolved Solids (TDS)**, were measured and provided in 2012 from the Department of Groundwater Resources, Ministry of Natural Resources and Environment, Thailand. All groundwater wells were divided into 2 quality groups, nonstandard (Group 0: 249 wells) and standard (Group 1: 67 wells), according to the chemical property standard of following 11 parameters for consumption⁷ as shown in Table 1.

The first step of data analysis was exploring and assessing each of independent variables with descriptive statistics. Secondly, Wilk's lambda statistic of multivariate analysis was applied to specify which independent variables were influential for classifying the quality groups of groundwater. Stepwise method of discriminant analysis was later employed to build the model for prediction of groundwater quality

group. If Box's M statistic of variance-covariance matrix between nonstandard and standard quality groups were equal, linear discriminant function was utilized. Otherwise, the quadratic discriminant function was interchanged⁸. Finally, the discriminant model was appraised in prediction of the quality groups of groundwater. A correct classification rate, one of favorite criterions, was chosen to evaluate the

model^{9,10} as $CCR = \frac{\sum_{k=0}^{C-1} CC_k}{n}$ where CC_k be the number of correctly allocated observations and n be the number of observation in the considered group.

Results

The descriptive statistics, mean, standard deviation, maximum and minimum, of independent variables displayed in Table 2.

The general characteristics of nonstandard quality group (Group 0) were the groundwater wells with the averages of depth 35.04 m, EC 1,299 microsiemens/cm and PH 7.31. The 6 independent variables affected and led to the nonstandard group were Fe, Mn, Cl, F, TH and TDS.

For the standard quality group (Group 1), were generally the groundwater wells with the averages of depth 32.15 m, EC 239.80 microsiemens/cm and PH 6.75. All 11 independent variables corresponded to the standard of chemical property of groundwater with the concentration averages of Fe 0.14, Mn 0.07, Cu 0.04, Zn 0.18, SO₄ 8.93, Cl 18.14, F 0.41, NO₃ 12.81, TH 78.92, NH 13.49 and TDS 154.3.

Table 1 — Chemical property of groundwater standard for consumption

Parameters	Maximum Acceptable Concentration (mg/l)	Maximum Allowable Concentration (mg/l)
Fe	<=0.5	1.0
Mn	<=0.3	0.5
Cu	<=1.0	1.5
Zn	<=5.0	15
SO ₄	<=200	250
Cl	<=250	600
F	<=0.7	1.0
NO ₃	<=45	45
TH	<=300	500
NH	<=200	250
TDS	<=600	1,200

Once Wilk's lambda statistic was applied, the test of difference between groups of groundwater quality was then shown in Table 3.

Based on small p-value (<0.05) of Wilk's lambda, there were 13 significant variables, EC, Ca, Mg, Na, Fe, Mn, Cu, SO₄, CO₃, HCO₃, F, TH and TDS, usefully influenced for classifying the quality groups of groundwater.

Table 2 — Descriptive statistics of independent variables employed for prediction of groundwater quality group

Variables	Group of Groundwater	Mean	Standard Deviation	Maximum	Minimum
Depth	0	35.04	35.36	1.00	150.0
	1	32.15	34.87	2.36	150.0
EC	0	1,299.00	3,942.00	6.79	3,2700.0
	1	239.80	176.00	12.00	700.0
PH	0	7.31	4.36	3.10	75.0
	1	6.75	0.83	4.60	8.4
Ca	0	63.39	111.63	0.00	1,100.0
	1	23.21	24.12	0.00	110.0
Mg	0	37.37	128.58	0.00	1,400.0
	1	5.16	4.67	0.00	23.0
Na	0	198.20	749.40	4.00	6,200.0
	1	16.60	15.58	4.00	86.0
K	0	16.63	55.25	0.20	610.0
	1	5.05	3.24	0.40	16.0
Fe	0	5.91	21.41	0.00	300.0
	1	0.14	0.14	0.00	0.5
Mn	0	1.11	2.09	0.00	19.0
	1	0.07	0.10	0.00	0.3
Cu	0	0.01	0.05	0.00	0.5
	1	0.04	0.05	0.00	0.1
Zn	0	0.58	3.71	0.00	49.0
	1	0.18	0.65	0.00	5.0
SO ₄	0	79.10	254.50	0.00	2,100.0
	1	8.93	15.33	1.00	80.0
Cl	0	361.60	1,454.10	1.50	12,000.0
	1	18.14	16.65	1.50	88.0
CO ₃	0	0.07	0.79	0.00	11.0
	1	0.69	3.98	0.00	30.0
HCO ₃	0	224.20	405.90	0.00	5,700.0
	1	92.60	93.00	2.00	343.0
F	0	1.09	1.63	0.40	15.0
	1	0.41	0.05	0.40	0.7
NO ₂	0	0.37	3.36	0.00	45.0
	1	0.03	0.04	0.00	0.3
NO ₃	0	13.21	29.55	0.30	240.0
	1	12.81	13.44	0.90	45.0
TH	0	323.40	798.30	10.00	6,600.0
	1	78.92	70.11	3.70	290.0
NH	0	163.30	668.00	0.00	5,200.0
	1	13.49	20.01	0.00	96.0
TDS	0	864.00	2,572.00	15.00	21,300.0
	1	154.30	113.70	8.00	455.0

For building the prediction model of groundwater quality group, the stepwise method was used as resulted in Table 4.

Only the 5 independent variables, Mn, Cu, F, Fe and CO₃, importantly influenced in classification of groundwater quality group because all these p-values of Wilk's lambda from stepwise method were closed to 0.

The assumption of multicollinearity was then required for testing as of Table 5 before building the discriminant model which was composed of all these five variables.

Since all VIF values were less than 10 also the p-values of t statistic were small (<0.05), it indicated there was no multicollinearity problem among these independent variables.

Table 3 — Test of difference between groups of groundwater quality

Variable	Wilk's lambda	F	df1	df2	P-value
Depth	.999	.355	1	314	.552
EC	.985	4.827	1	314	.029
PH	.996	1.118	1	314	.291
Ca	.973	8.554	1	314	.004
Mg	.987	4.193	1	314	.041
Na	.988	3.927	1	314	.048
K	.991	2.937	1	314	.088
Fe	.985	4.852	1	314	.028
Mn	.950	16.407	1	314	.000
Cu	.967	10.778	1	314	.001
Zn	.998	.756	1	314	.385
SO ₄	.984	5.074	1	314	.025
Cl	.988	3.729	1	314	.054
CO ₃	.983	5.271	1	314	.022
HCO ₃	.978	6.933	1	314	.009
F	.965	11.458	1	314	.001
NO ₂	.998	.712	1	314	.399
NO ₃	1.000	.011	1	314	.915
TH	.980	6.256	1	314	.013
NH	.989	3.360	1	314	.068
TDS	.984	5.081	1	314	.025

Table 4 — Selection of independent variables with stepwise method

Step	Enter	Stat	Wilks' lambda (df1, df2, df3)	Stat	Exact F (df1, df2)	p-value
1	Mn	.95	(1, 1, 314.00)	16.41	(1, 314.00)	.00
2	Cu	.91	(2, 1, 314.00)	15.81	(2, 313.00)	.00
3	F	.87	(3, 1, 314.00)	15.02	(3, 312.00)	.00
4	Fe	.85	(4, 1, 314.00)	13.72	(4, 311.00)	.00
5	CO ₃	.83	(5, 1, 314.00)	12.72	(5, 310.00)	.00

Regarding to p-value of Box's M statistic closed to 0, it meant the variance-covariance matrix between nonstandard and standard groups were not equal so the quadratic discriminant function should be chosen. Nevertheless, the linear discriminant function could be applied though occurring of heterogeneity of variance-covariance matrix as mention in¹¹. Therefore, the result of significance test of Wilks' Lambdastatistic for this discriminant function was appeared in Table 6.

From Table 6, the p-value of Wilks' Lambdaclosed to 0. That meant the linear discriminant function containing all these five independent variables was effectively discriminated the groups of groundwater quality.

Once the five influential variables were determined for building the prediction model, the estimated linear discriminant functions for each of the groundwater quality groups was formed as follows.

Nonstandard group:

$$\hat{Y}_0 = -1.192 + .003Fe + .311Mn + 3.575Cu - .012CO_3 + .537F$$

Where $\hat{Y}_0$ be the predicted score of nonstandard groundwater quality group.

Standard group:

$$\hat{Y}_1 = -1.169 - .026Fe + .043Mn + 21.148Cu + .201CO_3 + .137F$$

Where $\hat{Y}_1$ be the predicted score of standard groundwater quality group.

The predictive accuracy of discriminant model was

Table 5 — Test of multicollinearity among independent variables

Variable	Coefficient	Standard Error	t	p-value	VIF
Fe	-.004	.001	-3.016	.003	1.442
Mn	-.037	.012	-3.117	.002	1.169
Cu	2.451	.495	4.956	.000	1.257
CO ₃	.030	.011	2.751	.006	1.010
F	-.056	.014	-3.857	.000	1.007

Table 6 — Significance test of Wilks' Lambdastatistic for discriminant function of groundwater quality group

Test of Function(s)	Wilks' Lambda			
	Wilks' Lambda	Chi-square	df	p-value
1	.830	58.142	5	.000

Table 7 — Classification results of discriminant model

Put into group	Validation data set	
	True group	
	Nonstandard group	Standard group
Nonstandard group	222	40
Standard group	27	27
Total of observations	249	67
CCR	0.8916	0.4029
Average of CCR	0.7880	

finally evaluated with the CCR and average of CCR demonstrated in Table 7.

From Table 7, it indicated the predictive discriminant model rather performed a good performance with the average CCR 78.80% for both training and validation data set.

Discussion

This study investigated groundwater quality in the east of Thailand with discriminant analysis method. The quality of groundwater was classified into 2 groups, nonstandard and standard, in accordance with the chemical property standard. The results of study led to conclude and discuss as follows.

The stepwise discriminant determined only the five significant parameters (Fe, Mn, Cu, CO₃ and F), which discriminated the groundwater quality into 2 groups. That was in agreement with the results of descriptive statistics which indicated Fe, Mn and F were the causes leading to the nonstandard group. Therefore, the responsible agent could apply the obtained results for improving the quality of groundwater by emphasizing and remedying on these particularly three parameters.

Discriminant analysis method allowed a reduction in the dimensionality of the huge data set and illustrated a few significant parameters responsible for huge variations in groundwater quality that could reduce the number of sampling parameters. Thus, discriminant analysis method was useful and effective for groundwater quality management. In addition, the predictive discriminant model could be applied as an alternative tool for monitoring the groundwater quality instead of using only the chemical examination, i.e. argentometric which was complicated and expensive.

The obtained results could be extended by using this discriminant model for predicting the group of groundwater quality in any locations of Thailand or other countries which having similar topography or even applying other mathematical models, i.e. neural network.

Acknowledgement

This work was financially supported by the Research Grant of Burapha University through National Research Council of Thailand (Grant no. 81/2559). Authors sincerely thank to the Department of Groundwater Resources, Ministry of Natural Resources and Environment, Thailand for kindly furnishing all data.

References

- 1 Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M., & Fernandez, L., Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis, *Water Research*, 34(3)(2000) 807-816.
- 2 Bitton, G., Farrah, S. R., Ruskin, R. H., Butner, J., & Chou, Y. J., Survival of pathogenic and indicator organisms in ground water, *Ground Water*, 21(4)(1983) 405-410.
- 3 Cavalier, T. C., Lavy, T. L., & Mattice, J. D., Persistence of Selected Pesticides in Ground Water Samples, *Ground Water*, 29(2)(1991) 225-231.
- 4 Das, J., Sahoo, R. K., & Sinha, B. K., Urban ground water pollution: a case study in Cuttack city, India, *Groundwater Monitoring & Remediation*, 22(3)(2002) 95-103.
- 5 Joarder, M. A. M., Raihan, F., Alam, J. B., & Hasanuzzaman, S., Regression analysis of ground water quality data of Sunamganj District, Bangladesh, *Environmental Sciences Jagangirnagar*, 2(3)(2008) 291-296.
- 6 Daraigan, S. G., Wahdain, A. S., Ba-Mosa, A. S., & Obid, M. H., Linear correlation analysis study of drinking water quality data for Al-Mukalla City, Hadhramout, Yemen, *International Journal Of Environmental Sciences*, 1(7)(2011) 1692.
- 7 Law of Ministry of Natural Resources and Environment, *Royal Thai Government Gazette*, 2008, pp. 7-18.
- 8 Johnson, R. A., & Wichern, D. W., *Applied multivariate statistical analysis*, (Prentice Hall International. INC., New Jersey) 2007.
- 9 El-Sebakhy, E. A., Hadi, A. S., & Faisal, K. A., Iterative least squares functional networks classifier, *IEEE Transactions On Neural Networks*, 18(3)(2007) 844-850.
- 10 Oh, C., & Ritchie, S. G., Recognizing vehicle classification information from blade sensor signature, *Pattern Recognition Letters*, 28(9)(2007) 1041-1049.
- 11 Marks, S., & Dunn, O. J., Discriminant functions when covariance matrices are unequal, *Journal Of American Statistical Association*, 69(346)(1974) 555-559.